Evaluation of measurement properties of health-related quality of life instruments for burns: A systematic review

Catherine M. Legemate, MD, Inge Spronk, MSc, Lidwine B. Mokkink, PhD, Esther Middelkoop, PhD, Suzanne Polinder, PhD, Margriet E. van Baar, PhD,
and **Cornelis H. van der Vlies, MD, PhD**, *Amsterdam, the Netherlands*

BACKGROUND: Health-related quality of life (HRQL) is a key outcome in the evaluation of burn treatment. Health-related quality of life instruments with robust measurement properties are required to provide high-quality evidence to improve patient care. The aim of this review was to critically appraise the measurement properties of HRQL instruments used in burns.

METHODS: A systematic search was conducted in Embase, MEDLINE, CINAHL, Cochrane, Web of Science, and Google scholar to reveal articles on the development and/or validation of HRQL instruments in burns. Measurement properties were assessed using the Consensus-based Standards for the selection of health Measurement Instruments methodology. A modified Grading of Recommendations, Assessment, Development, and Evaluation analysis was used to assess risk of bias (prospero ID, CRD42016048065).

RESULTS: Forty-three articles covering 15 HRQL instruments (12 disease-specific and 3 generic instruments) were included. Methodological quality and evidence on measurement properties varied widely. None of the instruments provided enough evidence on their measurement properties to be highly recommended for routine use; however, two instruments had somewhat more favorable measurement properties. The Burn-Specific Health Scale—Brief (BSHS-B) is easy to use, widely accessible, and demonstrated sufficient evidence for most measurement properties. The Brisbane Burn Scar Impact Profiles were the only instruments with high-quality evidence for content validity.

CONCLUSION: The Burn Specific Health Scale—Brief (burn-specific HRQL) and the Brisbane Burn Scar Impact Profile (burn scar HRQL) instruments have the best measurement properties. There is only weak evidence on the measurement properties of generic HRQL instruments in burn patients. Results of this study form important input to reach consensus on a universally used instrument to assess HRQL in burn patients. (*J Trauma Acute Care Surg.* 2020;88: 555–571.

LEVEL OF EVIDENCE: Systematic review, level III.
KEY WORDS: Burn in injuries; health-related quality of life; outcome; measurement properties; PROM.

Because of the substantial advances in surgical and critical care management, the number of people surviving burns has increased during the past few decades.[1–3] As a result, more patients have to deal with lifelong disabilities and disfigurements, which are frequently a consequence of burn injury.[4] This has led to a shift in attention from clinician-led short-term outcomes, such as improvement of survival, to longer-term patient-centered outcomes of burn care focusing progressively on physical and psychological sequelae.[4–6] Therefore, perceived health-related quality of life (HRQL) of burn patients has become a key outcome in burn treatment.[7,8] Health-related quality of life is an outcome measure that reflects a patient's perception of his or her health condition on physical, psychological, and social well-being after an injury or disease.[9]

Patient-reported outcome measurement of HRQL offers an assessment of the patients' perspectives on burn care outcomes and is therefore useful in decision-making. Along with the variations in defining and operationalizing HRQL, a variety of patient-reported outcome measurement instruments (PROMs) to evaluate HRQL is currently available.[8,10] Measurement instruments to assess HRQL after burn injury are either generic (assessing general aspects of health) or disease-specific (covering aspects that are specifically relevant for burn patients), with benefits and disadvantages to the use of either type. Generic instruments allow comparison with the general population and other diseases, whereas burn-specific instruments include disease-specific items and may thus be better targeted to burn patients. Within burns, a subtype of burn-specific instruments has been introduced: instruments that assess the influence of burn scarring on HRQL.

Submitted: October 16, 2019, Revised: December 8, 2019, Accepted: December 25, 2019, Published online: January 17, 2020.
From the Department of Plastic, Reconstructive and Hand Surgery (C.M.L., I.S., E. M.), Amsterdam UMC, Vrije Univeristeit Amsterdam, Amsterdam Movement Sciences, Amsterdam; Association of Dutch Burn Centres (C.M.L., I.S., M.E.v. B., C.H.v.d.V.), Maasstad Hospital, Rotterdam; Department of Public Health (I. S., S.P., M.E.v.B.), Erasmus MC, University Medical Center Rotterdam, Rotterdam; Department of Epidemiology and Biostatistics (L.B.M.), Amsterdam UMC, Vrije Univeristeit Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, Amsterdam; Association of Dutch Burn Centres (E.M.), Red Cross Hospital, Beverwijk; and Trauma Research Unit, Department of Surgery (C.H.v.d.V.), Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands.
This study was presented at the 18th European Burn Association Congress, May 9, 2019, in Helsinki, Finland.
Supplemental digital content is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site (www.jtrauma.com).
Address for reprints: Cornelis H. van der Vlies, MD, PhD, Department of Surgery and Burn Centre, Maasstad Hospital, Maasstadweg 21, 3079 DZ, Rotterdam, The Netherlands; email: vliesc@maasstadziekenhuis.nl.

DOI: 10.1097/TA.0000000000002584

Selecting the best instrument to evaluate HRQL after burn injury requires the evaluation of specific instrument characteristics, feasibility of use (e.g. availability, patient compliance), and measurement properties. Measurement properties are quality aspects of a measurement instrument, such as reliability, validity, or responsiveness and provide information whether the results obtained by an instrument can be trusted. Health-related quality of life instruments with robust measurement properties in burn patients are required to draw valid conclusions about HRQL outcomes and, ultimately, to provide high-quality evidence to improve patient care. In this systematic review, the Consensus-based Standards for the Selection of health Status Measurement Instruments (COSMIN) methodology and guidelines[10–12] are used to critically appraise the measurement properties of HRQL instruments used in burn patients.

PATIENTS AND METHODS

This review was conducted according to the COSMIN methodology and the Preferred Reporting Items for Systematic Reviews and Meta-Analysis statement.[10,13] The protocol was registered a priori in the International Prospective Register of Systematic Reviews (CRD42016048065; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=48065).

Literature Search

A systematic literature search (no date or language restriction) was conducted in Embase, MEDLINE, CINAHL, Cochrane, Web of Science, and Google scholar on February 12, 2018. A medical librarian optimized the search strategy and performed the systematic search. The search strategy combined terms covering HRQL and the target population (patients with burn injury) (Supplemental Digital Content, http://links.lww.com/TA/B547). A combined library of the retrieved articles was created using Endnote, and duplications were excluded. The reference lists of included studies were hand searched for additional articles.

Article Selection and Data Extraction

Articles were included if they met the following criteria: (1) written in English, (2) published as full-text articles in a peer-reviewed journal, and (3) their purpose was the development and/or evaluation of the measurement properties of instruments that measure the construct HRQL in burn patients. Relevant articles were selected on the basis of title by one researcher (I.S.). Two researchers (C.M.L. and I.S.) independently screened a random sample of 10% of the abstracts. Because there was no disagreement between the reviewers, one reviewer (I.S.) appraised the remaining abstracts. At a second stage, two reviewers (C.M.L. and I.S.) assessed all full texts independently to identify studies evaluating measurement properties. Conflicts were resolved by consensus of the two reviewers and, if necessary, discussion with a third reviewer (M.E.v.B.). Data on characteristics of included studies and instruments, and results on measurement properties were extracted independently by two reviewers (C.M.L. and I.S.) and cross-checked. Evidence tables were used to summarize data.

Assessment of Methodological Quality of Included Studies

Two researchers (C.M.L. and I.S.) independently scored all quality assessment steps described hereafter. Any discrepancies were discussed and, if necessary, resolved with a third reviewer (M.E.v.B.). The COSMIN taxonomy was used to select which measurement properties of an instrument were evaluated (Table 1).[15] Because there is no criterion standard for HRQL, criterion validity was not considered. Individual articles may comprise more than one study if they evaluate more than one measurement property or the same measurement property for more than one HRQL instrument. The COSMIN Risk of Bias checklist was used to assess the methodological quality for each study.[10–12] Studies were stratified as having very good, adequate, doubtful, or inadequate methodological quality. More detailed information on the COSMIN Risk of Bias checklist can be found elsewhere (http://www.cosmin.nl).

Assessment of Measurement Property Results

The result of each study on a measurement property was rated against criteria for good measurement properties: sufficient (+), insufficient (−), or indeterminate (?) (Table 1). Evidence on relevance, comprehensiveness, and comprehensibility (aspects of content validity) was derived from development and content validity studies in which patients and/or professionals were involved. This was done first based on the methods and results of the instrument development study; second, based on each available content validity study of the specific instrument; and third, based on the reviewer's own rating of the content of the instrument (i.e., assessment of coverage of burn-specific consequences, which was a subjective assessment of both reviewers on all items in each included HRQL instrument because no precedent exists).[14] If instruments were not freely available, developers of the instrument were contacted. If they were not willing to distribute the instrument, the review team could not evaluate the content.

