A systematic review of patient reported outcome measures (PROMs) used in adult burn research
Griffiths, C* a, Guest, E a, White, P b, Gaskin, E a, Rumsey, N a, Pleat, J c, & Harcourt, D a

a Centre for Appearance Research (CAR), University of the West of England, Bristol, United Kingdom.
b Department of Engineering Design and Mathematics, University of the West of England, Bristol, United Kingdom.
c Department of Plastic Surgery, Southmead Hospital, North Bristol NHS Trust, Bristol, United Kingdom.

* Corresponding author at: Centre for Appearance Research (CAR), Department of Health and Social Sciences, University of the West of England, Frenchay Campus, Coldharbour Lane, Bristol, UK. BS16 1QY. Tel: +44117 32 83947.

Email addresses: Catrin.Griffiths@uwe.ac.uk (C. Griffiths; Research Fellow), emma.gaskin@uwe.ac.uk (E. Gaskin; Research Associate), Paul.White@uwe.ac.uk (P. White; Lecturer in Quantitative Research Methods and Statistics), Ella.Guest@uwe.ac.uk (E. Guest; Research Associate), Nichola.Rumsey@uwe.ac.uk (N. Rumsey; Co-director of CAR & Professor of Appearance Psychology), Jonathon.Pleat@nbt.nhs.uk (J. Pleat, Plastic surgeon), Diana2.Harcourt@uwe.ac.uk (D. Harcourt; Co-director of CAR & Professor of Appearance and Health Psychology).
A systematic review of patient reported outcome measures (PROMs) used in adult burn research
Abstract

Introduction: Patient reported outcome measures (PROMs) are vital for evaluating patient needs and therapeutic progress. This review aimed to identify the PROMs used in adult burn care and establish their quality.

Methods: Computerised bibliographic searches of Psychinfo, Social Sciences Citation Index, Cinahl, Psycharticles, AMED, Medline and HAPI were used to find English-language articles using English-language PROMs from January 2001 to September 2016. Psychometric quality assessment of the PROMs was conducted.

Results: 117 studies achieved the entry criteria and reported using 77 different PROMs (71 generic and 6 burns-specific). Overall, the psychometric quality of the PROMs was low; only 17 (13 generic and 4 burns-specific) had psychometric evidence appropriate to adults with burn injuries completing an English language version of the PROM.

Conclusions: Although this review identified a number of generic and burn-specific PROMs which have some psychometric evidence with adult burn patients, research is still needed to further examine these pre-existing measures and validate them in different languages. This will enable researchers and clinicians to better understand the potential impact of a burn injury on adults, and evaluate the effectiveness of therapeutic interventions.

Keywords: Systematic Review; Patient Reported Outcome Measures; PROM; Adult; Burn
In the UK alone, around 250,000 people suffer a burn injury each year [1]. Previously, such injuries resulted in high mortality rates, however significant advances in burns medical care over recent years mean that an increasing number of people are living with their injuries and may face lifelong physical, psychological and social rehabilitation. Whether sustained in childhood or adulthood, a burn can have a significant impact on the lives of those directly affected and those supporting them. Adults with a burn injury can experience physical symptoms such as pain, sensitivity and itching of the burn scar itself, together with psycho-social difficulties such as trauma symptoms, anxiety, body image distress and difficulties in work, romantic relationships and intimacy [2]. It is therefore important to identify the needs of adults with burn injuries, in order to ensure that they receive the most appropriate support and reduce the likelihood of experiencing lifelong difficulties.

Patient reported outcome measures (PROMs) are increasingly used in research and clinical settings to identify patients’ needs and therapeutic progress. They are standardised and validated health-related questionnaires which patients complete before and after they have received healthcare treatment. PROMs can be generic (assessing general aspects of health) or injury/condition specific (investigating patients’ health in relation to having a burn injury and/or associated treatment). Injury/condition specific PROMs tend to have greater face validity and sensitivity to change. Generic PROMs can be valuable for detecting general health outcomes; however they do not identify outcomes that are specific to a particular patient group. They may therefore lack the degree of sensitivity necessary to identify burn-specific health needs and treatment progress [3]. The United Kingdom (UK) National Health Service (NHS) Next Stage Review [4] highlighted the importance of using PROMs to evaluate healthcare services and to inform commissioning and regulatory decision making. However, the National Burn Care Review [1] identified that PROMs are not routinely collected in burn care and highlighted that the development of new patient reported outcome measures for this population was a priority. The need for rigorous outcome measurement has been reinforced by current UK National Burn Care Standards 2013 [5].
A number of previously published reviews have reported on the PROMs used in adult burn care research, but few included analysis of their psychometric properties [2,6-9]. Although existing PROMS may investigate issues that are relevant to the experiences of adult burn patients, their psychometric qualities cannot be assumed without formal testing with this patient group [10]. A recent review of the psychosocial consequences of burn scars identified a dearth of PROMs that have been validated with burns patients [2], and a systematic review of burn scar rating scales recently assessed the feasibility, reliability, validity, responsiveness and interpretability of these measures [11]. However this review focussed solely on scar assessment which, although important, is only one aspect of burn outcome. Only one scale in Tyack et al’s review [11] (the Patient and Observer Scar Assessment Scale, POSAS) [12] was designed for completion by patients, with the rest being objective measures. Another recent systematic review assessed the psychometric properties of self-reported outcome scales for measuring activities of daily living among burn patients [13], but did not capture scales measuring the impact of burns on wider quality of life.

The current authors previously conducted a systematic review of PROMs used in child and adolescent burn care, together with an assessment of their psychometric quality [14]. Of the 32 different PROMs identified (31 generic, 1 burns-specific), only two generic scales (the Perceived Stigmatization Questionnaire and the Social Comfort Scale) and one burn-specific scale (the Children Burn Outcomes Questionnaire for children aged 5 -18) had psychometric evidence relevant to child and adolescent burn patients. However since this review focussed on child and adolescent burn care, it remains unclear which PROMs are being used with adult burns patients and whether they are psychometrically valid for this population.

The current review was therefore conducted in order to identify and evaluate PROMs currently being used to assess health and well-being (e.g. anxiety, depression, pain, post-traumatic stress disorder, mobility) amongst adult burn patients.
Methods

This systematic review is described using the PRISMA checklist for reporting systematic reviews [15]. The PROSPERO systematic review database published the protocol for this review (http://www.crd.york.ac.uk/prospero) on 8th November 2013. This review used the same method as described in Griffiths et al (2015) [14].

Search strategy

A systematic review technique was used to identify and screen studies that have used PROMs in adult burn care.

