Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/burns

Clinimetric properties of a translated and culturally adapted Norwegian version of the Patient and Observer Scar Assessment Scale for use in clinical practice and research

Marit Hjellestad [a], *Liv Inger Strand* [b], *Geir Egil Eide* [b,c], *Ragnvald Brekke* [d], *Anne Nesheim* [d], *Bente Elisabeth Bassøe Gjelsvik* [a,*]

[a] *Department of Physiotherapy, Haukeland University Hospital, Bergen, Norway*
[b] *Department of Global Public Health and Primary Care, University of Bergen, Norway*
[c] *Centre for Clinical Research, Haukeland University Hospital, Bergen, Norway*
[d] *National Burns Centre, Haukeland University Hospital, Bergen, Norway*

ARTICLE INFO

Keywords:

Burn scar
Assessment
Inter-tester reliability
Intra-tester reliability
Test-retest reliability

ABSTRACT

Purpose: To translate and culturally adapt the Patient and Observer Scar Assessment Scale, POSAS, to Norwegian and explore its test-retest, intra- and inter-tester reliability.

Methods: POSAS was translated into Norwegian following international guidelines in collaboration with an international translation bureau. Twenty-six adults and 24 children were recruited from a burns outpatient clinic. Three observer-categories: doctor, nurse and physiotherapist, assessed the patients' scars and scored the Observer scale for estimating inter-tester reliability. Photos of the scars were taken and used to score the Observer scale a second time for examining intra-tester reliability. The patients or parents/next of kin rated their scar on the Patient scale at the clinic and after two days at home for examining test-retest reliability. Intraclass correlation (ICC) and Kappa were used for statistical analysis.

Results: A Norwegian version of POSAS (POSAS-NV) was developed. Inter-tester ICC of the Observer parameters varied between 0.203 and 0.728, and for the total sum score, ICC=0.528 (0.280–0.708). Intra-tester ICC of the Observer scale ranged between 0.575 and 0.858. The Patient scale demonstrated high test–retest reliability.

Conclusions: Intra-tester reliability of the Observer scale and test–retest reliability of the Patient scale of POSAS-NV were found satisfactory, but not inter-tester reliability of the Observer scale.

* *Corresponding author at*: Bergen Health Authority, Haukeland University Hospital, Department of Physiotherapy, Post Box 1400, Bergen 5021, Norway.
E-mail addresses: bente.elisabeth.bassoe.gjelsvik@helse-bergen.no, bebgjelsvik@gmail.com (B.E.B. Gjelsvik).

https://doi.org/10.1016/j.burns.2020.10.007

1. Introduction

Evaluation of burn scars benefits from a measurement tool that reflects the quality and appearance of the scar. Draaijers et al. developed the Patient and Observer Scar Assessment Scale (POSAS) for this purpose [1].

The scale comprises an Observer scale (OSAS) to be scored by an expert/health professional, and a Patient scale (PSAS) to be scored by the patient. For OSAS, six characteristics, called parameters, of the scar are evaluated and scored on ordinal scales: Vascularity, Pigmentation, Thickness, Relief, Pliability and Surface area. In addition, category items are available and used to define: Colour, Pigmentation, Thickness, Smoothness, Elasticity, and Distribution, which are considered clinically relevant for comprehensive evaluation. For PSAS, the following six items are scored: Degree of pain, Itching, and comparisons to their normal skin with regard to Colour, Stiffness, Thickness and Irregularity.

In a systematic review of clinimetric properties of burn scar rating scales, POSAS was the only one to receive a high quality rating regarding reliability of the vascularity parameter and the total sum score of OSAS [2]. POSAS was considered the most superior measure as compared to the other scales evaluated. However, its measurement properties with regard to validity, internal consistency and interpretability were found indeterminate, and the need for further testing to justify use of the total sum score was highlighted [2]. A study of POSAS using Rasch analysis, demonstrated reliability and validity showing that POSAS measures a single-construct, defined as scar quality, which justifies summation of scores to a total score [3].

Methodological studies of POSAS have mainly been examined in linear surgical scars, and only few studies have been performed on scars due to burn injuries. In burn injuries, the parameter Surface area may be challenging to determine, as scar quality may differ in different areas of the injured area. Rasch analysis demonstrated that this parameter did not demonstrate an adequate fit to the model [3], therefore more research is needed to determine how best to define Surface area, or remove this parameter from POSAS altogether. The category items of OSAS are not well defined, and we have not found other studies that have examined their reliability. The category items are not clearly enough defined to determine the difference between for example pale–pink–purple–mix (Vascularity) or more–less–mix (Relief). These categories need clearer definitions to be useful in determining scar quality, and as such, need further research. A project to improve the POSAS has recently been initiated and is currently underway [4].