Regarding hypothesis testing and responsiveness, we predefined that correlations with (domain scores of) other outcome measurements that aim to measure related constructs should be 0.30 or greater[16] and there should be significant differences in scores between relevant subgroups. Subgroups were based on the results of a previous systematic review on predictors of HRQL in burn patients and involved factors determining burn severity: percentage of total body surface area (TBSA) burned, length of hospital stay, and the necessity of surgery.[17]

SYNTHESIS OF EVIDENCE AND RECOMMENDATIONS

All results per measurement property of each HRQL instrument were checked for consistency, and seven were qualitatively summarized. These summarized results were evaluated against the criteria for good measurement properties to produce an overall rating (sufficient (+), insufficient (−), inconsistent (±), or indeterminate (?)) for each measurement property of each HRQL instrument.[10] The focus was on the HRQL instrument specifically, while in the previous steps the focus was on the single studies.

TABLE 1. Definitions (Mokkink et al.[14]) and Criteria for Good Measurement Properties (Prinsen et al.[10])

Domain Measurement Property	Definition	Rating	Criteria
Reliability	The degree to which the measurement is free from measurement error		
Reliability (extended definition)	The extent to which scores for patients who have not changed are the same for repeated measurements under several conditions		
Internal consistency	The degree of the interrelatedness among the items	+	At least low evidence for sufficient structural validity *and* Cronbach α's ≥0.70 for each unidimensional scale or subscale
		?	Criteria for "At least low evidence for sufficient structural validity" not met
		−	At least low evidence for sufficient structural validity *and* Cronbach α's ≥0.70 for each unidimensional scale or subscale
Reliability	The proportion of the total variance in the measurements which is due to *true* differences between patients	+	ICC *or* weighted κ ≥0.70
		?	ICC *or* weighted κ not reported
		−	ICC *or* weighted κ <0.70
Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured	+	SDC *or* LoA < MIC
		?	MIC not defined
		−	SDC *or* LoA > MIC
Validity	The degree to which an HRQL instrument measures the construct(s) it purports to measure		
Content validity	The degree to which the content of an HRQL of life instrument is an adequate reflection of the construct to be measured		
Relevance	The degree to which items in an HRQL instrument are relevant for the construct of interest within a specific population and context of use	+	≥85% of the items of the HRQL instrument fulfill the criterion*
		?	No (or not) enough information available or quality of (part of a) the study inadequate
		−	<85% of the items of HRQL instrument fulfill the criterion
Comprehensiveness	The degree to which key aspects of the construct are missing		Idem relevance*
Comprehensibility	The degree to which items are understood by patients as intended		Idem relevance*
Construct validity	The degree to which the scores of an HRQL instrument are consistent with hypotheses based on the assumption that the HRQL instrument validly measures the construct to be measured		
Structural validity	The degree to which the scores of an HRQL instrument are an adequate reflection of the dimensionality of the construct to be measured	+	CTT: CFA:CFI or TLI or comparable measure >0.95 *or* RMSEA <0.06 *or* SRMR <0.08 IRT/Rasch: no violation of unidimensionality (CFI or TLI or comparable measure >0.95 *or* RMSEA <0.06 *or* SRMR <0.08) *and* no violation of local independence (residual correlations among the items after controlling for the dominant factor <0.20 *or* Q3's <0.37) *and* no violation of monotonicity (adequate looking graphs *or* item scalability >0.30 *and* adequate model fit: IRT, χ^2 > 0.01; Rasch, infit and outfit mean squares ≥0.5 and ≤1.5 *or* *Z*-standardized values >−2 and <2)
		?	CTT: not all information for + reported IRT/Rasch: model fit not reported
		−	Criteria for + not met
Hypotheses testing	Item construct validity	+	At least 75% of the result is in accordance with the hypotheses[2] or no differences between groups reported[3]
		?	No correlations with instrument(s) measuring related construct(s) *or* no differences between groups reported[4]
		−	Criteria for + not met

Continued next page

TABLE 1. *(Continued)*

Domain Measurement Property	Definition	Rating	Criteria
Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted HRQL instrument is an adequate reflection of the performance of the items of the original version of the HRQL instrument	+	No important differences found between group factors (such as age, sex, language) in multiple group factor analysis *or* no important DIF for group factors (McFadden's $R^2 < 0.02$)
		?	No multiple group factor analysis *or* DIF analysis performed
		−	Important differences between group factors *or* DIF was found
Criterion validity	The degree to which the scores of an HRQL instrument are an adequate reflection of a criterion standard	+	Correlation with criterion standard ≥0.70 or AUC ≥0.70
		?	Not all information for + reported
		−	Correlation with criterion standard <0.70 *or* AUC <0.70
Responsiveness	The ability of an HRQL instrument to detect change over time in the construct to be measured	+	The result is in accordance with the hypothesis *or* AUC ≥0.70
		?	No hypothesis defined (by the review team)
		−	The result is not in accordance with the hypothesis *or* AUC <0.70.

1. + indicates sufficient; –, insufficient;?, indeterminate.
2. Correlations with instruments measuring the same construct >0.50 *or* at least 75% of the results are in accordance with the hypotheses.
3. Known groups were based on factors determining burn severity: percentage of total body surface area burned, length of stay, and surgery (yes or no).
4. No hypotheses defined.

*Criteria on relevance, comprehensiveness, and comprehensibility can be found on www.comsin.nl.

AUC, area under the curve; CFA, confirmatory factor analysis; CFI, comparative fit index; CTT, classical test theory; DIF, differential item functioning; ICC, intraclass correlation coefficient; IRT, item response theory; LoA, limits of agreement; MIC, minimal important change; RMSEA, root mean square error of approximation; SDC, smallest detectable change; SRMR, standardized root mean residuals; TLI, Tucker-Lewis index.

The Grading of Recommendations, Assessment, Development, and Evaluation approach was used to grade the quality of the evidence, determining the trustworthiness of the summarized results. For content validity, the evidence quality could be downgraded because of risk of bias (as determined using the COSMIN Risk of Bias checklist), inconsistency of results across studies and indirectness (i.e., evidence from different populations) (Supplemental Digital Content 2, http://links.lww.com/TA/B547). For the other measurement properties, the evidence quality could be downgraded because of risk of bias, imprecision (i.e., low sample size), inconsistency, and indirectness (Supplemental Digital Content 2, http://links.lww.com/TA/B547).[10]

To come to an evidence-based and transparent recommendation, the instruments were categorized in three categories.[10] According to the COSMIN guidelines, instruments with sufficient content validity and sufficient internal consistency can be recommended for use (category A), PROMs can have the potential to be recommended for use (category B), and PROMs with high-quality evidence for an insufficient measurement property should not be recommended for use (category C).[10] The COSMIN guidelines indicate that if all instruments fall in category B, the most important property of a measurement instrument is content validity, followed by structural validity and internal consistency. Subsequently, the results of the other measurement properties should be considered.

In addition, information on feasibility was appraised to determine the feasibility of use, so recommendations would not only be based on the measurement properties. Important aspects of feasibility were defined as length of the instrument, completion time, and ease of score calculation and access fee of an instrument.

RESULTS

Of the 7,246 records identified, 43 articles were considered eligible for assessment (Fig. 1, Table 2). These 43 articles evaluated 15 different HRQL instruments. Most articles studied more than one measurement property; the included articles comprised 118 separate studies. Of the HRQL instruments identified, 3 were generic, and 12 were disease specific. Of these 12 instruments, 4 instruments measured the impact of burn scarring on HRQL (Table 3). Six instruments were specifically developed for the use in children (one generic, five disease specific, of which three on burn scarring). The most frequently appraised instruments were all burn-specific HRQL instruments: the Burns Specific Health Scale—Brief (BSHS-B[16,31–49]), Burn Specific Health Scale—Abbreviated (BSHS-A[26–30]), and Burn Specific Health Scale—Revised (BSHS-R[50–53]) (Table 2; Supplemental Digital Content 3). Of the instruments that were specifically for the use in children, the Burn Outcome Questionnaire 5 to 18 (BOQ 5–18 y) was the one most frequently appraised[22–24] (Table 2; Supplemental Digital Content 3, http://links.lww.com/TA/B547).

General characteristics of the included articles and instruments are summarized in Table 2 and Table 3, respectively.[16,18–59] Table 3 also includes feasibility aspects of each HRQL instrument.