Computerised bibliographic searches were conducted using 7 databases (AMED, HAPI, Medline, Cinahl, Psychinfo, Psycharticles, and Social Sciences Citation Index). Journal articles published since the publication of the UK National Burn Care Review (2001) [1] were investigated. The original search criteria therefore identified articles published from January 2001 – March 2013. The search was re-run twice using this same method in May 2015 and September 2016, whilst submitting this article for publication, in order to ensure it is as up-to-date as possible at the time of review. The overall literature search aim was to identify articles related to outcomes and/or measures assessing the effects of treatment in burns care. Articles were then split depending on whether they reported using PROMs with adult or child/adolescent patients. This paper reports those used with adults; the child and adolescent PROMs are reported elsewhere [14]. A lower age limit of 18 years was chosen in line with the definition of adulthood. If an article also included patients who were slightly younger than the specified age range (e.g. aged 17), authors used the average age of participants in the study as the exclusion criteria (i.e. the mean age of participants had to be 18 years and over).
The search terms used were:

- Scale OR score OR instrument OR research instruments OR questionnaire OR inventory OR survey OR measure OR form OR patient reported outcome measure OR pro OR prom

- AND burn OR burns.

Reference follow-up was conducted to identify relevant articles which were not detected in the online bibliographic search.

Articles were screened based on the following criteria:

**Inclusion criteria**

- Articles using a PROM or PROMs
- Articles written in the English language
- Articles using an English language PROM
- Articles using the PROM with adults (aged 18 and over) with a burn injury, or the mean age of participants in the study was at least 18
- Articles using PROMs with published psychometric evidence of measurement reliability, validity or responsiveness
- Articles using PROMs with more than one item (question)
Exclusion criteria

- Articles using instruments that were not patient reported (e.g. parent or clinician reported)
- Articles written in a language other than English
- Articles using PROMs written in a language other than English
- Articles reporting data from participants with a burn injury who are under the age of 18, or have a mean age under 18
- Articles including data on other patient groups in addition to those with burn injuries
- Articles using PROMs without published psychometric evidence of measurement reliability, validity or responsiveness
- Articles using validated PROMs that have been modified and the modified version has not been re-validated
- Articles using a single item PROM
- Articles reporting data from patients who have not had a burn injury
- Articles published before January 2001

Data extraction procedure

The data was independently extracted by three reviewers (CG, EGa and EGu) which included study design, the country in which the study was conducted, participant information (e.g. number and characteristics); PROM type and characteristics (e.g. number of questions). Missing or unpublished information was requested from the study’s corresponding author when necessary. The reviewers discussed any discrepancies in the extracted data and this was resolved through consensus and the double checking of papers by the reviewers.
Quality assessment procedure

International guidelines for the development and validation of health outcome measures were used to assess the quality of the PROMs used in the identified studies. A three-stage development and validation process based on the guidelines and criteria outlined by the Scientific Advisory Committee of the Medical Outcomes Trust for the development and review of health outcome measures [16] is reported by Cano et al (2004) [17]. The Scientific Advisory Committee is an international group of PROM experts that rigorously developed a set of criteria to develop and review the quality of health instrument assessments. This involves a step-by-step process for item generation (developing a conceptual framework and using a literature review, qualitative interviews with patients and expert opinion), item reduction (using expert opinion and psychometric criteria such as factor analysis) and psychometric evaluation (using psychometric criteria). These guidelines are identified as the gold standard for developing and evaluating PROMs. This process is described in detail in Griffiths et al (2015) [14]. These guidelines were used to assess the quality of the evidence of the development and validation data related to each of the reviewed PROMs used with adults with a burn injury.

Two review authors (CG and PW) independently evaluated the quality of the included articles using the criteria detailed above. Any discrepancies between authors in the quality assessments of particular studies were resolved through discussion and the double checking of articles.

Results

Study selection

Figure 1- Flow diagram of systematic selection of articles in the review

********** Insert Figure 1- Flow diagram of systematic selection of articles in the review**********
Figure 1 shows the flow chart of the data screening process. A total of 6250 articles were identified. Twenty three of the final articles reported data from child and adolescent patients so were excluded from the current review (results from these articles are discussed in Griffiths et al (2015) [14]). A total of 67 articles met the inclusion criteria of using one or more PROMs with adults with a burn injury and were included in the review. This flow chart describes the original systematic review search which identified articles published from January 2001 to March 2013.

Two additional searches were conducted to update the adult review to include articles published from March 2013 to September 2016, and then bibliographic reference searches and grey literature searches were conducted. This paper reports the total 117 adult papers reviewed.

Study design

Table 1 details the 77 PROMs used in the 117 articles identified by the systematic search.

The majority of studies were conducted in the USA (n=73), with the remainder being from Australia (n=32), UK (n=8), Canada (n=3) and New Zealand (n=1). The studies used cross sectional (n=47), longitudinal (n=53), intervention evaluation (n=14) and experimental (n=2) designs.

Just under half of the articles (n=56) analysed data from less than 100 participants. The largest sample size was 1842 adult burn patients [18]. Patients completing the PROMs ranged between 13 and 96 years of age, and the mean age of participants ranged from 20.8 – 69.7.

Types of measures

Of the 77 PROMs reviewed, 71 were generic and only 6 were burns-specific. Generic and burns-specific measures assessed a range of outcomes. The most frequently measured overall domains
were psychological health (including anxiety, depression, mood and stress) [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], quality of life [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], physical abilities/functioning [43], [44], [45], [46], [47], [48], [49], pain [50], [51], [52], [53], [54], [55], [56], [57], [58], appearance [59], [60], [61], [62], [63], [64] trauma symptoms/post-traumatic stress disorder [65], [66], [67], [68], [69], personality [70], [71], [72], [73], [74], coping behaviours [75], [76], [77], [78], problem behaviours [79], fatigue [80], post-traumatic growth [81], experienced stigma from others [82], social comfort [82], perceived social support [83,84], community integration [85], perceived family setting [86], itching [87], exercise [88], suicide-related feelings and behaviours [89], alcohol [90], [91] and drug use [92]. See Table 1 for further detail of the domains measured by each scale.

The majority of the generic PROMs were only psychometrically validated with adults in the general population. Only thirteen of the generic PROMs (the Perceived Stigmatization Questionnaire (PSQ) [82], the Social Comfort Questionnaire (SCQ) [82], the Satisfaction with Appearance Scale (SWAP) [62], the Short Form 36-item Medical Outcomes Survey (SF-36) [35], the DASH [48], QuickDash [93], the POSAS [64], the LLFI-10 [49], the Community Integration Questionnaire [85], the Brief Cope [76], the McGill Pain Scale [53], the Brief Fatigue Inventory [80] and the Davidson Trauma Scale [67]) had evidence of validation data with English speaking adults with burn injuries.

