OPEN

PRS GLOBAL OPEN

ORIGINAL ARTICLE
Reconstructive

Development of the Observer Scales of the Patient and Observer Scar Assessment Scale: An International Delphi Study

Michelle E. Carrière, MD*†‡§¶
Lidwine B. Mokkink, PhD*‖
Jonathon Pleat, MD**
Zephanie Tyack, PhD††‡‡
Anouk Pijpe, PhD*†§¶
Henrica C.W. de Vet, PhD*‖
Paul P.M. van Zuijlen, MD, PhD‡§¶§§
The POSAS Group

Background: This Delphi study aimed to develop the Observer Scale of the Patient and Observer Scar Assessment Scale (POSAS) 3.0 for measurements of scar quality in patients with all types of scars. The POSAS 3.0 Patient Scale has been developed using an extensive qualitative study, which has been published separately.

Methods: An online, 3-round Delphi study was performed with 124 international scar experts from 23 countries spanning 7 continents. Panelists rated their (dis)agreement with proposals regarding the wording and relevance of scar characteristics on a 5-point scale ("strongly disagree" to "strongly agree"), and provided arguments for their decisions. Consensus was reached if ≥67% of the panelists agreed or strongly agreed to a proposal.

Results: Consensus was reached on the inclusion of 9 items. Two versions of the Observer Scale of the POSAS 3.0 have been developed, 1 generic version including 7 items, and 1 linear scar version, containing the same 7 items and 2 additional scar-specific items.

Conclusions: This is the first study using the Delphi process to develop an instrument for the assessment of scars. Panelists from all over the world reached a consensus on the content of 2 versions of the Observer Scale, the generic version and the linear scar version. By involving the potential end-user community of the new POSAS version in its development, we expect the developed instrument to have high content validity. For a comprehensive scar assessment, the Observer Scale should be used in combination with the Patient Scales of the POSAS 3.0.
(Plast Reconstr Surg Glob Open 2025;13:e6416; doi: 10.1097/GOX.0000000000006416; Published online 28 February 2025.)

INTRODUCTION

Scar quality is a construct referring to the visual, tactile, and sensory characteristics of the scar.[1,2] Scar characteristics measurements are necessary to evaluate the effectiveness of scar treatments and to monitor scar progression. The Patient and Observer Scar Assessment Scale (POSAS) is an instrument that measures scar quality using 2 individual perspectives: 1 designed for the patient and 1 for the health professional. The first version of the POSAS was introduced in 2004 and slightly adapted a year later, resulting in the POSAS 2.0. The latter POSAS version has been widely used over the past decades[1,3–5] and has been translated into more than 13 languages (www.posas.org). However, the POSAS 2.0 could be further improved given

*From the *Department of Epidemiology and Data Science, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; †Association of Dutch Burn Centers, Beverwijk, The Netherlands; ‡Department of Plastic, Reconstructive and Hand Surgery, Burn Center, Red Cross Hospital, Beverwijk, The Netherlands; §Department of Plastic, Reconstructive and Hand Surgery, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; ¶Amsterdam Movement Sciences (AMS) Institute, Amsterdam UMC, Amsterdam, The Netherlands; ‖Amsterdam Public Health Research Institute, Amsterdam, The Netherlands; **Department of Plastic and Reconstructive Surgery, Southmead Hospital, Bristol, United Kingdom; ††Centre for Children's Burns and Trauma Research, Child Health Research Centre, The University of Queensland, Brisbane, QLD, Australia; ‡‡Australian Centre for Health ServicesInnovation (AusHSI), School of Public Health & Social Work, Centre for Healthcare transformation,Queensland University of Technology, Brisbane, QLD, Australia; and §§Pediatric Surgical Centre, Amsterdam UMC, University of Amsterdam, Emma Children's Hospital, Amsterdam, The Netherlands.*

Received for publication August 26, 2024; accepted October 29, 2024.

Members of the PSOS Group are included in an Appendix at the end of the article.