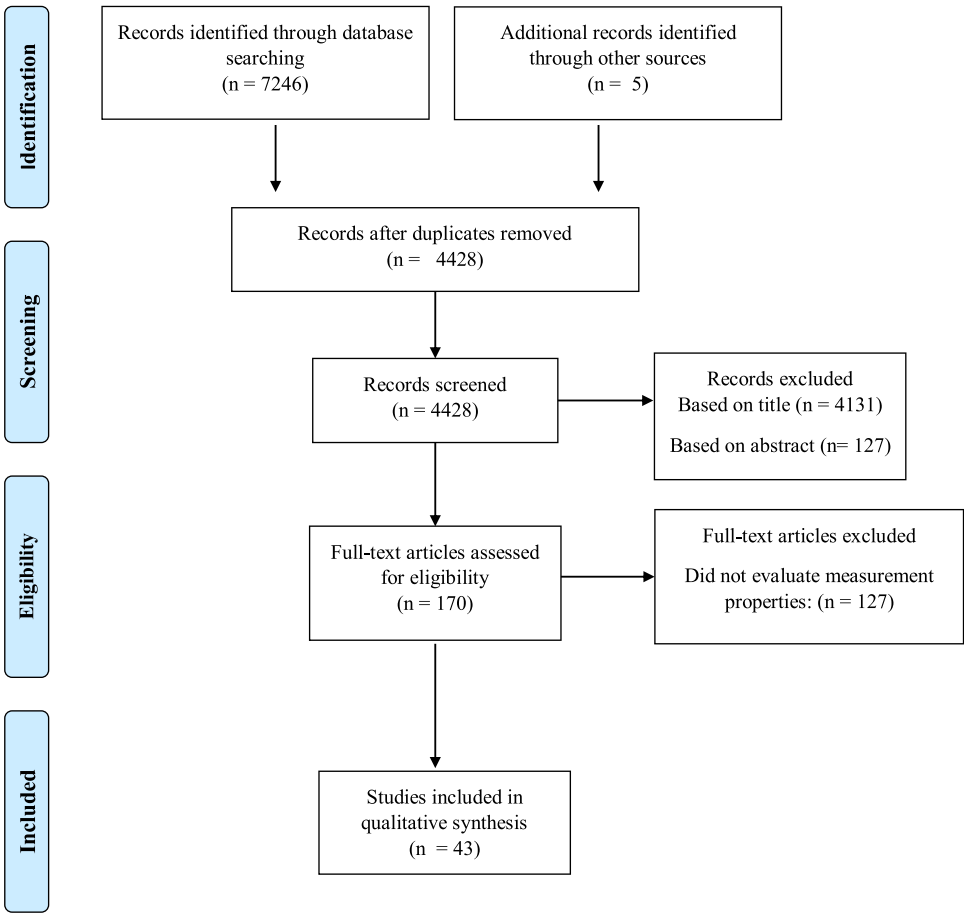

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-analysis flow diagram demonstrating the identification and screening of studies for inclusion.

The most commonly assessed measurement properties were internal consistency, hypotheses testing, and reliability. No study assessed measurement error. Methodological quality and evidence on measurement properties were variable (Supplemental Digital Content 3, http://links.lww.com/TA/B547).

Table 4 presents the results of the best evidence syntheses. All instruments were categorized as level B instruments: PROMs that have the potential to be recommended based on their measurement properties.

Measurement Properties of Generic HRQL Instruments

The three generic HRQL measurement instruments include the EuroQol 5 dimensions (EQ-5D), the 47-item short form Infant Toddler Quality of Life Questionnaire (ITQOL-SF47), and the 36-item short form survey (SF-36). There was only weak evidence on the measurement properties of generic HRQL instruments in burns. The comprehensiveness of all of these instruments was rated insufficient because these instruments did not cover all the aspects of HRQL that are relevant to patients with burn injury (e.g., problems related to scarring). There was high-quality evidence for sufficient hypotheses testing for construct validity of the EQ-5D and SF-36, but studies on other measurement properties in burns were lacking.[16,56,58] Both scales are widely available and especially the SF-36 is widely applied within the field of burns.[8] In terms of feasibility, a limitation of the SF-36 is the license fee. Structural validity and internal consistency of the ITQOL-SF47 were studied, but both were rated as indeterminate.[57]

Measurement Properties of Burn-Specific HRQL Instruments

The 12 disease-specific HRQL instruments were as follows: the Brisbane Burn Scar Impact Profile (BBSIP) for adults, BBSIP for children 8 to 18 years, BBSIP for caregivers of children younger than 8 years, BBSIP for caregivers of children 8 to 18 years, BOQ 0 to 4, BOQ 5 to 18 years, Burn Specific Health Scale (BSHS), BSHS-A, BSHS-B, BSHS-R, Dermatology Life Quality Index, and the Young Adult Burn outcome Questionnaire (YABOQ) (Table 3).

The different versions of the BBSIP focus on the impact of burn scarring on HRQL and are the only instruments with moderate to high-quality evidence for sufficient content validity, which is the most important measurement property according to the COSMIN guideline. The BSHS-B is the only instrument with high-quality evidence for internal consistency, which is (together with structural validity) the second important measurement property according to the COSMIN guideline. Therefore,

TABLE 2. Characteristics of the Included Articles

HRQL Instrument	Article	Year	Country (Language)	Measurement Properties Studied	Sample Size	Adults/ Children	Age at Study, Mean (SD) in years	Sex (Male)	Time Postburn	%TBSA Mean (SD)
BBSIP (adults)	Tyack et al.[18]	2015	Australia (English)	– Content validity	n = 10	≥8 y	49 (34)	60%	Median, 12.5 (31) mo	14 (20)
	Tyack et al.[19]	2017	Australia (English)	– Internal consistency – Reliability – Hypotheses testing – Responsiveness	n = 118	Adults (>18 y)	Median, 34	74%	Baseline, around the time of wound healing, 1 to 2 wk postbaseline, and 1 mo postbaseline	Median, 4
BBSIP (caregiver 0–8 y)	Tyack et al.[18]	2015	Australia (English)	– Content validity	n = 9	Children (0–8 y)	2.5 (1.5)	44%,	8 (18)	8 (14)
BBSIP (caregiver 8–18 y)	Tyack et al.[18]	2015	Australia (English)	– Content validity	n = 11	Children (8–18 y)	13 (1.8)	64%	Median, 10 (27) mo	Child, 7.5 (18)
BBSIP (children 8–18 y)	Tyack et al.[18]	2015	Australia (English)	– Content validity	n = 11	Children (8–18 y)	Child, 13 (1.8)	64%	Median, 10 (27) mo	7.5 (18)
BOQ 0–4 y	Kazis et al.[20]	2002	United States (English)	– Structural validity – Internal consistency – Reliability – Hypotheses testing	n = 184	Children (0–5 y)	2.55 (1.3)	54%	Baseline and 6 mo	17
	van Baar et al.[21]	2006	The Netherlands (Dutch)	– Content validity – Internal consistency – Reliability – Hypotheses testing	n = 194	Children (0–4 y)	36.7 (5–63) mo	54%	Mean, 17.5 (9.8) mo	6 (0–66)
BOQ 5–18 y	Daltroy et al.[22]	2000	United States (English)	– Internal consistency – Hypothesis testing	n = 86	Children (5–18 y)	10	71%	NR	22 (NR)
	van Baar et al.[23]	2006	The Netherlands (Dutch)	– Content validity – Internal consistency – Reliability – Hypotheses testing	n = 145	Children (5–15 y)	8.8 (3.6)	65%	Mean, 21.1 (10.3) mo	6.0 (2.0)
	Sveen et al.[24]	2012	Sweden (Swedish)	– Content validity – Internal consistency – Reliability – Hypotheses testing	n = 70	Children (5–18 y)	9.5 (3.5)	67%	Mean, 5.4 (2.4) y	10.5 (12.7)
BSHS	Blades et al.[25]	1982	United States (English)	– Internal consistency	n = 40	Adults	32.1 (Range, 18–55)	NR	Mean, 35.2 wk (range, 4–107 wk)	30.2 (Range, 11–80)
BSHS-A	Adam et al.[26]	2009	Turkey (Turkish)	– Internal consistency – Reliability – Hypotheses testing	n = 53	≥16 y	33.74 (13.2)	81%	2 wk	19.9 (12.5)
	Li et al.[27]	2014	China (Chinese)	– Content validity – Internal consistency – Reliability – Hypothesis testing	n = 457	Adults (≥18 y)	36.66 (15.3)	70%	Mean, 13.63 mo	39.6 (27.3)
	Moi et al.[28]	2003	Norway (Norwegian)	– Content validity – Internal consistency – Reliability – Hypothesis testing	n = 95	Adults (≥18 y)	43.7 (14.7)	82%	Mean, 47.0 (23.8) mo	18.5 (14.2)
	Munster et al.[29]	1987	USA (English)	– Internal consistency – Reliability	n = 70	Adults	NR	NR	NR	NR