Only four of the burn-specific PROMs had been validated in English with adults with a burn: the Burn Specific Health Scale–Abbreviated (BSHS-A) [32], the Burn Specific Health Scale-Brief (BSHS-B) [33], the Young Adults Burns Outcomes Questionnaire (YABOQ) [34] and the Burn Specific Pain Anxiety Scale (BSPAS) [50].
Table 1 PROMs used with adult burn patients

**********Insert Table 1 - PROMs used with adult burn patients***

Table 2 Quality assessment of adult PROMs with English language speaking adult burn patients

**********Table 2 - Quality assessment of adult PROMs**********

Table 2 shows the quality assessment of the 17 PROMs that had available evidence of their development/or validation with English speaking adults with a burn injury. Four of the PROMs were burn-specific: the Burn Specific Health Scale – Abbreviated (BSHS-A) [32], the Burn Specific Health Scale – Brief (BSHS-B) [33], the Young Adults Burns Outcomes Questionnaire (YABOQ) [34] and the Burn Specific Pain Anxiety Scale (BSPAS) [50]. Thirteen were generic PROMs: Perceived Stigmatization Questionnaire (PSQ) [82], the Social Comfort Questionnaire (SCQ) [82], the Satisfaction with Appearance Scale (SWAP) [62], the Short Form 36-item Medical Outcomes Survey (SF-36) [35], the DASH [48], the QuickDash [93], the POSAS [64], the LLFI-10 [49], the Community Integration Questionnaire [85], the Brief Cope [76], the McGill Pain Scale [53], the Brief Fatigue Inventory [80] and the Davidson Trauma Scale [67].

The Abbreviated Burn Specific Health Scale (BSHS-A) [32] is an abbreviated version of the Burn Specific Health Scale [209], a burn-specific quality of life questionnaire. It has 80-items on a five point Likert scale which measures four domains (physical, social, mental and general) and eight subdomains. The items in the original Burn Specific Health Scale [209] were generated using a literature review and expert clinician opinion and a conceptual framework was developed as per Cano et al’s guidelines [17]. Patient interviews were not conducted; however a group of burn patients reviewed a draft of the original scale and suggested additional items. Items were reduced
based on the opinion of clinical experts and burn patients. The BSHS-A was then shortened from 114 to 80 items in Munster and Horowitz’s (1987) [32] study, by the authors identifying duplicates and inconsistencies which were then eliminated. They do not report using item redundancy, endorsement frequencies, missing data, factor analysis or tests of scale assumptions to reduce the number of items.

Munster et al’s (1987) [32] and Munster et al’s (1996) [210] studies provided psychometric evidence for the use of the BSHS-A with adult burn patients, and showed evidence of high levels of internal consistency reliability on all major domains (coefficients: 0.86 = physical health, 0.86 = sexual health, 0.83 = body image and 0.92 = psychological health) [32]. The BSHS-A showed acceptable validity when compared with other pre-burn health and psychological scales, high test-retest reliability (R= 0.89, P< .01). Evidence of validity hypothesis testing indicated that the BSHS-A was sensitive to different outcomes in persons with a history of psychiatric illness and differentiated between outcomes in those employed vs not employed prior to burn injury [210]. Evidence of acceptability (e.g. level of missing data or time taken to complete scale), item total correlations and responsiveness of the scale with adult burn patients not reported.

The Burn Specific Health Scale – Brief (BSHS-B) is an abbreviated version of the BSHS-A and the BSHS-R (a revised version developed by Blalock et al (1994) [211]) which was reduced using factor analysis and validated in Swedish by Kildal et al in 2001 [33]. The scale has 40 items across 9 subscales (simple abilities, heat sensitivity, hand function, treatment regimens, work, body image, affect, interpersonal relationships, and sexuality). A study conducted in Sweden by Willebrand and Kildal (2008) [212] conducted a further second order factor analysis of the BSHS-B and identified three broader domain structures (affect and relations, skin involvement and function) and the work subscale was removed from the analysis because of double loadings. However it must be noted that the patients in this study were 10 years post burn at the time of completing the measure and all had
more severe burns (TBSA => 10%). The psychometric evidence of the English version of BSHS-B is still growing. A recent validation study by Finlay et al (2014) [213] included 927 burn patients who completed the English version of the BSHS-B and found that it had evidence of internal consistency reliability (Cronbach’s alpha= 0.95 for the total score and the subscales ranged from 0.88 to 0.95). Construct validity was identified by measuring the strength of the total scale and subscales with established indicators of severity within three months of injury. TBSA (p<.001), length of stay (p<.001) and surgical treatment (p=.03) significantly predicted the total score scale. The length of stay predicted each subscale, surgical treatment only predicted the treatment regimens and work subscales and TBSA predicted all but one subscales (affect, interpersonal relations and sexuality). A factor analysis identified a final structure of four main domains: skin involvement (heat sensitivity, treatment regimens and body image), physical function (simple abilities and hand function), work, and affect and relations (affect, sexuality and interpersonal relations). Evidence of criterion validity was identified using longitudinal data which showed that the BSHS-B total score scale improved significantly over 24 months (estimated average monthly change (EAMC)=3.48, p <.001). Skin involvement ((EAMC)=0.16, p<.001), affect and relations ((EAMC)=0.49, p<.001), work ((EAMC)=1.63, p<.001) and physical function ((EAMC)=1.09, p<.001) also significantly improved. There was only a small amount of missing data (7%), which indicates that the scale was acceptable for patients to complete. Evidence of responsiveness was shown in Edgar et al (2010) [145] which reported that the BSHS-B significantly identified clinical change between discharge and 1 month post burn, and between 1 and 3 months post burn. However the authors found the BSHS-B showed ceiling effects and a reduced ability to identify statistically significant clinical change from 6 months post burn. There is currently no evidence of item total correlations, test-retest reliability and validity hypothesis testing.

The Young Adult Burn Outcome Questionnaire (YABOQ) [34] measures health outcomes in young adults affected by burns. It has 47 items and 15 domains (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). The items in the YABOQ were originally generated from expert clinician opinion and a literature review (it used items from previously developed scales). Items were based on a conceptual framework. The authors do not report whether exploratory patient interviews (e.g. to identify adult burn patients’ experiences) were conducted. Factor analysis, testing of scale assumptions, and expert opinion informed the item reduction phase.

Ryan et al’s (2013) [34] study which included 153 adult burn patients provided psychometric evidence for the YABOQ. The results showed evidence of internal consistency reliability (Cronbach’s alpha’s ranged from 0.72 to 0.92), test-retest reliability (ranging from 0.29 - 0.94, which showed some change in the health status for some scales) and responsiveness (seven domains in the large range (Cohen’s effect size >0.8), six in the moderate range (>0.5 – 0.8) and two in the small range (=0.2). Factor analysis provided evidence of construct, convergent and divergent validity. The factor analysis identified 15 factors from the 47 items with factors providing non-trivial explanatory power. Items that loaded on to each factor were correlated with each other and were different from the items of other factors. Another study conducted by Ryan et al (2015) [157] provided evidence of validity hypothesis testing; for example, as the total burned surface area (TBSA) increased, nine of the fifteen domain scores reduced. In addition when TBSA increased to 20%, the physical function domain worsened by an effect size of 1.42.