DOI: 10.1097/GOX.0000000000006416

the fact that the Patient Scale was developed by professionals without input from patients living with scars. In addition, the Observer Scale was developed in the Netherlands and could be criticized for not reflecting a more global perspective. Therefore, we initiated the development of a third POSAS version, POSAS 3.0, in which we aimed to improve the content validity and generalizability by including many international experts encompassing patients and professionals in the field.[6]

A comprehensive scar assessment should include the evaluation of the observer's opinion and the patient's opinion. We promote using the Patient Scale complementary to the Observer Scale in clinical practice. The development of the Patient Scale using a qualitative study has been published separately.[2,7] This study aimed to develop the Observer Scale of the POSAS 3.0, suitable for measuring scar quality in patients with various types of scars, including those resulting from burns, trauma, surgery, infection (eg, necrotizing fasciitis), and acne, along the entire pathway to full maturation. As sensory characteristics can only be accurately reported by the patient, the Observer Scale contains items concerning visual and tactile characteristics of the scar. An international Delphi study was performed to obtain the opinion of many scar experts on the relevance, comprehensiveness, and comprehensibility of the content of the Observer Scale in a systematic manner. The Delphi method is a research method to obtain opinions, structure discussion, and establish a consensus from the spectrum of opinions of a group of experts, by using a series of surveys.[8–11]

Takeaways

Question: Which scar quality characteristics do scar experts value most in various types of scars?

Findings: The international Delphi process involved 3 rounds of online surveys with scar experts, resulting in two versions of the Patient and Observer Scar Assessment Scale 3.0 Observer Scale—a generic version for all scars and a linear scar version specifically for linear scars. The study's key findings include the consensus on 13 relevant scar characteristics for inclusion, a 5-point verbal scale for rating, and the exclusion of certain items such as keloid size due to measurement challenges.

Meaning: Panelists from all over the world reached a consensus on the content of 2 versions of the Observer Scale: the generic version and the linear scar version.

METHODS

Preparation of the Delphi Study

A systematic review was undertaken previously on the content of outcome measurement instruments that measure scar quality.[1] Outcomes from that review (scar quality characteristics) were used to formulate proposals included in the Delphi questionnaires. We aimed to develop a concise scale, which only contained items that could be assessed by looking at and touching the scar, without the need for measurement devices.

Steering Committee

The steering committee consisted of international experts with experience in instrument development and Delphi studies, clinicians working with scars, and qualitative researchers (ie, all co-authors of this publication). The steering committee was responsible for panelist selection, the content of questionnaires and feedback rounds, the analysis of responses, and reporting of the study. Members of the steering committee did not participate as panelists.

Panelists

Scar experts were defined as professionals working with scars in the field of burn care, plastic surgery, general surgery, and dermatology, or researchers reporting on scar evaluation. Panelists were recruited using a multifaceted approach. First, researchers who authored 3 or more publications on scar quality assessment were identified and enlisted.[1] Second, we recruited experts at sequential burn, scar, and plastic surgery conferences over a year (2018–2019). Third, members of the steering committee suggested experts from their own professional network. Fourth, panelists identified through 1 of the 3 previously mentioned recruitment methods were asked to suggest additional experts from their professional network. Panelists who agreed to participate were requested to participate in all rounds of the Delphi study. Panelists who missed 1 round were not invited to the next one. Panelists remained anonymous to all people involved in the study, except the lead author (M.E.C.) for administrative purposes. The responses and data were handled confidentially and securely. All participants who completed all 3 Delphi questionnaires were offered the opportunity to be acknowledged in this publication.

Design of the Study

This international Delphi study consisted of 3 rounds of online surveys sent between July 2019 and March 2021. Panelists were asked to rate their agreement with proposals and encouraged to provide arguments for their responses. Panelists rated their (dis)agreement with proposals on a 5-point rating scale, ranging from strongly agree to strongly disagree. In rounds 2 and 3, panelists received the results of the previous round in a feedback report, including all given arguments. This enabled panelists to revise their answers based on the arguments of other panelists. Based on previous experiences, the threshold of consensus was reached when ≥67% of the panelists agreed with a proposal.[12] If no consensus was reached upon a proposal, panelists were asked to consider the same proposal, or a slightly adapted proposal, again in

Disclosure statements are at the end of this article, following the correspondence information.

Related Digital Media are available in the full-text version of the article on www.PRSGlobalOpen.com.

the next round, accompanied by other panelists' ratings and arguments. When no consensus could be reached upon a proposal after 3 rounds, the steering committee made the final decision. Survalyzer software (Survalyzer AG, Utrecht, the Netherlands) was used for survey drafting and distribution.