	Salvador Sanz et al.[30]	1998	Spain (Spanish)	– Content validity – Internal consistency – Reliability – Hypotheses testing	n = 115	Adults (16–73 y)	40.5 (16.7)	54%	Mean, 783 (801.4) d	14 (11.6)
BSHS-B	Finlay et al.[31]	2014	Australia (English)	– Structural validity – Internal consistency – Hypotheses testing	n = 224	≥16 y	36 (16–84)	83%	1 mo and ≥6 mo	4 (Range, 1–60)
	Gandolfi et al.[32]	2016	France (French)	– Internal consistency – Reliability – Hypotheses testing	n = 53	Adults (18–70 y)	46.4 (15.9)	66%	Range, 2–4 y	26.9 (15.9)
	Goudarzian et al.[33]	2017	Iran (Persian)	– Structural validity – Internal consistency	n = 410	Pregnant women	39.36 (12.30)	0%	NR	19.1 (2.16)
	Hwang et al.[34]	2016	Taiwan (Chinese Taiwanese)	– Internal consistency – Reliability – Hypotheses testing	n = 108	Adults	42.1 (13.3)	64%	Mean, 565 d (range, 3–3,965 d)	23.3 (25.4)
	Kildal et al.[35]	2001	Sweden (Swedish)	– Structural validity – Internal consistency – Hypotheses testing	n = 248	Adults (≥18 y)	36.8 (16.1)	80%	Mean, 9.3 y	23.1 (16.2)
	Ling-Juan et al.[36]	2012	China (Chinese)	– Content validity – Structural validity – Internal consistency	n = 208	Adults (≥18 y)	40.42 (13.16)	77%	Mean, 37.12 mo	40.1 (27.4)
	Meirte et al.[16]	2017	The Netherlands and Belgium (Dutch)	– Hypothesis testing	n = 184	Adults	39.0 (12.8)	71%	9 mo	11.8 (10.2)
	Mulay et al.[37]	2015	India (Hindi)	– Content validity – Structural validity – Internal consistency – Reliability	n = 20	Adults (18–65 y)	31.0 (22–55)	40%	Between 6 mo and 12 mo	39.8 (Range, 20–60)
	Muller et al.[38]	2015	Germany (German)	– Structural validity – Internal consistency – Hypothesis testing	n = 141	Adults (≥18 y)	49.62 (15.16)	65%	Mean, 45.01 mo	12.9 (10.3)
	Piccolo et al.[39]	2015	Brazil (Brazilian Portuguese)	– Content validity – Internal consistency – Reliability – Hypotheses testing	n = 92	Adults (≥18 y)	37.12 (12.12)	52%	Mean, 4.38 y (range, 1–30 y)	19.1 (18.8)
	Pishnamazi et al.[40]	2013	Iran (Persian)	– Content validity – Structural validity – Internal consistency – Reliability – Hypotheses testing	n = 200	Adults (≥18 y)	25 (6.8)	38%	NR	34.9 (2.0)
	Sideli et al.[41]	2010	Italy (Italian)	– Internal consistency – Hypotheses testing	n = 50	Adults	40.12 (15.67)	72%	1 mo after admission	31.8 (17.6)
	Sideli et al.[42]	2014	Italy (Italian)	– Internal consistency – Hypotheses testing	n = 131	Adults (18–65 y)	40.21 (12.33)	53%	Within 6 mo from admission	16.8 (12.2)
	Stampolidis et al.[43]	2012	Greece (Greek)	– Internal consistency – Reliability – Hypotheses testing	n = 40	Adults	52.20 (18.49)	74%	During the first month of hospitalization	15.6 (13.0)
	Stavrou et al.[44]	2015	Israel (Hebrew)	– Content validity – Internal consistency – Reliability – Hypotheses testing	n = 86	Adults (≥18 y)	38.04 (18.22)	79%	12.71 (13.29) mo	11.1 (11.6)

Continued next page

TABLE 2. *(Continued)*

HRQL Instrument	Article	Year	Country (Language)	Measurement Properties Studied	Sample Size	Adults/ Children	Age at Study, Mean (SD) in years	Sex (Male)	Time Postburn	%TBSA Mean (SD)
	Stolle et al.[45]	2017	Germany (German)	– Structural validity – Internal consistency – Reliability – Hypotheses testing	n = 364	Adults (≥18 y)	44.7 (16.8)	69%	Range, 1–50 y	9.5 (12.2)
	Szczechowicz et al.[46]	2014	Poland (Polish)	– Content validity – Internal consistency – Reliability – Hypotheses testing	n = 190	Adults (≥17 y)	46.49 (18.35)	66%	NR	23.6 (15.7)
	Van Loey et al.[47]	2013	Swedish and Dutch	– Cross-cultural validity	Swedish, n = 231; Dutch, n = 275	Swedish, 16–93 y; Dutch, 18–88 y	Swedish, 45.6 (18.5); Dutch, 39.3 (13.2)	Swedish, 72.7%; Dutch, 73.5%	Swedish, 9 mo; Dutch, 12 mo	Swedish, 21.6 (18.2); Dutch, 11.8 (11.2)
	Willebrand and Kildal[48]	2008	Sweden (Swedish)	– Structural validity – Internal consistency	n = 334	Adults (≥18 y)	46.4 (16.0)	78%	Mean, 7.9 (4.8) y	21.6 (16.0)
	Willebrand and Kildal[49]	2011	Sweden (Swedish)	– Structural validity – Internal consistency – Hypotheses testing	n = 94	Adults (≥18 y)	44.4 (15.8)	73%	6 mo, 12 mo, and 24 mo	23.4 (19.6)
BSHS-R	Blalock et al.[50]	1992	United States (English)	– Content validity	n = 38	Adults (≥18 y)	43.2 (16.8)	81.6%	Mean, 362 d	26.5 (14.4)
	Blalock et al.[51]	1994	United States (English)	– Structural validity – Internal consistency – Hypotheses testing	n = 254	Adults	39.3 (14.2)	74%	Mean, 313 d	19.2 (15.1)
	Ferreira et al.[52]	2008	Brazil (Brazilian Portuguese)	– Structural validity – Internal consistency – Hypotheses testing	n = 115	Adults (≥18 y)	31.8 (13.4)	66%	NR	19.3 (16.0)
	Nicolosi et al.[53]	2013	Brazil (Brazilian Portuguese)	– Internal consistency	n = 63	12–20 y	16.0 (2.9)	30%	NR	23.8
DLQI	Finlay and Khan[54]	1994	United Kingdom (English)	– Content validity	n = 120	Adults (15–17 y)	Median, 42 y	42%	NA	NA
	Mazharinia et al.[55]	2007	Iran (Persian)	– Structural validity – Internal consistency	n = 109	Adults (≥16 y)	28.9 (11.43)	44%	NR	NR
EQ-5D	Oster et al.[56]	2009	Sweden (Swedish)	– Hypothesis testing	n = 78	Adults (≥18 y)	43.6 (15.1)	78%	At baseline, 3 mo, 6 mo, and 12 mo	24.3 (19.7)
	Meirte et al.[16]	2017	The Netherlands and Belgium (Dutch)	– Hypothesis testing	n = 184	Adults	39.0 (12.8)	71%	9 mo	11.8 (10.2)
ITQOL-SF47	Landgraf et al.[57]	2013	The Netherlands (Dutch)	– Structural validity – Internal consistency	n = 194	Children	36.7 (5–63) mo	54%	Mean, 17.5 mo	6 (0–66)
SF-36	Edgar et al.[58]	2000	Australia (English)	– Hypotheses testing	n = 280	≥16 y	37.4 (15.1)	81%	1 mo, 3 mo, 6 mo, and 12 mo	8.9 (11)
	Meirte et al.[16]	2017	The Netherlands and Belgium (Dutch)	– Hypothesis testing	n = 184	Adults	39.0 (12.8)	71%	9 mo	11.8 (10.2)
YABOQ	Ryan et al.[59]	2013	United States (English)	– Content validity – Structural validity – Internal consistency – Reliability	n = 153	Adults (19–30 y)	24.7 (3.6)	73%	At baseline contact, 2 wk, and 6 mo and 12 mo	11 (14)

BDI, Beck Depression Inventory; ITQOL—SF 47, Infant Toddler Quality of Life Questionnaire, the 47-item short-form; NA, not applicable (e.g., dermatology life quality index [DLQI], not developed in burn patients); NR, not reported.

these instruments will be discussed in more detail. Regarding the other instruments, it is of note that these are not necessarily inadequate but that their measurement properties are merely not or scarcely investigated in literature.