Further information relating to the development and validation of the YABOQ was not reported in the validation paper but was identified through personal communication with the authors of the scale (C. Ryan, personal communication, 16 December 2015); adult burn patients had reviewed draft versions of the scale during the item generation stage. The YABOQ was based on the conceptual frameworks outlined by Wilson and Cleary (1995) [214] and the Medical Outcomes study [35],
however these conceptual frameworks do not describe the full range of domains/subscales in the YABOQ. These frameworks were developed for measuring quality of life for people in the general population and were not developed with or for adult burn patients. Item reduction was based on missing data, item redundancy (using item deletion techniques with Cronbach’s alpha statistics) and endorsement frequencies. In the development study, no items had more than 5% missing data, which indicates that the YABOQ was acceptable to adult burn patients. This is consistent with Cano et al’s (2004) guidelines.

The Burn Specific Pain Anxiety Scale - Abbreviated (BSPAS-A) [51] measures anxiety related to pain experienced during or after medical treatment for a burn. It has 5-items on a 0 - 100 visual analogue scale and is a shortened version of the Burn Specific Pain Anxiety Scale (BSPAS) [50]. The items in the original BSPAS-A were generated from adult burn patient interview data, but the authors do not report whether a literature review was conducted, whether a conceptual framework was developed and/or expert opinion was accessed. In the item reduction phase, retained items were identified using item-total correlations [50] and confirmatory factor analysis [51]. Expert opinion, item redundancy, missing data and tests of scaling assumptions were not mentioned by the authors.

Only one study was identified which provided psychometric evidence of the abbreviated scale with English speaking adult burn patients [131]. Aaron et al’s (2001) [131] study involved 27 adult burn patients and the results provided evidence of criterion validity. The BSPAS-A uniquely predicted procedural pain ($F(1, 24) = 4.63, p=.04$) compared to other general anxiety measures. Additionally, the BSPAS-A was the only significant predictor to add unique explanatory variance (15%) in the degree to which physical health limited function in activities of daily living (measured by the SF-36 physical role functioning subscale), after controlling for TBSA ($F(1,22) = 4.28, p=.05$). However the sample (n=27) size was very low and was underpowered for the statistical analysis that was conducted (Tabachnick & Fidell, 2007) [215]. The BSPAS-A has no evidence of acceptability, internal
consistency reliability, item total correlations, test-retest reliability, validity within the scale, validity hypothesis testing or responsiveness data with English speaking adult burn patients.

The Perceived Stigmatization Questionnaire (PSQ) [82] measures stigmatising behaviours from others that are commonly reported by people with a visible difference, such as burn injuries. It has 21-items on a 5-point Likert scale. There are three subscales: absence of friendly behaviour, confused/staring behaviour, and hostile behaviour. The Social Comfort Questionnaire (SCQ) [82] evaluates the extent to which people feel comfortable in social situations. It has 8-items on a 5-point Likert scale. Item generation for both measures was based on a literature review and expert clinician opinion. The authors do not report whether, in line with Cano et al’s (2004) recommendations for generating PROM items, patient interviews were conducted and/or a conceptual framework was developed. They do report eliciting feedback from adult burn survivors and clinical experts on draft versions of the PROMs, and the questions were amended in line with their feedback. Item reduction was based on psychometric criteria e.g. factor analysis, item redundancy, endorsement frequencies and testing scale assumptions. This is consistent with Cano et al’s (2004) [17] item reduction guidelines.

A validation study involving 361 patients aged over 18 (mean age 44.1) provided psychometric evidence for the use of the PSQ and SCQ with adult burn patients [82]. The study did not report missing data or the time it took adult patients to complete both measures, therefore their acceptability to patients is unclear. Confirmatory factor analysis indicated that the PSQ and SCQ had good internal consistency (Cronbach’s coefficient alpha = .93 and .91, inter-item correlation = .40 (SD = .15) and .57 (SD = .08) and corrected item-total correlations = .60 (SD = .08) and .72 (SD = .06), respectively). Both scales had adequate evidence of construct validity with all factor loadings < .30 and the PSQ cross loadings were < 0.20.
Both scales also demonstrated convergent and discriminant validity (validity with other scales). Each of the PSQ subscales had high negative correlations with the subscales of the Body-Esteem Scale for Adolescents and Adults (BES) [59] (i.e. “appearance evaluation” (correlations ranged from -.48 to -.29, p < .01) and “others’ evaluation of one’s appearance” (correlations ranged from -.48 to -.29, p<.01). The PSQ subscales showed modest negative correlations with the weight satisfaction subscale (correlations ranged from -.30 to -.20, p<.01). The PSQ subscales had a moderate negative relationship with social support (measured by subscales of the Interpersonal Support Evaluation List 12 Question Version (ISEL-12) [84] (all correlations ranged from -.54 to -.27, p<.01). Social Comfort had a moderately high correlation with the ISEL-12 subscales (correlations ranged from -.54 to -.27, p<.01). Both the PSQ subscales and the SCQ had a moderately high correlation with depression (measured by the Short Mood & Feelings Questionnaire Adult Version (SMFQ)) (correlations ranged from .45 to .58, p<.01 for the PSQ and were -.69, p<.01 for the SCQ) [25] and were relatively unrelated to the subscales of the Importance of Appearance Scale (IAS) [60]. However, currently, there is no published evidence of responsiveness, test-retest reliability, item total correlations, or validity hypothesis testing with adult burn patients.

The SWAP [62] measures appearance satisfaction. It has 14-items on a 7 point scale. The items were originally generated from expert clinician opinion and a literature review (using items from a previously developed scale). The authors do not report whether a conceptual framework was developed or whether patient interviews were conducted. However adult burn survivors reviewed draft versions of the scale and amendments were made to the scale based on their feedback. Exploratory factor analysis, item redundancy and the opinions of adult burn survivors and expert clinicians informed the item reduction phase, consistent with Cano et al’s (2004) measurement development guidelines.
Psychometric evidence for the SWAP with 165 adult burn patients is shown in Lawrence et al’s (1998) [62] validation study. The study did not report missing data or the time it took participants to complete the SWAP, therefore it is unclear from this paper whether the measure is acceptable to adult burn patients. Content validity of the SWAP was ascertained by adult burn survivors and clinician experts reviewing the measure to identify the relevance and representativeness of its items.

The SWAP demonstrated evidence of tests of scaling assumptions. There was a high level of internal consistency (.87). The mean inter-item correlation was adequate (.32) and the item-total scale correlation coefficients were relatively high (mean total item-total correlation .53 and the lowest item-total correlation .31). The SWAP therefore had evidence of validity within the scale. The test-retest reliability was identified from a subsample of 84 participants but was relatively low (.59), possibly due to the length of time between tests (2 months).

The SWAP had evidence of convergent validity, showing moderate correlations with measures of body image (Physical Appearance State Trait Anxiety Scale (r = .63, p < .01)), depression (Beck Depression Inventory (r = .51, p < .01)), post-traumatic stress disorder (Davidson Trauma Scale (r = .37, p < .01)), anxiety (Beck Anxiety Inventory (r = .30, p < .01)), quality of life (emotional functioning) (SF-36 Role Emotional (r = -.26, p < .01)), quality of life (social health) (SF-36 social functioning (r = -.40, p < .01)), quality of life (vitality) (SF-36 vitality (r = -.42, p < .01)), and quality of life (mental health) (SF-36 mental health (r = -.43, p <.01)). Evidence of divergent validity showed that the SWAP had no significant relationships with the SF-36 physical functioning (r = -.05, ns) and SF-36 general health (r = -.09, ns) (after controlling for depression). However, as yet, there is no published evidence of responsiveness or validity hypothesis testing with adult burn patients.