Content of Rounds

This Delphi study consisted of 3 rounds (Fig. 1; Table 1). Panelists received an invitation e-mail with a link to the questionnaire. The first questionnaire covered the relevance and preferred wording to describe scar characteristics. In addition, we enquired about the panelists' professional backgrounds. Based on the results of a systematic review of the literature,[1] we proposed alternative terms (with a range of 2–7 different options) for each scar characteristic and asked panelists which one they found most appropriate. For an example, see Figure 2. After panelists selected the preferred term, they were asked whether it should be included in a scale that measures scar quality from a professional perspective. Next, we enquired whether important scar characteristics were missing. In the second round, proposals regarding the wording, descriptive categories, answering options, and instruction of potential items were provided based on the arguments given in the previous round. Panelists were also asked to rate the relevance of each identified characteristic for different types of scars (scars caused by burns, infection, acne, trauma, surgery, and keloids). In the third round, new proposals were made based on the results of the previous round. In addition, a preliminary version of the Observer Scale was included.

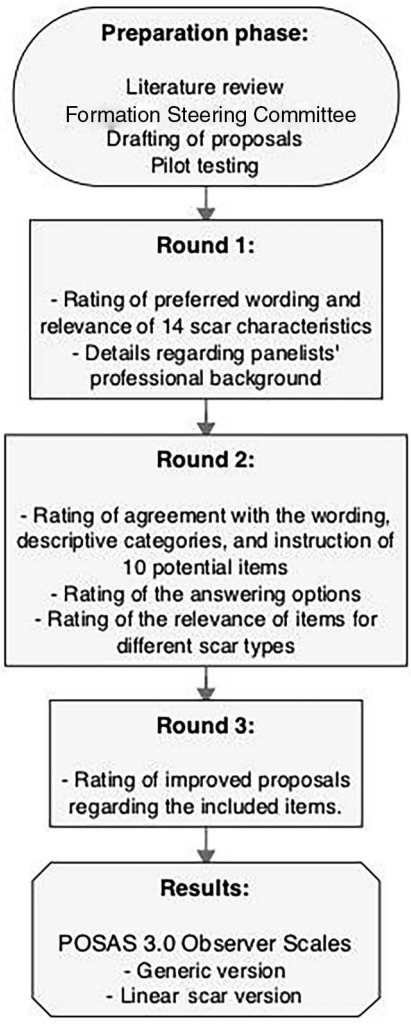

Fig. 1. Flowchart of Delphi procedure.

Table 1. Demographics of Panelists

	Value, *N*	%
Panelists round		
1	124	100
2	68	55
3	55	44
Country		
Australia	29	23
The Netherlands	20	16
The United Kingdom	19	15
The United States	10	8
Canada	7	6
Italy	7	6
Other*	32	26
Focus area		
Burn care	65	52
Plastic surgery	57	46
Rehabilitation	27	22
General surgery	13	11
Wound repair/regeneration	12	10
Dermatology	3	2
Other	6	5
Expertise in scar types		
Burn	102	82
Traumatic	67	54
Surgical	67	54
Keloids	44	36
Necrotizing fasciitis	38	31
Acne	2	2
Other	7	6
Scar scales, frequency of use		
Always, standard of care	31	25
Regularly	43	35
Sometimes	30	24
A few times	12	10
Never before	4	3
Other	4	3

*The category "other" consists of 16 countries in Europe (7), Asia (6), Africa (1), South America (1), and Oceania (1).

RESULTS

Panel Members

We invited 268 experts to participate, of whom 124 (46%) agreed to participate in the first round. Of the 124 panelists who participated in the first round, 68 (55%) also agreed to participate in the second round, and 55 (44%) in the third round. Panelists' details regarding

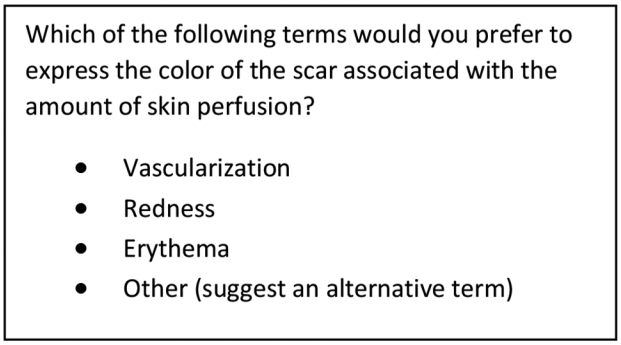

Fig. 2. Example of a question included in the first Delphi questionnaire in which panelists were asked to choose the term which best fit the specific scar characteristic described.

demographics, background, and experience are presented in Table 1.