Brisbane Burn Scar Impact Profile

The BBSIP was developed in 2013 to assess burn scar–specific HRQL in burn patients at risk of or with burn scars.[18] Multiple versions were developed for different age groups (Table 3). International scar management experts and patients were involved in the development of the items, and cognitive interviews were done to understand how patients interpreted the items.[18] Nevertheless, the overall rating of comprehensiveness was judged to be doubtful for all versions because patients were not asked about the comprehensiveness of the final developed forms. Other content validity studies were not encountered.

BBSIP Adult Version

The adult versions of the BBSIP consists of 66 items divided into 7 subscales (Table 3). One study reported that Cronbach α was 0.7 or greater for all subscales,[19] but the quality of the study was rated doubtful and the overall rating of internal consistency was indeterminate because there were no studies on structural validity (Supplemental Digital Content 3, http://links.lww.com/TA/B547; Table 4). Reliability and hypotheses testing for construct validity were sufficient; however, the evidence was graded as moderate as a consequence of downgrading for risk of bias (i.e., only one study of adequate quality was available). One study provided high-quality evidence for sufficient responsiveness.

BBSIP Child Versions

The version of the BBSIP for children 8 to 18 years consists of 58 items divided into 7 subscales. The BBSIP for caregivers of children younger than 8 years and the BBSIP for caregivers of children 8 to 18 years both comprise an extra subscale to measure parent and family concerns and consist of 58 and of 62 items, respectively.[18] No studies on other measurement properties of the child or caregiver versions of the BBSIP were revealed in our systematic search.

Regarding the feasibility of the different versions of the BBSIP, currently, all versions are only available in English, but validated translation studies may emerge in the future. To reach the level A status, it is vital that structural validity is assessed to determine if the item on the scales sufficiently measures the same construct.

Burn-Specific Health Scale—Brief

The 40-item BSHS-B was derived from items of the BSHS and BSHS-R in 2001.[29,35,51] Despite a development process that involved patients and featured a pilot study, comprehensibility was the only aspect of content validity that was rated sufficient (Table 4).

Relevance and comprehensiveness of the BSHS-B were rated inconsistent as a result of conflicting results of multiple studies (Supplemental Digital Content 3, http://links.lww.com/TA/B547).[31–46,48,49] The BSHS-B consists of nine subscales that have been confirmed in one study that used confirmatory factor analysis with an adequate sample size, which was therefore of very good quality.[45] Nevertheless, some studies that were of inferior quality because they used exploratory factor analysis and/or had an inadequate sample size showed other results, and therefore, the overall quality of structural validity was graded moderate.[33,38] The BSHS-B carries high-quality evidence for sufficient internal consistency, reliability, and very low–quality evidence for sufficient cross-cultural validity. Furthermore, moderate quality evidence for sufficient hypotheses testing for construct validity was found.

The BSHS-B carries the best evidence for sufficient measurement properties (Table 4). It has been studied extensively (Table 1), resulting in good-quality evidence for sufficient structural validity, internal consistency, and cross-cultural validity. The instrument is relatively short and freely available in 14 languages. Nevertheless, there is only low to moderate evidence on sufficient content validity (which is the most important measurement property according to the COSMIN guidelines). Of note is that especially relevance and comprehensives of the BSHS-B should therefore be investigated further.

DISCUSSION

This systematic review provides a comprehensive overview of all available studies on measurement properties of instruments used to assess HRQL in burn patients. Recently updated, consensus-based standards, developed by the COMSIN initiative,[10,11,15] were used to ascertain sufficient quality of this review. This review comprised 118 different studies on the measurement properties of 15 different instruments. The methodological quality of the studies varied widely. Most of the measurement properties reported in the studies were rated sufficient; only 11 (9%) were rated insufficient (Supplemental Digital Content 3, http://links.lww.com/TA/B547), which might indicate publication bias because positive results are more likely to be published.

According to the COSMIN guidelines, PROMs with evidence for sufficient content validity and at least low-quality evidence for sufficient internal consistency can be recommended for use and results obtained with these PROMs can be trusted.[10] None of the instruments provided enough evidence on their measurement properties to be highly recommended for routine use.

All instruments were categorized as level B instruments: PROMs that have the potential to be recommended based on their measurement properties. Further validation studies are needed before one instrument can be highly recommended, although two instruments (the BSHS-B and the different versions of the BBSIP) currently have favorable measurement properties compared with the rest.

The BSHS-B was studied most and possessed the strongest evidence for sufficient quality of most of the measurement properties assessed. Moreover, it seemed the most feasible instrument as is relatively short and freely available in 14 languages. However, the analysis of content validity showed that adding items or item refinement seems necessary before the BSHS-B can be highly recommended. Inconsistency in the results of content validity studies made it difficult to define the *true* gaps in the content of its items. Further validation of the content should therefore be obtained by systematically asking patients and professionals (e.g., clinicians, researchers) about the relevance and comprehensiveness of the items. Also, data

TABLE 3. Characteristics of the HRQL Instruments

Name	No. Items	Adults or Children?	HRQL Construct Definition	Subscales	Number/Types of Response Options	Scoring Algorithm	Feasibility (Completion Time)	Language Version	Administration Costs
Generic instruments									
EQ-5D[16,56,60]	5	Both	HRQL	5 Subscales: mobility, self-care, usual activities, pain/discomfort and anxiety/depression, and a VAS for general health	Three response levels (no problems [1] to extreme problems [3])	3 Different levels per dimension and the VAS score from 0 to 100, or a single (weighted) index utility score (−0.281 [worst] to 1 [best])	Few minutes	176 Translations†	Licensing fees dependent on the type of study/trial/project, funding, source, sample size, and number of requested languages
ITQOL-SF47[57,61]	47	Children 0–5 y of age	Health: a complete physical, mental, and social well-being and not merely the absence of disease	2 Subscales comprising 13 concepts: infant (38 items)—physical abilities, growth and development, bodily pain, temperament and moods, and behavior; parent (9 items)—emotional impact, impact time, and family cohesion	5-Point Likert scale	Sum score from 47 items and transformation to a scale from 0 (worst health) to 100 (best health)	10 min	50 Languages¶	Licensing fee varies according to use
SF-36[16,58,60]	36	Adults	Health, 8 concepts: physical functioning, social and role functioning, mental health, general health, perception, bodily pain, and vitality	8 Subscales: physical functioning, role limitations—physical, bodily pain, general health, vitality, social functioning, role limitations emotional, and mental health	3-Point Likert scale, 5-point Likert scale, 6-point Likert scale	Transformed mean domain scores (0 [the worst] to 100 [the best].	5–10 min	>170 Translations	Licensing fee dependent on the type of organization
Disease-specific instruments									
BBSIP			(adults)[18,19]	66	Adults	Impact of scarring on a person's life experience	7 Subscales: overall impact of scars; impact of itch, pain and other sensations; work and daily activities (mobility and daily activities items); friendship and social interaction; appearance; emotional reactions; and physical symptoms	7-Point Likert scale Dichotomous/numeric	The total score is the summed score of individual items divided by the number of applicable items.
ND		English	Free						

Continued next page

BBSIP (caregivers of children 0–8 y)[18]	58 items	Children	Impact of scarring on a person's life experience	8 Subscales: overall impact of scars; impact of itch, pain, and other sensations; school, play, and daily activities (mobility and daily activities items); friendships and social interaction; appearance; emotional reactions; physical symptoms; and parent and family concerns	5-Point Likert scale Dichotomous/ numeric	The total score is the summed score of individual items divided by the number of applicable items.	ND	English	Free
BBSIP (caregivers of children 8–18 y)[18]	62 items	Children	Impact of scarring on a person's life experience	8 Subscales: overall impact of scars; impact of itch, pain, and other sensations; school, play, and daily activities (mobility and daily activities items); friendships and social interaction; appearance; emotional reactions; physical symptoms; and parent and family concerns	5-Point Likert scale Dichotomous/ numeric	The total score is the summed score of individual items divided by the number of applicable items.	ND	English	Free
BBSIP (children 8–18 y)[18]	58 items	Children	Impact of scarring on a person's life experience	7 Subscales: Overall impact of scars and treatment; impact of itch, pain, and other sensations; daily activities; friendship and social interaction; appearance; emotional reactions; and physical symptoms	5-Point Likert scale Dichotomous/ numeric	The total score is the summed score of individual items divided by the number of applicable items.	ND	English	Free
BOQ <5 y[20,21]	55	Children <5 y	Health status; no definition given	10 Subscales: play, language, fine motor, gross motor, behavior, family, pain/ itching, appearance, satisfaction, and concern/worry	3-Point Likert scale/5-point Likert scale	Domain scores (0 [worst] to 100 [best])	16 min	English, Dutch	ND