The Short Form 36-item Medical Outcomes Survey (SF-36) [35] is a 36-item questionnaire that covers 8 general domains of quality of life (physical functioning, role functioning, bodily pain, general
health, vitality, social functioning, role emotional and mental health). The SF-36 was originally
developed from items of established quality of life measures. The authors do not report whether
expert opinion was elicited or patient interviews were conducted. A conceptual framework was
developed. It is not reported whether the item reduction phase involved expert opinion, item
redundancy, endorsement frequencies, missing data, factor analysis and tests of scaling
assumptions. The SF-36 has been well validated in the general population [35], but until recently
little research has investigated its psychometric properties with a burns population; Edgar et al
(2010) [145] conducted a validation study of the SF-36 with 280 adult burn patients. The results
found that the measure’s subscales and total score scale showed moderate to good correlations
with the total score of the Burn Specific Health Scale-Brief (correlation coefficients ranged from 0.37
to 0.79). This provides evidence of validity comparison with other measures. The SF-36 was explored
in its ability to be sensitive to change over 1 to 24 months post injury, and demonstrated the most
significant change in scores during the period between 1 and 3 months follow-up. However after 6
months post-burn the SF-36 showed a ceiling effect and a reduced ability to measure significant
clinical change. A further paper by Edgar et al (2013) [146] provided evidence of validity hypothesis
testing. The findings showed that age negatively affected recovery in the role emotional, role
physical, physical functioning, role physical and vitality domains. Age had a positive effect on bodily
pain. The total missing scores for the SF-36 ranged from 0.5- 4.4% which indicated that the scale
was acceptable to participants. However currently there is no evidence of item total correlations,
test-retest reliability and validity hypothesis testing in a burns population.

The Patient and Observer Scar Assessment Scale (POSAS) measures the severity of a scar [64]. It
consists of two separate scales, the patient scale and a clinician (observer) scale. The patient scale
(version 2.0 English) has 7 items and measures patients’ evaluation of the scar’s physical qualities.
The first six questions relate to specific characteristics of a scar (e.g. pain, itch, colour, thickness),
whilst the seventh question asks the patient to rate their overall opinion of the scar. The POSAS was
originally developed in Dutch and items were generated from expert opinion and a literature review. No methods of item reduction were reported in the original development paper [64]. The corresponding author for the POSAS confirms wide scale expert opinion was sought in its development, and, although formal quantitative item reduction methods were not used, both item redundancy and item completeness were considered in its development. Additionally, an international study is planned to further test the psychometric properties of the POSAS.

The only study using the POSAS that met the inclusion criteria for this review tested the psychometric properties of the English language version (patient form) with 358 adult burn patients [158]. The authors report using POSAS version 2 (the 7 item scale); however they do not include the seventh item in the structural analysis. The results showed that the POSAS (patient form) had evidence of validity within the scale. Confirmatory factor analysis identified that a two dimensional model was superior to a unidimensional model. The two dimensions were the physical scar (colour, stiffness, thickness and irregularity) and the sensory scar (pain and itch). The POSAS has currently no published evidence of acceptability, internal consistency reliability, item total correlations, test-retest correlations, validity comparison with other measures, validity hypothesis testing and responsiveness with English speaking adult burn patients.

The Lower Limb Functional Index-10 (LLFI-10) measures the functional status of patients with a lower limb condition. It is a 10-item shortened version of the original 25 item LLFI which was developed and validated with patients with musculoskeletal conditions [49]. Participants are asked to respond to questions using the following criteria: a mark when in agreement with the question, a ½ mark when in partial agreement and the question should be left blank if it does not relate to the participant. The original items in the LLFI were generated using a literature review and expert opinion. Items were then reduced using expert opinion, patient (non-burns) opinion and factor analysis. The LLFI was then shortened to the LLFI-10 using expert opinion, item redundancy and factor analysis.
Gittings et al (2016) [150] tested the psychometric properties of the LLFI-10 with 739 adult burn patients and showed evidence of internal consistency reliability (Cronbach’s alpha = 0.85-0.86) and validity within the scale with principal components analysis indicating the LLFI-10 to have a single component structure. There were significant associations with the Burns Specific Health Scale-Brief and Short Form-36 (Spearman’s rho = .56 to .72, p < .001) and observer assessments: the Timed Up and Go test (rho = .41, p < .001) and ankle range of motion (rho = .31 to .35, p < .001). The LLFI showed associations (p < .001) with time since injury (rho = .29), age (rho = .12) and TBSA (rho = .12). Evidence of validity hypothesis testing was identified using a multivariable regression model which showed that changes in the LLFI-10 score were associated with time since burn, age and TBSA. These associations indicate a recovery of function after the burn. The authors also conducted Rasch analysis on the LLFI-10, which demonstrated misfit (Andersen LR p < .001, R1c p < .001). The survey item relating to sleep disturbance was then removed, which resulted in a good fit to the Rasch model (Andersen LR p = .124, R1c p = .219).

Ryland et al (2016) [173] consider the test-retest properties of the LLFI-10, for both the 10-item scale and an additional single item for patients to indicate current percentage of pre-injury performance on a scale from 0 to 100%. Analysis using the intra-class correlation coefficient on a sample of n = 28 indicates excellent test-retest properties for the 10-item short form of the LLFI (ICC = 0.98, 95% CI 0.96 to 0.99) and good test-retest properties for the single item (ICC = 0.88, 95% CI 0.79 to 0.94) with testing at 24 hours and assuredly no later than 48 hours. The LLFI-10 and the single item also showed good internal consistency (Spearman r = -0.83) based on all available data from the two issues of the instrument. Ryland et al (2016) [173] further considered the minimum detectable change (MDC) and estimate an MDC = 1.27. The LLFI-10 is scored in 0.5 increments and consequently a change of >= 1.5 points in the total score indicates a real change in patient’s lower limb function (95% confidence level). For the single pre-injury performance item the MDC is
estimated to be approximately a 30% change, largely demonstrating the increased variability in a single item scale. The LLFI-10 has currently no evidence of acceptability, item total correlations, or validity hypothesis testing with English speaking adult burn patients.

The Disabilities of the Arm, Shoulder and Hand measure (DASH) is a 30-item patient-reported questionnaire which measures upper extremity health status [48]. It was developed for patients with a variety of musculoskeletal diseases and conditions of the upper extremity, rather than specifically for burn patients. It measures domains including function, symptoms (pain, tingling, weakness, stiffness), social activities, and self-image. The items in the DASH were based on a conceptual framework and generated from a literature review of outcome measures for disorders of the arm, shoulder and hand but these were not burn-related. Items were also generated by a group of patients with musculoskeletal difficulties who reviewed the questionnaire items to assess content validity.