Relevance of the Scar Characteristics

In round 1, panelists considered 13 of 14 proposed scar characteristics relevant enough to be included in a scale to measure scar quality from a clinician's perspective (Table 2; Fig. 3). Table 2 shows the percentages of agreement for each scar characteristic, as well as the preferred terms. The item *shine* was excluded because panelists did not consider it relevant enough for the construct scar quality (29% agreed on inclusion). The item *marks resulting from skin closure* also failed to meet the threshold for consensus (56% agreed on inclusion) and was reassessed for relevance in round 2. Panelists pointed out that some of the 13 relevant characteristics contained overlap in their meaning. Panelists also pointed out that some characteristics were part of the same concept but referred to different extremes within a spectrum. For example, elevation and depression are both terms that describe the level of scar in comparison to unaffected skin but imply 2 opposite directions. In the first round, these characteristics were discussed separately to enable panelists to judge their relevance individually. In the following round, we proposed to merge such characteristics within 1 item by adding descriptive categories. For example, the item *surface level* includes the following descriptive categories: elevated and depressed.

In round 2, panelists reached a consensus on the wording, descriptive categories, answering options, and instruction of 7 of 11 items (Table 2; Fig. 3). For the 4 remaining characteristics, no consensus could be reached yet (ie, the difference in texture, pliability/firmness, tension and/or movement restriction, and fragility). In this round, panelists also rated the relevance of proposed items for different types of scars. (**See table, Supplemental Digital Content 1,** which displays the panelists' rating of the relevance of proposed scar characteristics for different types of scars, **http://links.lww.com/PRSGO/D759**.) These data showed that most characteristics were relevant for all scar types, whereas some characteristics were only applicable to specific scar types (eg, surgical marks for linear scars). Consequently, the steering committee decided to develop different versions of the Observer Scale of the POSAS 3.0 for different scar types. At this time, the aim was to develop 3 versions of the POSAS 3.0: a version for scars caused by burns, infection, and accidents; a version specific for linear scars; and a keloid-specific version.

In the third Delphi round, panelists agreed to exclude the item *fragility* because it is a characteristic that can only accurately be reported by the patient and is hard to identify or quantify by a professional. In addition, small modifications were proposed to improve the items that were already approved in the previous round. For the majority of the discussed items, we received a high consensus rate (>80%). On 1 remaining item, that is, marks resulting from surgical wound closure, we did not reach a consensus for exclusion (66% agreed on exclusion). The steering committee decided to include the item *marks resulting from surgical wound closure* because these marks are often a concern for patients and an indication for surgical revision.

On another item, *keloid size in comparison to the size of the original lesion*, a consensus was reached on the wording and instruction (73% and 71% agreed on inclusion, respectively). However, in the open-text options, many panelists expressed their doubts about whether this item can accurately measure keloid size. As a result, the decision regarding the inclusion of these 2 items was determined by the steering committee. It decided to exclude the item *keloid size*. The main reason for exclusion was the imprecision in assessing the original lesion and the current keloid size. Observers would be guessing the size of the original lesion to be able to answer the item, which greatly decreases its accuracy and practical utility. Keloid size can be more accurately measured using measurement devices, the use of which was contrary to the aim of this study (see Methods section). As a consequence of excluding the item *keloid size*, it was decided to refrain from developing a keloid-specific POSAS version. Therefore, there were 2 final versions of the POSAS 3.0 Observer Scale, the generic version (Supplemental Digital Content 2) and the linear scar version (Supplemental Digital Content 3). (**See figure, Supplemental Digital Content 2,** which displays the generic version of the POSAS 3.0 Observer Scale, applicable to all types of scars, except linear scars, **http://links.lww.com/PRSGO/D760**.) (**See figure, Supplemental Digital Content 3,** which displays the linear scar version of the POSAS 3.0 Observer Scale, applicable to linear scars, **http://links.lww.com/PRSGO/D761**.)

Wording of the Included Items of the POSAS 3.0 Observer Scale

In Supplemental Digital Content 4, the details regarding the wording of each included item in a consecutive order are discussed. (**See table, Supplemental Digital content 4,** which displays details of the discussions regarding the wording of each included item of the Observer Scale of the POSAS 3.0, **http://links.lww.com/PRSGO/D762**.)