Continued next page

TABLE 3. *(Continued)*

Name	No. Items	Adults or Children?	HRQL Construct Definition	Subscales	Number/Types of Response Options	Scoring Algorithm	Feasibility (Completion Time)	Language Version	Administration Costs
BOQ 5–18 y[22–24]	53	Children 5–18 y	Function, physical appearance, and other relevant outcomes	12 Subscales: Upper extremity function, physical function and sports, transfers and mobility, pain, itch, appearance, compliance, satisfaction with current state, emotional health, family disruption, parental concern, and school reentry	Numeric/Likert scales/ dichotomous	Domain scores (0 [worst] to 100 [best])	Parents: mean, 30 min Adolescents: mean, 45 min	English, Swedish, Dutch	ND
BSHS[25]	114	Adults	Dysfunction and distress/ HRQL; no definition given	6 Subscales: physical health, body image, psychological health, sexual health, physical activities, and family/social relationships	ND	ND	ND	English, Spanish	ND
BSHS-A[26–30]	80	Adults	HRQL; no definition given	8 Subscales: mobility and self-care, family/friends, body image, affective, hand function, sexual activity, role activities, and general functioning	Ordinal score (0–4)	Dividing the total score for a domain by the total possible score; resultant scores range from 0.00 (best) to 1.00 (worst).	31 min	English, Chinese, Norwegian, Turkish	Free
BSHS-B[16,31–46,48,49]	40	Adults	HRQL; no definition given	9 Subscales: simple abilities, interpersonal relationships, body image, affect, hand function, sexuality, heat sensitivity, treatment regimens, and work	Ordinal score (0–4)	Mean scores per domain; higher scores reflect a higher perceived functioning.	10–15 min	Chinese, Dutch, English, French, Hindi‡, Taiwanese, Swedish, German, Portuguese, Iranian, Italian, Greek, Hebrew, Polish	Free

BSHS-R[36,51–53]	31	Adults	The impact of burn injury	6 Subscales: simple functional abilities, interpersonal relationships, body image/affect, heat sensitivity, treatment regimens, and work	Ordinal score (1–5)	Mean scores per domain and sum score (range, 31 [worst] to 155 [best])	ND	English, Brazilian Portuguese	Free
DLQI[54,55]*	10	Adults	Quality of life: no definition given	6 Subscales: symptoms and feelings, daily activities, leisure, work and school, personal relationships, and treatment	4-Point Likert scale	Sum score (range, 30 [best] to 0 [worst])	2 min	115 Languages§	Free for clinicians, free for nonacademic research (not funded externally); external funded trial fees dependent on sample size
YABOQ[59]	47	Adults <30 y	Long-term burn recovery	14 Subscales: physical function, fine motor function, pain, itch, social function limited by physical function, perceived appearance, social function limited by appearance, sexual function, emotion, family function, family concern, satisfaction with symptom relief, satisfaction with role, work reintegration, and religion	Likert scales/ numeric/ dichotomous	For each domain, scores are standardized to a mean of 50 (reference group), and a higher score denotes a better health.	10 min	English	Free

*Disease-specific for all patients with a skin disease.
†www.euroqol.org.
‡Number of items and items modified from original version.
§www.cardiff.ac.uk/dermatology/quality-of-life/dermatology-quality-of-life-index-dlqi/.
¶https://www.healthactchq.com/translation/itqol.
ND, not determined. VAS = Visual Analogue Scale.

TABLE 4. Evidence Synthesis (Rating and Quality of the Evidence) on Measurement Properties of HRQL After Burn Injury

	Content Validity			Internal Structure			Construct Validity			
	Relevance	Comprehensiveness	Comprehensibility	Structural Validity	Internal Consistency	Reliability	Hypotheses Testing	Cross-cultural Validity	Responsiveness	Category†
Generic instruments										
EQ-5D	+ Very low	− Very low	+ Very low				+ High			B
SF-36	+ Very low	− Very low	+ Very low				+ High			B
ITQOLSF-47*	+ Very low	− Very low	+ Very low	? Moderate	? High					B
Disease-specific instruments										
BBSIP (adults)	+ High	+ Moderate	+ High		? Low	+ Moderate	+ Moderate		+ High	B
BBSIP (caregivers 0–8 y)*	+ High	+ Moderate	+ High							B
BBSIP (caregivers 8–18 y)*	+ High	+ Moderate	+ High							B
BBSIP (children 8–18 y)*	+ High	+ Moderate	+ High							B
BOQ 0–4*	± Moderate	+ Very low	+ Very low	? Moderate	? Moderate	+ Moderate	− Moderate			B
BOQ 5–18*	+ Moderate	± Low	± Moderate		? High	± Moderate	− Low			B
BSHS					? Very low					B
BSHS-A	+ Very low	+ Very low	± Low		? Low	+ High	+ High			B
BSHS-B	± Moderate	± Low	+ Moderate	+ Moderate	+ High	+ High	+ Moderate	+ Very low		B
BSHS-R	+ Very low	− Very low	+ Very low	? Moderate	? High		+ Moderate			B
DLQI†	± Very low	+ Very low	± Very low	? Moderate	? Low					B
YABOQ	± Very low	+ Very low	+ Very low	? Very low	? Low	? Very low				B

Rating of evidence: Results were qualitatively summarized in an overall conclusion that was either sufficient (+), insufficient (−), inconsistent (±), or indeterminate (?).

Quality of evidence: The quality of the evidence contributing to rating of results was graded according to the modified GRADE approach adapted for this type of review into: high, moderate, low, or very low. + indicates sufficient; –, insufficient; ±, moderate;?, indeterminate.

*Developed for the use in children with burns.

†A, PROMs that have the potential to be recommended as the most suitable PROM for the construct and population of interest (HRQL instruments with evidence for sufficient content validity (any level) and at least low quality for sufficient internal consistency). B, PROMs that may have the potential to be recommended, but further validation studies are needed (HRQL instruments categorized not in A or C). C, PROMs that should not be recommended (HRQL instruments with high-quality evidence for an insufficient measurement property).[10]

GRADE, Grading of Recommendations, Assessment, Development, and Evaluation.

on measurement error of the BSHS-B are lacking and should be investigated to determine if the measurement errors are small enough to obtain important differences in change scores and to determine the importance of (change) scores in an individual.

The four versions of the BBSIP were more recently developed than the other HRQL instruments. Hereby, the developers of these instruments were the only one able to use modern state-of-the-art methods to develop the instruments (Supplemental Digital Content 3, http://links.lww.com/TA/B547).[18,62] This may have contributed to the fact that these were the only instruments that met the high standards for high-quality PROM development and content validity. It is of note to mention that the BBSIP versions were developed to measure HRQL for people at risk of or with burn scars; all questions are asked in relation to scarring, while domains like work and daily activities or emotional reactions may be also influenced by other trauma-related factors and not all patients may (only) suffer from scarring.[14,18,19] The BBSIP versions have to be translated and validated further before they can be highly recommended based on their measurement properties. The outcomes of the questionnaires are the sum score of all items divided by the number of completed items. Future studies should preferably focus on structural validity to determine if this method allows for a meaningful interpretation of scores and to identify whether or not treatment effects are influenced by some scales or items and not others.

All other instruments showed moderate to very low–quality evidence for the aspects of content validity. This was likely the result of poorly performed development studies (no patient involvement or insufficiently sized qualitative interview groups) and a general paucity of studies that analyzed the content of the instruments. Regarding the other measurement properties of the other instruments, it is of note that these are not necessarily inadequate, but they are merely not or scarcely investigated in literature.

The generic instruments EQ-5D and SF-36 are helpful for making a comparison with population norms and other patient groups.[8] Both instruments score moderate to high-quality evidence for sufficient hypotheses testing, which suggests that these instruments can adequately determine differences between groups that differ in burn severity.[16,56,58] However, they seem to miss important content that is relevant for patients after burns; items related to scarring (self-esteem, stigmatization, physical appearance) are missing. Therefore, it cannot be assured that the patient's perspective on HRQL is comprehensively captured in the outcomes.