Item reduction was based on the opinion of expert clinicians and patients with upper extremity problems (not burn patients) and psychometric data e.g. equidiscriminatory item total correlations, endorsement frequencies and factor analysis [216]. Chapman et al’s (2008) [217] study provided evidence of the validation of the DASH within a study of 211 adult burn patients who were in the military and had received a hand burn injury. Only a subset of study participants (n= 61) completed data on the DASH. These patients took 10 - 15 minutes to complete the questionnaire, suggesting that the DASH was acceptable to patients with a burn. Evidence of validity with other measures showed that DASH correlated well with scores on the Greenleaf EVAL computer-assisted upper extremity evaluation system (which measures severity of impairment) (AMA), with a moderate correlation ($r = 0.50$) between AMA and DASH scores at time 1 and a moderately high correlation ($r = 0.74$) at time 2. Evidence of validity hypothesis testing showed that the DASH was able to discriminate between patients who returned to duty compared to those that did not, with those
who did not return to duty reporting significantly higher DASH scores compared to those that did, at time 1 (54 vs. 33, p=.0002) and time 2 (41 vs. 12, p<.0001). The DASH also had evidence of responsiveness showing a statistically significant change (p< .001) in scores between time points with a large effect size (Cohen's d > 0.8). Chapman et al’s (2008) [217] study provided psychometric evidence for the DASH with adults with a hand burn. However the sample did not have any burns to the arm or shoulder (which the PROM is also intended to measure), so the psychometric properties of the DASH for patients with arm or shoulder burns is unclear. There is currently no published evidence of test-retest reliability, internal consistency reliability, item total correlations and validity within the scale with adult burn patients.

The QuickDASH [93] is an abbreviated version of the DASH and measures upper extremity disability caused by various upper limb disorders. It has 11-items on a 5-point Likert scale. The QuickDASH was developed with patients with upper extremity problems, not burn patients. Although the authors do not report whether patient interviews and expert opinion were used to inform item generation, a conceptual framework was developed, as per Cano et al’s (2004) guidelines.

Item reduction was based on the opinion of expert clinicians and patients with upper extremity problems (not burn patients) [48] and psychometric data e.g. equidiscriminatory item total correlations [216]. In order to develop the QuickDASH, the original 30-item DASH items were reduced to 11 by means of patient feedback using a concept-retention approach [216], and Cano et al’s recommended process does not appear to have been followed.

Wu et al’s (2007) validation study of the QuickDASH involved 85 adult burn patients [46]. Neither missing data nor the time it took patients to complete the QuickDash were reported, so its acceptability to this population is unclear. The QuickDASH correlated well with scores on the Burn-Specific Health Scale (BSHS) at four time points (coefficients ranged from - 0.79 and - 0.89) and the
physical domain of the BSHS (coefficients ranging between – 0.82 and -0.90). Further evidence of validity with other measures was identified in a study by Clifford et al (2013) which showed that the QuickDash correlated significantly with the Grip Strength Dynamometry (an objective device that measures hand-grip) in both the right ($b=0.17$, $p = .002$) and left ($b=0.14$, $p = .002$) hands. Evidence of validity hypothesis testing showed that the QuickDASH was able to identify different patient groups i.e. those with more than 25% TBSA with full thickness burns, inpatients, and those who had undergone surgical interventions who reported higher scores than the remaining patients (Wu et al, 2007). It showed excellent test-retest reliability (coefficients= 0.91, ICC = 0.93) and responsiveness (effect sizes at three follow up time points were large and ranged from 0.6 to 0.8) [46]. There is no published evidence of internal consistency reliability, item total correlations and validity within the scale with adult burn patients.

The Brief Cope (BCOPE) measures coping behaviours [Carver et al, 1989] [76]. It has 28 items and 14 domains (active coping, planning, positive reframing, acceptance, humour, religion, using emotional support, using instrumental support, self-distraction, denial, venting, substance, behavioural disengagement, self-blame). It is a shortened version of the COPE [77]. The items in the original COPE were based on a literature review (it used items from previously developed scales). The authors have not reported the development of a conceptual framework, use of expert opinion and patient interviews.

The COPE was reduced down to the BCOPE based on the feedback from non-student populations that used the scale and factor analysis (Carver et al, 1997) [77]. This is consistent with Cano et al’s (2004) [17] measurement development guidelines, but endorsement frequencies, missing data and test of scaling assumptions during the item reduction phase are not reported.
Amoyal et al’s (2011) study which included 362 adult burn patients provided psychometric evidence for the BCOPE with this population. The results showed evidence of internal consistency reliability (alpha’s ranged from 0.55 to 0.86) and test-retest reliability (ranging from 0.16 to 0.63). Exploratory factor analysis identified a 7-factor solution (active coping, avoidance coping, humour, religion, emotional support, venting and acceptance) that accounted for 51% of the variance. Spearman correlations between the BCOPE, the SF-12, the Davidson Trauma Scale and the Satisfaction with Appearance Scale provided evidence of construct validity. Active coping and avoidance factors were each significantly and positively associated with total DTS scores at 6 months (active .44, avoidance .53) and 12 months (active .44, avoidance .59) after discharge. Avoidance coping was significantly negatively related to the SF-12 at 6 months (-.28), 12 months (-.20) and 24 months (-.26) since discharge. Avoidance coping was significantly positively related to SWAP scores at 12 months (.43) and 24 months (.26) and active coping was significantly positively associated with SWAP scores at 12 months (.36) and 24 months (.41). However currently there is no evidence of acceptability (e.g. amount of missing data or time it took patients to complete the scale), validity hypothesis testing and responsiveness.

The Short Form McGill Pain Questionnaire is a 15-item measure of pain (SF-MPQ) (Melzack et al, 1987) [53]. It is a shortened version of the McGill Pain Questionnaire (MPQ) (Melzack et al, 1983) [52]. The authors report how items in the original MPQ were based on a literature review and expert and non-burn patient opinion, and do not mention patient interviews being conducted or a conceptual framework being developed. The MPQ was reduced down to the SF-MPQ based on non-burn patient feedback (Melzack, 1987) [53]. Expert opinion, endorsement frequencies, missing data, expert opinion, factor analysis and test of scaling assumptions are not reported as being employed during the item reduction phase.
Mason et al’s (2008) [186] study tested the SF-MPQ with 338 adult burn patients. The results showed evidence of internal consistency reliability (alpha= 0.86) and acceptability (e.g. there was less than 2.7% missing data). There was also some limited evidence of construct validity identified using a confirmatory factor analysis of the 15-item scale which identified a 2-factor solution with standardised factor loadings ranging from 0.34 to 0.70 for the sensory factor and from 0.59 to 0.64 for the affective factor. The covariance between factors was 0.77. Future research needs to explore the construct validity of the SF-MPQ with adult burn patients in greater depth. Currently there is no evidence of item total correlations, test-retest reliability, validity comparison with other measures, validity hypothesis testing and responsiveness of the SF-MPQ with adult burn patients.