POSAS has been translated and adjusted to several languages, and is available online in Dutch (original version) and English [4]. A reliable and valid scale available in the user's native language is needed to evaluate individual scars and changes in scars over time. In addition, for further development of scar treatment, more research comparing outcomes between different types of treatments as well as between burn units in different parts of the world, is needed. To use the assessment tool in different countries with different languages, POSAS needs to be translated and culturally adapted, and the translated versions need to be examined for measurement properties.

The National Burn Centre of Norway is located at Haukeland University Hospital (HUH) in Bergen. The primary role of the physiotherapists at the Burn Unit is to prevent scar contracture, loss of range of movement and to regain function as soon as possible, as scars have a major effect on the short- and long-term mobility of the patient. Therefore, physiotherapists play a leading role in the evaluation of scars and scar development. This study was undertaken in collaboration between the National Burn Centre of Norway and the Department of Physiotherapy, HUH, and carried out at the multidisciplinary outpatient clinic of the Department for Plastic-, Hand- and Reconstructive surgery (Burns Outpatient Clinic, BOC, HUH).

The purpose of the study was to achieve a translated and culturally adapted Norwegian version of POSAS with good clinimetric properties for use in both clinical practice and in research. The aims of the study were first to translate the POSAS into Norwegian, and then to examine its intra- and inter-tester reliability as well as test–retest reliability.

2. Materials and methods

2.1. *Translation process*

The developers of POSAS [4] were contacted for permission to translate the assessment tool to Norwegian, and approval was given in 2013. POSAS was translated into Norwegian following guidelines of the ISPOR TCA Task Force for Principles of Good Practice for Translation and Cultural Adaptation [5]. Four bi-lingual, experienced burn physiotherapists translated POSAS into separate Norwegian versions. The versions were compared and discussed in relation to the English version, until consensus was achieved.

Cognitive debriefing of the agreed-upon Norwegian version of POSAS was performed to ensure that the translation was comprehensible to the target patient population, i.e. patients with burn injuries, as well as to professionals assessing and treating these patients.

Participants in the debriefing process were four adult patients; two males and two females, as well as the father of a baby patient, in addition to five professionals; two experienced male plastic surgeons and three experienced female wound nurses. They were asked whether they found any words difficult to understand or easy to misunderstand, unclear or offensive.

A bi-lingual colleague and a native speaking English person with no experience with POSAS translated it back into English.

2.2. *Design of reliability study*

A cross-sectional design was used to examine inter-tester and intra-tester reliability of the OSAS, and a longitudinal design to examine test–retest reliability of the PSAS.

2.3. *Scoring criteria*

OSAS and PSAS are scored separately, each including six parameters. Each parameter is scored on a 1–10-point rating scale (10 indicating the worst imaginable scar or sensation).

PSAS and OSAS are summarised individually, each giving a total sum score ranging from six to 60. In addition, there are scores of the observer's and the patient's overall opinion of the scar as compared to the patient's normal skin, scored on a 1–10-point rating scale (10=the worst imaginable scar).

2.4. Patients

Fifty patients were recruited consecutively from BOC by the physiotherapist observer. They were informed verbally and in writing about the study and asked to participate. Based on a written informed consent, they were included in the study. In the case of children, informed consent was derived from a parent or next of kin.

Exclusion criteria were severe cognitive dysfunction, too poor skills in the Norwegian language to understand information and instruction given and to score the Patient scale, as well as mature scars.

The patients or next of kin were asked to choose the scar or part of a scar that they experienced as most distressing (e.g. itchy); an area approximately 3×3cm large, as recommended by the POSAS group [4] and Kabuk et al. (2017) [6]. The patient or parent/next of kin evaluated this area by use of PSAS, first during a regular hospital visit, and again after two days at home. The same parent/next of kin evaluated the child's scar on both occasions. The patients were given a PSAS form and a stamped envelope for returning the filled-in form to the hospital.

2.5. Observer

Three observer-categories (wound nurse, plastic surgeon and physiotherapist) evaluated the same scar, each individually by OSAS, once for inter-tester reliability and twice for intra-tester reliability. Due to their work schedule, altogether six plastic surgeons and three wound nurses were involved in testing of patients. Only the physiotherapist (first author) assessed all patients. The first assessment was performed during a regular outpatient visit, and one photo was taken of the scar chosen by the patient or next of kin. The photo was taken at the same time and under exactly the same condition as when the scar was evaluated during the outpatient visit. The second assessment for intra-tester reliability was planned to be performed after one week, and was based on observation of the photos. The parameter Pliability could not be re-assessed by photo.