Burn injury comprises a wide range of patients with mild to severe injury and can affect all domains of physical, psychological, and social functioning.[14,63] Unfortunately, there is no consensus on what items should be included in an instrument to measure HRQL after burn injury.[14] Apart from further studies on the measurement properties of the identified instruments, there is a need to reach consensus on the definition of HRQL for burn patients, as well as on the best instrument to measure HRQL. In a broader perspective, it would be valuable to come to worldwide consensus on a core outcome set (COS) (agreed minimum set of outcomes) that should be measured in burn patients. Former studies found that a variety of different PROMs have been used to assess a range of outcomes, which covered psychological and physical health domains.[64,65] Recently, the development of a COS for clinical trials in burns has been initiated by Young et al.,[66] proposing HRQL as one of the outcomes. The combination of the COSMIN Risk of Bias checklist and criteria for good measurement properties to form a summary of the evidence base for each PROM is crucial to determine which outcome measurement instruments should be included in a COS. Results of current review can therefore guide these recommendations.[67]

Limitations

The COSMIN risk of bias checklist and criteria for good measurement properties are strict, require high standards for reporting, and call for distinct reporting of results. Some of the studies may be of higher quality than rated in this review as a result of incomplete reporting, even though researchers may perform extensive studies. In addition to the quality of measurement instruments, the specific construct as measured by the measurement instrument, feasibility, and interpretability are important aspects when selecting the most suitable measurement instruments. In Table 3, we described the completion time and aspects of feasibility, but the assessment of interpretability (e.g., floor and ceiling effects, minimal important changes) went beyond the scope of this review. Current review focused on instruments that aimed to measure HRQL. As a consequence, other PROMs that may assess only specific aspects of HRQL have not been included. For example, the Life Impact Burn Recovery Evaluation profile that aims to measure social participation includes items on social role and personal relationships, which may also be important to measure HRQL.[68]

CONCLUSIONS

This is the first systematic review to critically appraise the measurement properties of instruments that measure HRQL after burn injury using internationally accepted standards. It showed that the BSHS-B (burn-specific HRQL) and the BBSIP (burn scar HRQL) instruments have the best measurement properties compared with other burn-specific HRQL instruments and that there is only weak evidence on the measurement properties of generic HRQL instruments in burns. This systematic review provides guidance on the HRQL instrument with the best measurement properties. There is a need for consensus on what specific symptoms or aspects are relevant and need to be included in an instrument to comprehensively assess HRQL after burn injury. The overview provided in this review forms important input to reach consensus on a universally used instrument to assess HRQL in burns. In time, this will ultimately provide high-quality evidence to improve patient-centered care.

AUTHORSHIP

C.M.L. conceptualized and designed the study, collected data, analyzed and interpreted data, drafted the initial article, and reviewed and revised the article. I.S. conceptualized and designed the study, collected data, analyzed and interpreted data, and reviewed and revised the article. L.B.M. conceptualized the study, interpreted data, and reviewed and revised the article. M.E.v.B. conceptualized and designed the study, analyzed and interpreted data, and reviewed and revised the article. S.P. and C.H. v.d.V. conceptualized and designed the study, interpreted data, and reviewed and revised the article. All authors approved the final article as submitted and agree to be accountable for all aspects of the work.

ACKNOWLEDGMENTS

We thank Wichor Bramer (Biomedical Information Specialist, Medical Library, Erasmus MC) for performing the database search.

DISCLOSURE

For all authors, no conflicts are declared. This study was funded by the Dutch Burn Foundation (grant numbers 15.101 and 15.102). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article.

REFERENCES

1. Brusselaers N, Hoste EA, Monstrey S, Colpaert KE, De Waele JJ, Vandewoude KH, Blot SI. Outcome and changes over time in survival following severe burns from 1985 to 2004. *Intensive Care Med*. 2005;31(12):1648–1653.
2. Bloemsma GC, Dokter J, Boxma H, Oen IM. Mortality and causes of death in a burn centre. *Burns*. 2008;34(8):1103–1107.
3. Tompkins RG. Survival from burns in the new millennium: 70 years' experience from a single institution. *Ann Surg*. 2015;261(2):263–268.
4. van Baar ME, Essink-Bot ML, Oen IM, Dokter J, Boxma H, van Beeck EF. Functional outcome after burns: a review. *Burns*. 2006;32(1):1–9.
5. Pereira C, Murphy K, Herndon D. Outcome measures in burn care. Is mortality dead? *Burns*. 2004;30(8):761–771.
6. Klein MB, Goverman J, Hayden DL, et al. Benchmarking outcomes in the critically injured burn patient. *Ann Surg*. 2014;259(5):833–841.
7. Stavrou D, Weissman O, Tessone A, Zilinsky I, Holloway S, Boyd J, Haik J. Health related quality of life in burn patients–a review of the literature. *Burns*. 2014;40(5):788–796.

8. 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;13(5):e0197507.
9. Guyatt GH JR, Feeny DH, Patrick DL. Measurements in clinical trials: choosing the right approach. In: Spilker B, ed. 2nd ed. Quality of Life and Pharmacoeconomics in Clinical Trials. Philadelphia, PA: Lippincott-Rave; 1996.
10. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–1157.
11. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res*. 2018;27(5):1171–1179.
12. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, Bouter LM, de Vet HCW, Mokkink LB. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27(5):1159–1170.
13. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
14. Kool MB, Geenen R, Egberts MR, Wanders H, Van Loey NE. Patients' perspectives on quality of life after burn. *Burns*. 2017;43(4):747–756.
15. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. 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; 19(4):539–549.
16. Meirte J, Van Daele U, Maertens K, Moortgat P, Deleus R, Van Loey NE. Convergent and discriminant validity of quality of life measures used in burn populations. *Burns*. 2017;43(1):84–92.
17. Spronk I, Legemate CM, Dokter J, van Loey NEE, van Baar ME, Polinder S. Predictors of health-related quality of life after burn injuries: a systematic review. *Crit Care*. 2018;22(1):160.
18. Tyack Z, Ziviani J, Kimble R, Plaza A, Jones A, Cuttle L, Simons M. 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;41(7):1405–1419.
19. Tyack Z, Kimble R, McPhail S, Plaza A, Simons M. Psychometric properties of the Brisbane Burn Scar Impact Profile in adults with burn scars. *PLoS One*. 2017;12(9):e0184452.
20. Kazis LE, Liang MH, Lee A, 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;23(3):196–207.
21. van Baar ME, Essink-Bot ML, Oen IM, Dokter J, Boxma H, Hinson MI, van Loey NE, Faber AW, van Beeck EF. Reliability and validity of the Health Outcomes Burn Questionnaire for infants and children in the Netherlands. *Burns*. 2006;32(3):357–365.
22. Daltroy LH, Liang MH, Phillips CB, et al. American Burn Association/Shriners Hospitals for Children burn outcomes questionnaire: construction and psychometric properties. *J Burn Care Rehabil*. 2000;21(1 Pt 1):29–39.
23. van Baar ME, Essink-Bot ML, Oen IM, Dokter J, Boxma H, Hinson MI, van Loey NE, Faber AW, van Beeck EF. 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; 27(6):790–802.
24. Sveen J, Huss F, Sjoberg 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;33(6):e286–e294.
25. Blades B, Mellis N, Munster AM. A burn specific health scale. *J Trauma*. 1982;22(10):872–875.
26. Adam M, Leblebici B, Tarim MA, Yildirim S, Bagis S, Akman MN, Haberal M. Validation of a Turkish version of the burn-specific health scale. *J Burn Care Res*. 2009;30(2):288–291. discussion 92-3.
27. Li DW, Liu WQ, Wang HM, Ying S, Cui L, Zhao FF. The Chinese language version of the abbreviated burn specific health scale: a validation study. *Burns*. 2014;40(5):1001–1006.
28. Moi AL, Wentzel-Larsen T, Salemark L, Hanestad B. Validation of a Norwegian version of the Burn Specific Health Scale. *Burns*. 2003;29(6):563–570.
29. Munster AM, Horowitz GL, Tudahl LA. The abbreviated Burn-Specific Health Scale. *J Trauma*. 1987;27(4):425–428.
30. Salvador Sanz JF, Sanchez-Paya J, Rodriguez Marin J. Spanish version of the Burn-Specific Health Scale. *J Trauma*. 1998;45(3):581–587.
31. Finlay V, Phillips M, Wood F, Hendrie D, Allison GT, Edgar D. Enhancing the clinical utility of the burn specific health scale-brief: not just for major burns. *Burns*. 2014;40(2):328–336.
32. Gandolfi S, Auquit-Auckbur I, Panunzi S, Mici E, Grolleau JL, Chaput B. Validation of the French version of the burn specific health scale-brief (BSHS-B) questionnaire. *Burns*. 2016;42(7):1573–1580.
33. Goudarzian AH, Taebei M, Soleimani A, Tahmasbi M, Ahmadi M, Madani MH. Burn Specific Health Scale-Brief (BSHS-B) in pregnant burned women: translation and psychometric evaluation of the Persian version. *Int J of Ped*. 2017;5(7):5391–5400. DOI:10.22038/ijp.2017.24227.2041.
34. Hwang YF, Chen-Sea MJ, Chen CL, Hsieh CS. Validation of a Taiwanese version of the burn-specific health scale-brief. *J Burn Care Res*. 2016; 37(4):e310–e316.
35. Kildal M, Andersson G, Fugl-Meyer AR, Lannerstam K, Gerdin B. Development of a brief version of the Burn Specific Health Scale (BSHS-B). *J Trauma*. 2001;51(4):740–746.
36. Ling-Juan Z, Jie C, Jian L, Xiao-Ying L, Ping F, Zhao-Fan X, Jian-Ling H, Juan H, Feng Z, Tao L. Development of quality of life scale in Chinese burn patients: cross-cultural adaptation process of burn-specific health scale — brief. *Burns*. 2012;38(8):1216–1223.
37. Mulay AM, Ahuja A, Ahuja RB. Modification, cultural adaptation and validation of burn specific health scale-brief (BSHS-B) for Hindi speaking population. *Burns*. 2015;41(7):1543–1549.
38. Muller A, Smits D, Jasper S, Berg L, Claes L, Ipaktchi R, Vogt PM, de Zwaan M. Validation of the German version of the Burn Specific Health Scale-Brief (BSHS-B). *Burns*. 2015;41(6):1333–1339.
39. Piccolo MS, Gragnani A, Daher RP, Scanavino Mde T, de Brito MJ, Ferreira LM. Validation of the Brazilian version of the Burn Specific Health Scale-Brief (BSHS-B-Br). *Burns*. 2015;41(7):1579–1586.
40. Pishnamazi Z, Rejeh N, Heravi-Karimooi M, Vaismoradi M. Validation of the Persian version of the Burn Specific Health Scale-Brief. *Burns*. 2013; 39(1):162–167.
41. Sideli L, Prestifilippo A, Di Benedetto B, Farrauto R, Grassia R, Mule A, Rumeo MV, Di Pasquale A, Conte F, La Barbera D. Quality of life, body image, and psychiatric complications in patients with a burn trauma: preliminary study of the Italian version of the burn specific health scale-brief. *Ann Burns Fire Disasters*. 2010;23(4):171–176.
42. Sideli L, Di Pasquale A, Prestifilippo A, Benigno A, Bartolotta A, Cirrincione CR, La Barbera D. Validation of the Italian version of the burn specific health scale-brief. *Burns*. 2014;40(5):995–1000.
43. Stampolidis N, Castana O, Nikiteas N, Vlasis K, Koupidis SA, Grammatikopoulos IA, Mantzari E, Pallantzas A, Kourakos P, Papadopoulos O. Quality of life in burn patients in Greece. *Ann Burns Fire Disasters*. 2012;25(4):192–195.
44. Stavrou D, Haik J, Wiser I, Winkler E, Liran A, Holloway S, Boyd J, Zilinsky I, Weissman O. Validation of the Hebrew version of the Burn Specific Health Scale-Brief questionnaire. *Burns*. 2015;41(1):188–195.
45. Stolle A, Ripper S, Magdanz J, Honer B, Struckmann V, Kneser U, Harhaus L. Validation of the Ludwigshafen German version of the Burn Specific Health Scale-Brief. *J Burn Care Res*. 2018;39(2):252–260.
46. Szczechowicz J, Lewandowski J, Sikorski J. Polish adaptation and validation of burn specific health scale - brief. *Burns*. 2014;40(5):1013–1018.
47. Van Loey NE, Van de Schoot R, Gerdin B, Faber AW, Sjoberg F, Willebrand M. The Burn Specific Health Scale-Brief: measurement invariant across European countries. *J Trauma Acute Care Surg*. 2013;74(5): 1321–1326.
48. Willebrand M, Kildal M. A simplified domain structure of the burn-specific health scale-brief (BSHS-B): a tool to improve its value in routine clinical work. *J Trauma*. 2008;64(6):1581–1586.
49. Willebrand M, Kildal M. Burn specific health up to 24 months after the burn—a prospective validation of the simplified model of the Burn Specific Health Scale-Brief. *J Trauma*. 2011;71(1):78–84.
50. Blalock SJ, Bunker BJ, Moore JD, Foreman N, Walsh JF. The impact of burn injury: a preliminary investigation. *J Burn Care Rehabil*. 1992;13(4): 487–492.

51. Blalock SJ, Bunker BJ, DeVellis RF. Measuring health status among survivors of burn injury: revisions of the Burn Specific Health Scale. *J Trauma*. 1994;36(4):508–515.
52. Ferreira E, Dantas RA, Rossi LA, Ciol MA. The cultural adaptation and validation of the "Burn Specific Health Scale-Revised" (BSHS-R): version for Brazilian burn victims. *Burns*. 2008;34(7):994–1001.
53. Nicolosi JT, de Carvalho VF, Sabates AL, Paggiaro AO. Assessment of health status of adolescent burn victims undergoing rehabilitation: a cross-sectional field study. *Plast Surg Nurs*. 2013;33(4):185–191.
54. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)—a simple practical measure for routine clinical use. *Clin Exp Dermatol*. 1994;19(3): 210–216.
55. Mazharinia N, Aghaei S, Shayan Z. Dermatology Life Quality Index (DLQI) scores in burn victims after revival. *J Burn Care Res*. 2007; 28(2):312–317.
56. Oster C, Willebrand M, Dyster-Aas J, Kildal M, Ekselius L. Validation of the EQ-5D questionnaire in burn injured adults. *Burns*. 2009;35(5):723–732.
57. Landgraf JM, Vogel I, Oostenbrink R, van Baar ME, Raat H. Parent-reported health outcomes in infants/toddlers: measurement properties and clinical validity of the ITQOL-SF47. *Qual Life Res*. 2013;22(3):635–646.
58. Edgar D, Dawson A, Hankey G, Phillips M, Wood F. Demonstration of the validity of the SF-36 for measurement of the temporal recovery of quality of life outcomes in burns survivors. *Burns*. 2010;36(7):1013–1020.
59. Ryan CM, Schneider JC, Kazis LE, et al. Benchmarks for multidimensional recovery after burn injury in young adults: the development, validation, and testing of the American Burn Association/Shriners Hospitals for Children young adult burn outcome questionnaire. *J Burn Care Res*. 2013;34(3): e121–e142.
60. Chiarotto A, Terwee CB, Kamper SJ, Boers M, Ostelo RW. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain: a systematic review. *J Clin Epidemiol*. 2018;102:23–37.
61. HealthActCHQ (HACHQ) Available at: https://www.healthactchq.com/survey/itqol. Accesed April 24, 2018.
62. Streiner DL, Norman GR, Cariney J. Health measurement scales: A practical guide to their development and use. Oxford Medicine Online. January 2015. Oxford Universitiy Press.
63. Falder S, Browne A, Edgar D, Staples E, Fong J, Rea S, Wood F. Core outcomes for adult burn survivors: a clinical overview. *Burns*. 2009;35(5): 618–641.
64. Griffiths C, Armstrong-James L, White P, Rumsey N, Pleat J, Harcourt D. A systematic review of patient reported outcome measures (PROMs) used in child and adolescent burn research. *Burns*. 2015;41(2):212–224.
65. Griffiths C, Guest E, White P, Gaskin E, Rumsey N, Pleat J, Harcourt D. A systematic review of patient-reported outcome measures used in adult burn research. *J Burn Care Res*. 2017;38(2):e521–e545.
66. Young A, Brookes S, Rumsey N, Blazeby J. Agreement on what to measure in randomised controlled trials in burn care: study protocol for the development of a core outcome set. *BMJ Open*. 2017;7(6):e017267.
67. Boers M, Kirwan JR, Wells G, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. *J Clin Epidemiol*. 2014;67(7): 745–753.
68. Kazis LE, Marino M, Ni P, et al. Development of the life impact burn recovery evaluation (LIBRE) profile: assessing burn survivors' social participation. *Qual Life Res*. 2017;26(10):2851–2866.