The Brief Fatigue Inventory (BFI) (Mendoza et al, 1999) [80] has 9 items which are answered on a 0 - 10 point numeric scale. It measures level of fatigue and its potential impact on activity, walking, work, relationships, mood, walking and enjoyment of life. The BFI was originally developed with cancer patients and was based on the Brief Pain Inventory [218]. The authors report that its development involved a literature review, but do not report whether patient interviews, expert opinion and a conceptual framework were used. Item reduction was conducted by way of expert opinion and factor analysis, but this was with a sample of cancer patients and it is not evident whether item redundancy, endorsements frequencies, missing data or tests of scaling assumptions were employed.

Only one study has provided psychometric evidence for the use of the BFI with adult burn patients; Toh et al’s (2014) [135] validation study involved 587 adult burn patients. The Cronbach’s alpha for the BFI at one, 6 and 12 months after burn ranged from 0.96 to 0.99, indicating excellent internal consistency reliability. All item-rest correlations were >0.73 providing further evidence of good internal consistency. Factor analysis showed evidence that all items of the BFI significantly mapped
on to a single domain at one, 3, 6 and 12 months after burn (eigenvalues 6.67, 7.21, 6.83 and 6.50, accounting for 94.9%, 94.8%, 95.3% and 88.2% of variance, respectively).

Validity comparison with other measures was identified at one month post burn with a negative correlation between the BFI and the Burn Specific Health Scale-Brief (p<.001). Responsiveness evidence showed that the BFI was significantly responsive to change from baseline to one, 3, 6 and 12 months post burn (-.58, -.61, -.91, p<.001, respectively). The BFI also had evidence of validity hypothesis testing, identifying the difference between major and minor burns, with major burn fatigue being measured as greater than minor burn during the first 12 months after burn (p<.001). Women also reported higher fatigue levels than men (p<.001). The study did not report levels of missing data; therefore it is not clear how acceptable the BFI was to burn patients. There is currently no available evidence of test-retest reliability.

The Davidson Trauma Scale (DTS) (Davidson et al, 1997) [67] has 17 items that measure the symptoms of post-traumatic stress disorder (PTSD). Item generation was based on a literature review (e.g. using the definition of symptoms for PTSD outlined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV, 1994) [219]). The DTS was based on the general symptoms of PTSD and not related to any burn specific experiences. Contrary to Cano et al’s (2004) [17] guidelines, the authors do not report whether a conceptual framework was developed, expert opinion was elicited or whether patient interviews were conducted. Item reduction was based on factor analysis of data from patients with PTSD from a range of traumas but these did not include burn injuries. This is consistent with Cano et al’s (2004) [17] measurement development guidelines. However the use of endorsement frequencies, missing data and test of scaling assumptions during the item reduction phase are not reported.
Mason et al’s (2013) [220] study provided psychometric evidence for the DTS with 299 adult burn patients. The results showed evidence of internal consistency reliability (Cronbach’s alphas ranged from 0.93 to 0.95). Confirmatory factor analysis provided some limited evidence of construct validity and identified a 4-factor ‘numbing model’ (re-experiencing, active avoidance, numbing and hyper-arousal) which provided a good fit to the data ($\chi^2 (113, N =299) = 156.84$, $p < .001$, TLI = .96, CFI = .97, RMSEA= .04, SRMR=.04). Further research is needed to ascertain more detailed evidence of construct validity of the DTS with adult burn patients. Currently there is no evidence of acceptability, item total correlations, test-retest reliability, validity comparison with other measures, validity hypothesis testing and responsiveness with adults who have had a burn injury.

The Community Integration Questionnaire (CIQ) [85] has 15 items measuring the extent to which an individual feels integrated in their community. It has three subscales: home integration, social integration and productive activities. The scale was originally developed with and for patients with a brain injury. The items in the original CIQ were generated by this patient group and experts in the field. Items were based on a conceptual framework. Item reduction was based on item-subscale correlations and factor analysis. The only study to test the psychometric properties of the CIQ involved 492 adult burn patients [206] and shortened the original 15 items down to 13 items (with two factors: self/family care in the home, and social integration outside the home). Item reduction was based on expert opinion (burn care professionals), factor analysis and tests of scaling assumptions (Configural invariance was maintained when the Exploratory Factor Analysis was stratified by gender, TBSA, and ethnicity. Similarly, Item Response Theory analysis indicated that most items had their location of maximum information within the range of 0 to 1 and only a few items had maximum information at a location 0 or 1).

Gerrard et al’s (2015) [206] study also showed that the CIQ had evidence of validity within the scale (exploratory factor analysis showed loadings of < 0.40 for all items on their respective factors) and
internal consistency reliability (Cronbach’s alpha was 0.79). The CIQ has currently no evidence of acceptability, item total correlations, test-retest reliability, validity with other measures, validity hypothesis testing or responsiveness data with English speaking adult burn patients.

**Discussion**

This systematic review sought to identify the PROMs that have been used in adult burn care research and to establish the quality of the psychometric evidence for their use with this population. The results showed that a variety of different PROMs have been used to assess a range of outcomes. Most of the PROMs eligible for inclusion in this review were generic as opposed to burns-specific, and covered a range of psychological and physical health domains including anxiety, depression, quality of life, physical functioning, post-traumatic stress symptoms, pain, appearance, and coping behaviours. Most of the generic measures reviewed had only been validated with adults derived from the general population, meaning they may not be sufficiently sensitive to identify health outcome changes in an adult burn population. Only 17 PROMs (13 generic and 4 burn-specific) had been subjected to some level of psychometric development and/or testing with adult burn patients and they varied in the extent to which they appeared to have been developed and validated in line with the Medical Outcomes Trust guidelines for PROM development [16].

**PROM development**

The Scientific Advisory Committee of the Medical Outcomes Trust [16] and Cano et al (2004) [17] recommend that patient interview data should be used as a key source when generating PROM items, in order to ensure that the items reflect the experiences of the specific population. Recent guidelines from the Patient Reported Outcomes (PRO) Content Validity Good Research Practices Task Force [221,222] have outlined the importance of including target population input during item generation and for reviewing draft versions of the PROM to assess the content validity of the scale. During the first stage of PROM development, patient interviews are exploratory. These normally
Psychometric evaluation
The level of psychometric validation of the 17 PROMs was generally strong. Those with the most validation data were the BSHS-A, BSHS-B, YABOQ, PSQ, SCQ, SWAP, LLFI, DASH, BCOPE, BFI and the MPS. The SF-36, MPQ, DTS, POSAS, CIQ, and the BSPAS-A were lacking important psychometric evidence with adult burn patients. The majority of PROMs lacked evidence of test-retest data (excluding the BSHS-A, BSHS-B, SWAP, LLFI, QuickDash and BCOPE) and responsiveness (excluding the BSHS-B, YABOQ, SF-36, LLFI DASH, QuickDASH and the BFI). Evidence of responsiveness is particularly imperative for PROMs, since their aim is to identify clinically significant changes in health over time. Therefore further research is needed in order to ascertain their full psychometric properties.