2.6. Equipment

A Canon Ixus 132, Compact Camera, was used by a nurse to take the photographs. The camera had no fixed settings, but adjusted automatically to the light condition in the room.

2.7. Statistical analyses

Demographic information about age, gender, ethnicity (skin type is important for scar development [2], percent total body surface area (TBSA), and type/cause and depth of injury (scald-, flame-, contact- or chemical burn, superficial, dermal, deep dermal or full thickness burn), time since injury, initial and present treatment of the relevant scar areas, were collected. For descriptive statistics, we used n (%) or mean values and standard deviation (SD).

Cronbach's alpha was used for analysis of internal consistency. Intraclass correlation coefficient (ICC) with 95% confidence interval (CI) was calculated to examine reliability of each parameter, the total sum and the Overall Opinion. An ICC value $\geq$0.70 is required for sufficient reliability. Values of 0.70–0.89 is considered high reliability, and 0.90–1.00 very high reliability [7]. Standard error of measurement (SEM) and Smallest Detectable Change (SDC) were calculated as well as Limits of agreement, considering both systematic and random error.

Kappa statistic (κ) was used to examine reliability of the individual category items. Kappa values should be at least moderate. Reference values for kappa are: <0.20=poor, 0.21–0.40=weak, 0.41–0.60=moderate, 0.61–0.80=high, and 0.81–1.0=very high [7].

2.8. Ethical considerations

As the study was considered part of regular quality improvement work at the hospital, it did not need approval by the Regional Committee for Medical and Health Research Ethics in Western Norway, and a waiver was received (2016/193). It was approved by the Data Protection Officer at HUH (2016/4364). There was no conflict of interest.

3. Results

3.1. Translation

The translation, back translation and cognitive debriefing followed the guidelines from the ISPOR TCA Task Force for Principles of Good Practice for Translation and Cultural Adaptation [5]. The translation, back translation and cognitive debriefing followed the guidelines from the ISPOR TCA Task Force for Principles of Good Practice for Translation and Cultural Adaptation [5]. The four experienced burn physiotherapists fluent in the English language, started working on the translation of POSAS in 2013. When consensus was reached on a joint Norwegian version of the scale, the developers of POSAS informed that they, by mistake, also had given permission to a Swedish translation bureau (Facit), to translate POSAS into Norwegian. A collaboration between the present Norwegian group and Facit was undertaken, and in 2015 this resulted in an official joint Norwegian version (POSAS-NV).

Neither the patients nor the relatives who participated in the cognitive debriefing of the preliminary POSAS version had any remarks or comments regarding the interpretability of the scales, and no cultural issues arose. The professionals indicated a few problems and suggestions for change:

- The explanatory note on the item/category Thickness is incorrect, as we cannot see down to the subcutical dermal border with the human eye.
- The parameters are not written in the same order on the Observer and Patient form, why?

- There should be a box for the sum score of the six items.

The suggestions for change were, however, not incorporated in POSAS-NV, as it would have caused the scales to differ from the English version. The developers accepted the back translation early in 2016, supporting content validity of the Norwegian version in relation to the English version of POSAS. The inclusion of patients into the reliability study started in August 2016. See Appendix A in Attachment.

3.2. Patients

Of the 27 adults and 24 children/next of kin who were invited to participate (see Flowchart of inclusion process in Fig. 1), only one adult declined participation. Three patients dropped out, i.e. did not return the re-test form. The inclusion period was 2016–2018.

At the time of the burn accident, the mean (SD) age of the children was 4.8 (3.9) years, and for adults 35.2 (18.5) years. Inclusion time was mean (SD) 16.9 (46.2) months after the accident. Demographic information is given in Table 1. Characteristics of the burn injuries, and initial and current treatment are given in Table 2.

3.3. Inter-tester reliability – Observer scale

For the parameters of OSAS, the inter-tester reliability between the three testers was generally poor with ICC-values <0.70 (Table 3). For each of the pair-wise analyses, only one parameter, Pliability, for Observer 1 vs. 2 and Observer 1 vs. 3, demonstrated adequate inter-tester reliability (ICC>0.70). ICC values were low across all pairs of observers for the parameter Pigmentation, and for two pairs of observers for Surface area. For Observer 2 vs. 3, only Overall opinion demonstrated adequate inter-tester reliability. SEM for Overall Opinion was 1.30 and SDC=3.6. Inter-tester reliability between all three observers for the total sum scores of OSAS was not found satisfactory, with ICC (95%CI)=0.528 (0.280–0.708) (data not shown). The category items, except for Colour (moderate reliability) and Thickness (very high reliability), demonstrated generally low inter-tester reliability, see Table 4.