The vast majority of PROMs used with adult burn patients in the current literature were not validated with adult burn patients. However this review identified a growing number of both generic and burn-specific PROMs that have gained or are in the process of gaining psychometric evidence with this population. This therefore provides clinicians and researchers with a range of options for PROMs that they can use with adult burn patients. The BSHS-A, BSHS-B, YABOQ and the BSPAS-A offer burn-specific PROMs which measure aspects of quality of life and pain that are related to a burn injury. These measures have the benefit of asking patients directly about the ways in which their burn injury may have affected them and therefore are likely to be sensitive to the burn-specific needs of this patient group. The BSHS-B and YABOQ both also have evidence of responsiveness, which is vital for outcome measures that are intended to be used with patients multiple times. However the reviewed generic PROMs also had psychometric evidence for their use with adult burn patients. In particular the SF-36 was found to be significantly responsive to changes in adult burn patients’ health up to 6 months post injury [145]. This suggests that a generic PROM such as the SF-36 could be a valuable asset to outcome assessment when identifying general levels of quality of life in adult burn patients.
The advantages and disadvantages of using generic vs condition/injury-specific PROMs, and how they should be used in conjunction (if at all), are much debated in most fields of healthcare [224,225]. This review has identified a number of burn-specific and generic PROMs available to healthcare professionals working with adult burn patients. Using a combination of both generic and burn-specific PROMs (which have been validated with adult burn patients) would allow patients’ burn-specific needs to be identified, while at the same time assessing their general health status, thereby permitting comparisons with normative data from non-affected populations.

This review also identified that shorter validated versions of some scales are available, such as the BSHS (80 items) and the shortened version BSHS-B (40 items). Both versions of the BSHS had psychometric evidence for their use with adult burn patients. The brief version was developed to be easier to use and therefore more clinically relevant than the original 80 item version. Our review identified that the brief version of this scale had been used by more than double the number of studies that had used the original (BSHS), suggesting the shorter scale might be more practical and suitable for clinicians, researchers, and patients alike.

It must be noted that there are limitations in relation to the psychometric approach employed to develop the PROMs reviewed in the current study. The 17 reviewed PROMs were psychometrically tested using classical test theory (CTT). This is the most popular psychometric approach used to develop PROMs [82]. However, developing PROMs using CTT results in total scores that can only be used to compare groups of patients rather than measuring individual patients [10]. This may be challenging for clinicians and researchers working in burn care, since the purpose of PROMs is often to measure individual patient progress.

Similarly, PROMs developed using CTT provide ordinal rather than interval data. This means that measurement invariance (e.g. the tenet that the relationship between the latent variables and
questions needs to remain consistent across patient groups) is not proven [226]. Therefore it is not known whether any identified changes in patients’ scores are the result of genuine differences in the measured latent variable or due to discrepancies in the way different groups of patients interpret the items [14]. Finally, the psychometric properties of the PROM such as responsiveness, reliability or validity can vary depending on the type of sample completing the scale. Therefore it is hard to accurately compare subgroups of the same patient group because the probability distribution of PROM scores can fluctuate between these subgroups [226].

Rasch item analysis offers an alternative method for psychometric scale development. The advantages of Rasch analysis over CTT are described by Cano et al (2011) [226] and have been discussed in relation to burns in a previous paper by the current authors [14]. In summary, Rasch analysis identifies questions that are independent of the sampling distribution of items (and patients) which permits appropriate individual patient and subsample level measurement [10]. Rasch analysis also creates interval level data allowing for measurement invariance to be tested and for valid total scores to be created, which increases the potential for the PROM to identify clinical change. Few studies to date have used Rasch analysis on PROMs in burn care. For example, of the papers included in the current review, only the POSAS [227] and the LLFI-10 [49] were developed using this approach. Researchers developing new PROMs for use in adult burn care should consider using Rasch to ensure they are suitable for measuring the health of both individual patients and subgroups.

Apart from the BSPAS-A and the POSAS (which were originally developed in Dutch), the scales in this review were developed and validated in English speaking countries. It is important to recognise that cultural differences between, and within, countries may mean that the PROMs are not universally relevant. Future research should conduct additional intra-lingual validation if these PROMs are to be used with a population from a different culture from which the scale was originally created [228].
Additionally, the development of suitable PROMs for use in low and middle income countries warrants particular consideration.

Limitations

This review only included English language PROMs, and only those that had been validated in English were subjected to the quality analysis. It may therefore miss relevant research being conducted in languages other than English, yet it is clear that PROM research is being conducted in burn care around the world, in a variety of countries and languages. For example, the Coping with Burns Questionnaire (CBQ) [229] was developed in Swedish and it has not yet been validated in English. Similarly, the EQ-5D is a generic quality of life measure which has only been validated with adult burn patients in Sweden [230]. The Brisbane Burn Scar Impact Profile (BBSIP) [231] is a set of new patient reported and burn scar-related quality of life measures being developed in Australia. At the time this review was conducted, no published psychometric evidence or scoring guidelines for the BBSIP were available, so it did not meet the criteria for inclusion. It is clear from the disclaimer on the scale’s website (http://www.coolburns.com.au/brisbane-burn-scar-impact-profile) that validity testing is in progress, and the team caution against basing decisions solely on the BBSIP at this stage in its development. Given the current amount of activity in this field, it would be advantageous for researchers to work together, in order to progress the development and validation of PROMs for burn care and ensure they are being appropriately translated and validated.

Lastly, this review only identifies PROMs that have been included in published research or unpublished grey literature identified in the online and manual literature searches. It is possible that clinicians working in burn services might be using additional PROMs with their patients but their use has not been formally documented.

Conclusions
Despite the large number and variety of PROMs being used in adult burn care research, only 17 that met the criteria for our comprehensive review have been psychometrically validated with adult burn patients. Therefore, further research is needed in order to investigate the psychometric properties of all PROMs used in adult burn care and to assess their suitability with this population. Additionally, using one universal set of quality criteria to assess and develop PROMs in burn care research would allow for more consistent comparisons of measures used across the field.

Acknowledgements

This study was funded by the Restore Burn and Wound Research charity as part of a program of research funded by Restore Burn and Wound Research, together with The Children’s Burns Research Centre and Dan’s Fund for Burns. The Children’s Burns Research Centre, part of the Burns Collective, is a Scar Free Foundation initiative. The views expressed are those of the authors, and not necessarily those of the funding bodies.

References


27. Spielberger CD. Test anxiety inventory. Wiley Online Library; 1980.


74. Tellegen A. Brief manual for the Multidimensional Personality Questionnaire. Unpublished manuscript, University of Minnesota, Department of Psychology: Minneapolis, MN 1982.


171. Hoskins JL. Life Satisfaction Over the First Five Years Following Burn Injury, Texas A&M University; 2012.


222. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity–establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical