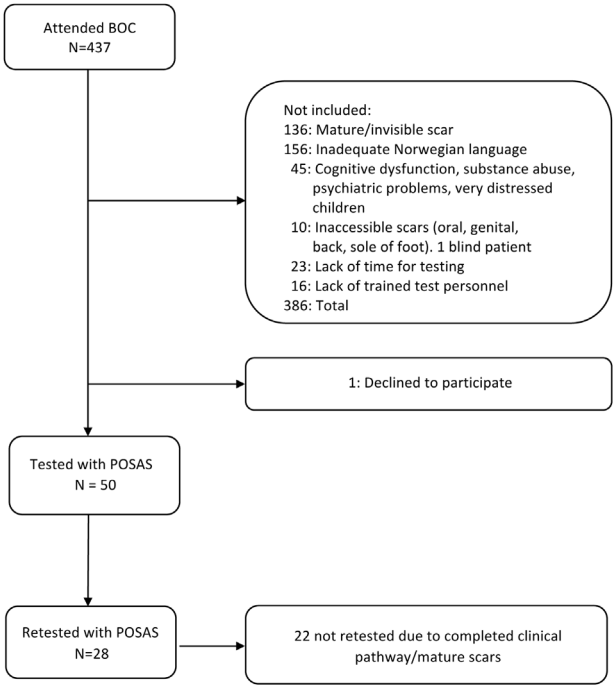

Fig. 1 – Flowchart of the inclusion process, from August 18th 2016 to November 12th 2018.

Table 1 – Characteristics of 50 patients with burn injuries.

Variables	Estimates
Time since accident, months; *mean (SD)*	16.9 (46.2)
Adults, *n (%)*	26 (52)
Children,[a] *n (%)*	24 (48)
Age at accident, years; mean (SD)	
Adults[b]	35.2 (18.5)
Children	4.8 (3.9)
Ethnicity, n (%)	
Caucasian	45 (90)
Asian	5 (10)

Abbreviation: SD=standard deviation.
[a] At most 16 years.
[b] 1 missing.

3.4. Intra-tester reliability – Observer scale

For the Observers, there were mean (SD) 11 (19.05) days, range 4–129 between the first and second assessment, the last by use of photo.

The parameter Pliability could not be assessed by photo and was excluded from the analyses. Intra-tester reliability for OSAS varied between the individual parameters and between the testers (Table 5). Observer 2 demonstrated high intra-tester reliability for all parameters, and intra-tester reliability for Observers 1 and 3 was high for four of six parameters. Vascularity scores demonstrated poor intra-tester reliability for both Observers 1 and 3. Reliability for Overall opinion was high for all Observers, ICCs from 0.797 to 0.827. SEM for Overall opinion was 0.81 and SDC=2.4.

The category items demonstrated varying reliability, from poor (Relief) to high (Thickness) (Table 6).

3.5. Test–retest reliability – Patient scale

For PSAS, there were mean (SD) 3 (2.6) days, range 2–13 between the first and the second assessment. Data from three children were missing. PSAS demonstrated high test–retest reliability for all parameters (ICC2.1≥0.728), with highest intra-tester reliability for Overall opinion (ICC2.1=0.848) (Table 7). SEM for Overall opinion was 0.92 and SDC=2.6.

4. Discussion

The purpose of the study was to achieve a translated Norwegian version of POSAS, POSAS-NV, and to examine its reliability. POSAS was translated according to internationally

Table 2 – Characteristics of the burns, initial and current treatment, in 50 patients with burn injuries.

Variables	*n*	(%)
Location		
Face	5	(10)
Neck	1	(2)
Anterior trunk	8	(16)
Posterior trunk	1	(2)
Anterior thigh	5	(10)
Anterior calf	2	(4)
Posterior calf	3	(6)
Dorsal foot	6	(12)
Outside upper arm	3	(6)
Inside Upper arm	3	(6)
Anterior lower arm	3	(6)
Posterior lower arm	2	(4)
Dorsal hand	4	(8)
Palmar hand	4	(8)
Right	25	(50)
Cause		
Scalding	26	(52)
Flame	18	(36)
Chemical	2	(4)
Contact	2	(4)
High voltage	1	(2)
Necrotizing faciitis	1	(2)
Depth		
Dermal	7	(14)
Deep dermal	14	(28)
Full thickness	29	(58)
Initial treatment		
Tubigrip	21	(42)
Interim garment	17	(34)
Mainat[1]	9	(9)
Silicon	2	(4)
Paper tape	1	(2)
Current treatment		
Tubigrip	8	816)
Interim garment	5	(10)
Mainat[1]	26	(52)
Silicon	7	(14)
Paper tape	5	(10)
Scar contracture	15	(30)
Scar band close to scar	11	(22)
Hypotrophic scar	3	(6)

Abbreviation: [1]Mainat: made to measure-pressure garment.

accepted guidelines for translation and cultural adaptation in collaboration with an international translation bureau, and the Norwegian version was accepted by the POSAS group who developed the mother instrument, supporting content validity of POSAS-NV in relation to the English version. Translation should ensure cross-cultural adaptation, and altogether 10 people were involved in this process. POSAS was developed in the Netherlands, which is a North-European country and culturally similar to Norway. We believe that a good translation that reflected the developers' intention, was achieved, which is supported by the accepted back-translation of the tool to English.

Reliability of POSAS was examined in scars from burn injuries in adults and children. Inter-tester reliability of the Observer scale was rather low between all three pairs of observers. Intra-tester reliability for OSAS was moderate to high for all parameters, while Overall opinion demonstrated high to very high reliability. Intra-tester reliability of the category items of OSAS varied, and ranged between *Kappa*=0.034 (Pigmentation) and 0.811 (Thickness). PSAS demonstrated high to very high test–retest reliability. The results indicate that POSAS is most reliable when the same observer (health personnel and patient/next of kin) reassesses the scar.

We have only found two studies on reliability of POSAS in burn scars [1,6]. Draaijers et al. [1] explored inter-tester reliability and Kabuk et al. [6] both inter- and intra-tester reliability of POSAS. None of these studies reported results on reliability of the individual parameters. Most studies on POSAS have explored its clinimetric properties in linear scars after surgery [8–12], also applying Exploratory Factor Analysis [13] or Rasch analysis [3,14]. In the original study by Draaijers et al. [1], inter-tester reliability was examined by four observers, all medical doctors, who assessed 49 scars from burn injuries in 20 patients (age 15–73). Inter-tester reliability of the OSAS total score was found very high (ICC>0.92). Kabuk et al. [6] translated POSAS to Turkish and inter-tester reliability was examined by a doctor and a nurse, assessing 53 adult patients with burn scars. Very high inter-tester reliability (ICC>0.90) of Overall opinion, was demonstrated. Inter-tester reliability for OSAS total scores in our study was only moderate (ICC=0.53), and moderate reliability between all observers was also demonstrated for Overall opinion. All in all, nine doctors and nurses with varying experience in scar assessment took part as observers, as it was difficult to have the same assessors score all patients because of their work schedules. The physiotherapist was the only observer to assess all scars. However, the photos were reassessed by the same observer who examined the patient at the BOC. Liu et al. [8] assumed that the variable levels of experience among the three observers who assessed the linear facial surgical scars in their study, might have contributed to lower inter-tester reliability on different POSAS items. This may explain the lower inter-tester reliability found in our study as compared to Draaijers et al. [1] and Kabuk et al. [6]. More training together as a group, on scar assessment in general and the POSAS scale in particular, before we started inclusion of patients, might have contributed to a higher inter-tester reliability. However, training of observers was not addressed in either of the two previous studies [1,6].

Measurement error by SEM and SDC must be taken into consideration when judging differences between or changes in scores. We have found no previous studies that calculate SEM and SDC of Overall Opinion. In our study we found intra-tester reliability and test–retest reliability to be satisfactory by ICC values. For OSAS, SDC for intra-tester reliability was 2.4, and SDC for test–retest reliability of PSAS was 2.6. This means that the score of Overall Opinion must change by at least these values to be 95% confident that the change is true, and not only measurement error. This also demonstrates that measurement error is less when the same rater (Observer or next of kin) assesses and reassesses the scar, as the measurement error was larger for inter-tester reliability of OSAS.

Table 3 – Inter-tester reliability of POSAS Observer Scale parameters and Overall opinion in 50 patients with burn injuries.

PSAS	Observer 1: 2			Observer 1: 3			Observer 2: 3		
Parameters	ICC2.1	95% CI	*p*	ICC2.1	95% CI	*p*	ICC2.1	95% CI	*p*
Vascularity	0.415	(0.155, 0.621)	0.001	0.398	(0.060, 0.637)	<0.001	0.399	(0.026, 0.651)	<0.001
Pigmentation	0.236	(−0.190, 0.470)	0.032	0.203	(0.078, 0.466)	0.006	0.293	(0.019, 0.527)	0.005
Thickness	0.644	(0.448, 0.781)	<0.001	0.638	(0.289, 0.811)	<0.001	0.601	(0.137, 0.808)	<0.001
Relief	0.428	(0.171, 0.630)	<0.001	0.426	(0.133, 0.700)	<0.001	0.654	(0.462, 0.787)	<0.001
Pliability	0.711	(0.543, 0.825)	<0.001	0.728	(0.532, 0.843)	<0.001	0.589	(0.376, 0.743)	<0.001
Surface area	0.374	(0.114, 0.588)	0.001	0.290	(0.006, 0.536)	0.003	0.672	(0.477, 0.802)	<0.001
Overall opinion	0.557	(0.332, 0.722)	<0.001	0.587	(0.303, 0.761)	<0.001	0.708	(0.453, 0.842)	<0.001

Abbreviations: POSAS: Patient and Observer Scar Assessment Scale; ICC:2.1: intraclass correlation coefficient a la Streiner and Normann p. 177: absolute agreement (ICC2.1); CI: confidence interval.

Table 4 – Inter-tester reliability of categorical items of the POSAS Observer Scale (OSAS) in 50 patients with burn injuries.

OSAS Category items	Observer 1: 2 Kappa	Observer 1: 3 Kappa	Observer 2: 3 Kappa
Colour	0.437	0.425	0.410
Pigmentation	0.194	0.034	0.261
Thickness	0.811	0.494	0.634
Relief	0.220	0.138	0.035
Pliability	0.390	0.222	0.198
Surface area	0.294	−0.054	0.073

Abbreviation: POSAS: Patient and Observer Scar Assessment Scale.

Previous studies have demonstrated conflicting levels of reliability when using photo to evaluate burn scars and split-thickness skin graft scars by POSAS (Kee 2015, Lindeboom 2009 and Brölmann 2013 in [9]). Shao et al. [9] examined reliability of POSAS when used with scar photos after surgery and concluded that POSAS can be reliable also with use of photo. Only one photo of each scar was taken during the outpatient visit in our study. As no default camera setting was used, the photos were taken under the exact same condition as the live assessment. Norway has only one (national) burn unit, and all patients suffering from burn injuries are therefore treated and followed up at this unit. Patients travel from all parts of the country for their scars to be followed-up. It was not reasonable to ask the patients to stay over one or two nights for reassessment. We therefore think it was justified to use photos for exploring intra-tester reliability although the conditions for assessment were somewhat different.

The time between the first and second assessment (by photo) for intra-tester reliability in our study, was planned to be one week. This time was considered adequate to minimize recall bias, as the assessments were performed during an ordinary busy outpatient clinic. Shao et al. [9] discussed that recall bias might have influenced the results in their study, despite a two-week period between seeing patients in person and grading scars by photo. Our results show that mean 11 days (range 4–129) passed between assessments, and we therefore believe that recall bias might be minimal. We found intra-tester reliability to be acceptable (moderate to very high) between all observers for all parameters, and high to very high for Overall opinion. Kabuk et al. [6] assessed 25 patients again after two weeks for intra-tester reliability, and found that Overall opinion of OSAS demonstrated very high reliability (ICC>0.90). Seen together, these two studies imply that OSAS is most reliable when the same observer re-assesses the same scar.

We found the test–retest reliability of PSAS to be high, which is in line with the results from Kabuk et al. [6]. The patients and parents or next of kin in our study chose the scar and part of the scar that they found most distressing, which is in contrast to Kabuk et al. [6] who chose to evaluate the patients' most visible scar. Draaijers et al. [1] found that itching and thickness of the scar mainly influenced the patients' Overall opinion, which indicates that patients or parents/next of kin have a strong and precise attention towards a self-

Table 5 – Intra-tester reliability for parameters of the POSAS Observer Scale in 50 patients with burn injuries.

OSAS	Observer 1			Observer 2			Observer 3		
Parameters	ICC2.1	95% CI	*p*	ICC2.1	95% CI	*p*	ICC2.1	95% CI	*p*
Vascularity	0.575	(0.168, 0.781)	<0.001	0.779	(0.479, 0.894)	<0.001	0.640	(0.440, 0.779)	<0.001
Pigmentation	0.638	(0.396, 0.789)	<0.001	0.764	(0.611, 0.862)	<0.001	0.858	(0.759, 0.918)	<0.001
Thickness	0.854	(0.667, 0.928)	<0.001	0.758	(0.546, 0.869)	<0.001	0.833	(0.724, 0.902)	<0.001
Relief	0.793	(0.668, 0.877)	<0.001	0.758	(0.609, 0.855)	<0.001	0.728	(0.566, 0.836)	<0.001
Surface area	0.771	(0.630, 0.863)	<0.001	0.778	(0.639, 0.867)	<0.001	0.536	(0.303, 0.708)	<0.001
Overall opinion	0.815	(0.447, 0.921)	<0.000	0.827	(0.691, 0.903)	<0.001	0.797	(0.627, 0.888)	<0.001

Abbreviations: POSAS: Patient and Observer Scar Assessment Scale; ICC2.1: intraclass correlation coefficient a la Streiner og Normann s. 177: absolute agreement (ICC2.1); CI: confidence interval.

Table 6 – Intra-tester reliability for category items of the POSAS Observer Scale (OSAS) in 50 patients with burn injuries.

OSAS Category items	Observer 1 Kappa	Observer 2 Kappa	Observer 3 Kappa
Colour	0.487	0.368	0.400
Pigmentation	0.249	0.400	0.275
Thickness	0.668	0.634	0.540
Relief	0.475	0.159	0.230
Surface	0.516	0.375	0.247

The parameter Pliability was excluded from analysis, as this characteristic cannot be reassessed by photo. *Abbreviation*: POSAS: Patient and Observer Scar Assessment Scale.

Table 7 – Test–retest reliability of the POSAS Patient Scale (PSAS) in 47 patients with burn injuries.

PSAS items	ICC 2.1	95% CI	*p*
Pain	0.798	(0.661, 0.883)	<0.001
Itching	0.872	(0.781, 0.927)	<0.001
Colour	0.746	(0.586, 0.850)	<0.001
Stiffness	0.775	(0.630, 0.868)	<0.001
Thickness	0.728	(0.546, 0.842)	<0.001
Irregularity	0.730	(0.554, 0.842)	<0.001
Overall opinion	0.848	(0.621, 0.929)	<0.001

Abbreviations: POSAS: Patient and Observer Scar Assessment Scale; ICC2.1: intraclass correlation coefficient a la Streiner and Normann p. 177: absolute agreement (ICC2.1); CI: confidence interval.

chosen scar area. Only a mean of three days passed between the first and second assessments in our study. We were concerned that the parent/next of kin would forget to fill in and return the form if we allowed more time between test and retest. We assumed that travelling back home and engaging in home life would bring about enough disruptions for recall bias to be at its minimum, although some recall bias cannot be excluded. We believe that our result is mostly due to the patients' heightened attention towards the distressing experience with the self-chosen scar area.

In burn injuries, the scar may vary in quality in the affected area. Although the patients chose a rather small part (3 × 3 cm) of their total scar area for assessment, we, the observers, found that the parameters were difficult to score as the scar quality differed even within such a small area. This made it challenging to estimate the chosen scar area in relation to the original wound area, which in most cases was both larger and scarred as well. When the chosen scar was a small scar equal in size to the original wound area, there was no category item giving a choice for equal or same size. Van der Wal et al. [3] found that the parameter Surface area did not demonstrate an adequate fit to the Rasch model. Our results show low inter-tester reliability between two pairs of observers for Surface area. More research is therefore needed to determine how best to define Surface area, or alternatively, remove this parameter from POSAS altogether.

The inclusion of 50 patients in this study took more than two years, although as many as 437 patients attended the BOC during this period and only one patient declined to participate. However, 386 patients were not included due to a variety of causes (Fig. 1). We experience that many persons who sustain burn injuries have challenges with understanding or ability to follow-up instructions due to inadequate knowledge of Norwegian, cognitive dysfunction, substance abuse or psychiatric problems. This may explain the long inclusion period. Also, many children and/or their parents/next of kin were distressed and it was deemed unethical to ask them for participation. Other causes for non-inclusion were mature or invisible scars or inaccessible scars for patients' own evaluation, and one patient was blind. In some situations, patients were not included due to lack of trained test-personnel or lack of time for testing. The context of the study was the ordinary day-to-day work in a busy BOC, where the allocated time for each patient is only 20 min. In addition to normal routines, like taking photos of scars, discussing all kinds of scar related problems with the patients, explaining possible treatment/solutions, assessing range of movement and function, dressing wounds if present, measuring for pressure garments as needed, adjusting splints and charting, introducing the POSAS study was often difficult. Therefore, our study possibly includes a selected patient sample, which may be considered a limitation. Nevertheless, the included scars with regard to cause, type and ethnicity seem to be representative of the burns population in our clinic, and are therefore, in our experience, representative of patients suffering from burn injuries.

We included scars from 26 adults and 24 small children. We have not found previous POSAS reliability studies including children with burn injuries. Many children suffer burn injuries. About 30–40% of the patient population of our BOC are children. A reliable assessment scale is therefore important to assess scar quality also for this population, although the use of parents/next of kin are not encouraged by the POSAS group if children are very young. Although the children were not able to evaluate their own scars, collecting the opinion of the parent/next of kin gives valuable information about how they perceive and consistently evaluate the quality of their child's scar. Our study showed that they were able to give rather consistent scores across two assessments.

This study has had a positive impact on the teamwork in our BOC. The attention towards scar assessment and scar management has always been strong, but the need for more collaboration to become more similar in quality assessment of the scars across observers has been highlighted.

5. Conclusion

A Norwegian version of POSAS, POSAS-NV, has been developed. Reliability was found satisfactory only when the same observer or the patient or parent/next of kin assessed and re-assessed the scar area.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Declaration of interest

None.

Conflict of interest statement

The authors declare no conflict of interest.

Acknowledgements

The authors wish to thank the Department for Plastic-, Hand- and Reconstructive surgery, National Burns Centre of Norway, the Burns Outpatient Clinic, and the Department of Physiotherapy, HUH, as well as all patients and parents/next of kin to children who were willing and able to participate in this study. Thank you to the members of the expert panel, physiotherapists Cathrine Sivertsen, Kristine Tillung and Trude Sangolt for valuable input and active participation in the translation process, Keith Jones for back-translating the POSAS-NV, and to librarian Regina Kaufner Lein for help with literature searches.

Appendix A

Attachment
POSAS-NV.

REFERENCES

[1] Draaijers LJ, Tempelman FR, Botman YA, Tuinebreijer WE, Middelkoop E, Kreis RW, et al. The patient and observer scar assessment scale: a reliable and feasible tool for scar evaluation. Plast Reconstruct Surg 2004;113(7)1960–5 discussion 6–7.
[2] Tyack Z, Simons M, Spinks A, Wasiak J. A systematic review of the quality of burn scar rating scales for clinical and research use. Burns 2012;38(1):6–18.
[3] van der Wal MB, Tuinebreijer WE, Bloemen MC, Verhaegen PD, Middelkoop E, van Zuijlen PP. Rasch analysis of the Patient and Observer Scar Assessment Scale (POSAS) in burn scars. Qual Life Res 2012;21(1):13–23.
[4] POSAS Group. POSAS. The Patient and Observer Scar Assessment Scale. The Netherlands: The POSAS Group; 2016 Available from: http://www.posas.org.
[5] Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8 (2):94–104.
[6] Kabuk A, Kazan EE, Aydogan C. Reliability and validity of the Turkish version of patient and observer scar assessment scale in patients with burns. Turk Dermatol Dergisi 2017;11(4):179–83.
[7] Altman DG. 1st ed. Practical statistics for medical research, 1991. London: Chapman & Hall; 1991.
[8] Liu X, Nelemans PJ, Van Winden M, Kelleners-Smeets NW, Mosterd K. Reliability of the Patient and Observer Scar Assessment Scale and a 4-point scale in evaluating linear facial surgical scars. J Eur Acad Dermatol Venereol 2017;31 (2):341–6.
[9] Shao K, Parker JC, Taylor L, Mitra N, Sobanko JF. Reliability of the Patient and Observer Scar Assessment Scale when used with postsurgical scar photographs. Dermatol Surg 2018;29:29.
[10] Truong PT, Lee JC, Soer B, Gaul CA, Olivotto IA. Reliability and validity testing of the Patient and Observer Scar Assessment Scale in evaluating linear scars after breast cancer surgery. Plast Reconstruct Surg 2007;119(2):487–94.
[11] van de Kar AL, Corion LU, Smeulders MJ, Draaijers LJ, van der Horst CM, van Zuijlen PP. Reliable and feasible evaluation of linear scars by the Patient and Observer Scar Assessment Scale. Plast Reconstruct Surg 2005;116(2):514–22.
[12] Vercelli S, Ferriero G, Sartorio F, Cisari C, Bravini E. Clinimetric properties and clinical utility in rehabilitation of postsurgical scar rating scales: a systematic review. Int J Rehabilitat Res 2015;38(4):279–86.
[13] DeJong HM, Phillips M, Edgar DW, Wood FM. Patient opinion of scarring is multidimensional: an investigation of the POSAS with confirmatory factor analysis. Burns 2017;43(1):58–68.
[14] van der Wal MB, Tuinebreijer WE, Lundgren-Nilsson A, Middelkoop E, van Zuijlen PP. Differential item functioning in the Observer Scale of the POSAS for different scar types. Qual Life Res 2014;23(7):2037–45.