Answering Options

In the first round, many panelists mentioned in the open-text comments that items should be rated by comparing the scarred skin to healthy, surrounding skin. In

Table 2. The Percentages of Consensus Among Panelists Regarding the Proposed Scar Characteristics Throughout Subsequent Delphi Rounds

Characteristic	Round 1	Round 2	Round 3	Final Decision
1. The color of the scar associated with the amount of skin perfusion	99% Agreed on inclusion Preferred term: vascularization (43%)	Item: 92% agreement Cat: 81 % agreement Instr: 80% agreement	Item: 97% agreement Cat: 95 % agreement Instr: 93% agreement	N/A
2. The color of the scar associated with factors other than the amount of skin perfusion (eg, melanin content)	89% Agreed on inclusion Preferred item: pigmentation (87%)	Item: 95% agreement Cat: 94% agreement Instr: 94% agreement	Item: 95% agreement Cat: 99% agreement Instr: 97% agreement	N/A
3.The increased level of the scar compared with the surrounding skin	99% Agreed on inclusion Preferred term: thickness 28%	Combined: Item: 93% agreement Cat: 85% agreement Instr: 76% agreement	Item: 98% agreement Cat: 96% agreement Instr: 88% agreement	N/A
4. The decreased level of the scar compared with surrounding skin	81% Agreed on inclusion Preferred term: depression 45%			
5. The uneven surface of the scar	88% Agreed on inclusion Preferred term: irregularity (44%)	The majority of panelists preferred to combine both concepts into 1 item (54%)	Item: 81% agreement Cat: 73% agreement Instr: 73% agreement	N/A
6. The even surface of the scar	70% Agreed on inclusion Preferred term: smoothness (54%)			
7. Biomechanical properties of the scar	96% Agreed on inclusion Preferred term: pliability (58%)	68% of the panelists agreed with the wording of the instruction 41% favored pliability 36% favored firmness	Item: 88% agreement Instr: 80% agreement	N/A
8. How the scar relates to underlying tissues	84% Agreed on inclusion Preferred term: adherence (84%)	Item: 92% agreement Instr: 84% agreement	Instr: 89% agreement	N/A
9. How the scar relates to surrounding tissues and a reduction in scar size that can occur over time	70% agreed on inclusion preferred term: contraction	44% preferred to include both tension and movement restrictions as separate items 41% preferred to combine both concepts in 1 item	Item: 85% agreement Instr: 83% agreement	N/A
10. A reduction in scar size that can occur over time	79% Agreed on inclusion Preferred term: reduction in surface area			
11. An increase in scar size that can occur over time	78% Agreed on inclusion Preferred term: enlargement in surface area (46%)	Item: 74% agreement Cat: 71% agreement Instr: 67% agreement	Keloid size* Item: 73% agreement Instr: 71% agreement	71% agreed on exclusion
			Widening scar edges* Item: 91% agreement Instr: 85% agreement	N/A
12. The aspect of the surface of the scar (eg, shine, gloss, reflection)	29% Agreed on inclusion This item was excluded			
13. The marks resulting from wound closure*	58% Agreed on inclusion Preferred term: suture marks (71%)	Item: 78% agreement Cat: 56% agreement Instr: 81% agreement	66% agreed to exclude this item	71% agreed on inclusion
14. The diminished quality of the skin, causing ulcerations, blisters, and open wounds in the scar	80% Agreed on inclusion Preferred term: fragility (43%)	62% agreed on exclusion, 12% was neutral	81% agreed on exclusion This item was excluded	
The observers' overall opinion of the scar	67% Agreed on inclusion	Item: 83% agreement Cat: 74% agreement Instr: 71% agreement	Instr: 84% agreement	N/A

*The marked characteristics are scar-specific characteristics, which are included in scar-specific versions of the POSAS 3.0.
Cat, categories; Instr, instructions.

the second round, it was proposed that the answering options reflect the degree to which the scarred skin is different from the unaffected skin. A 7-point verbal scale was suggested, ranging from similar (minimum score) to extremely different (maximum score). Panelists did not agree with this proposal. The main reason for the panelists' rejection of the proposal was the inadequate distinction between the scaling statements "a bit different" and "slightly different," as well as between, "significantly different" and "very different." Most panelists preferred to use a 5-point scale instead of a 7-point scale. Others suggested that a 4-point scale would suffice or mentioned that they preferred to use the familiar 10-point scale that is intrinsic to the POSAS 2.0 scale. In the third round, we proposed 3 different options: (1) a verbal 5-point scale with all points labeled, (2) a 7-point scale with only the endpoints labeled, or (3) the 10-point scale as included in the POSAS 2.0. After this round, 77% of the panelists agreed on the use of a 5-point verbal scale to rate all included items (ie, "not," "minimally," "moderately," "severely," and "extremely different" in comparison to unaffected skin.

Final Versions of the Observer Scale of the POSAS 3.0

Two versions of the POSAS 3.0 Observer Scale were developed. The generic version is applicable to all scars,

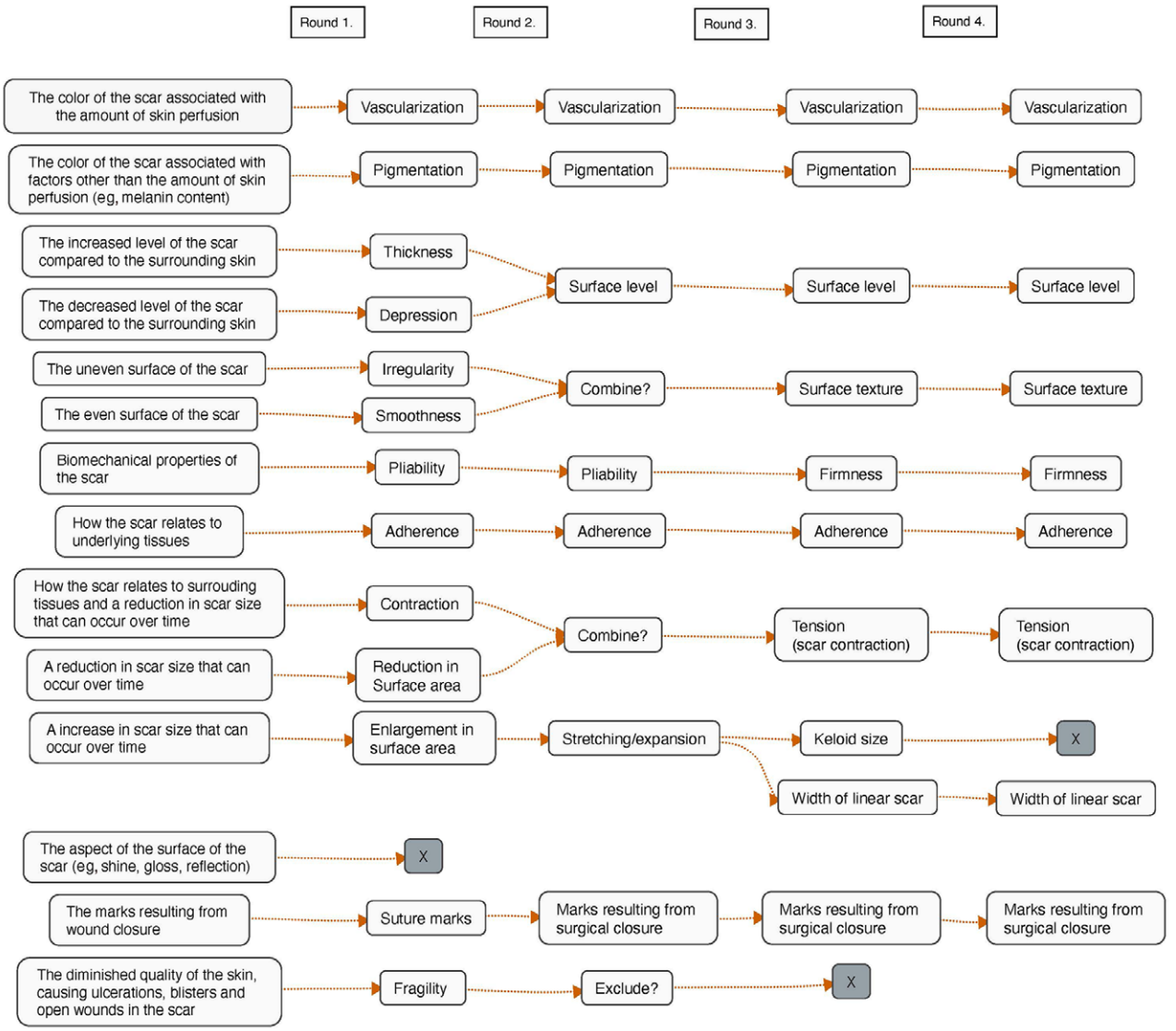

Fig. 3. Proposed scar characteristics and how these evolved throughout subsequent Delphi rounds.

except linear scars, and includes 7 items (**Supplemental Digital Content 2,** http://links.lww.com/PRSGO/D760). The linear scar version (**Supplemental Digital Content 3,** http://links.lww.com/PRSGO/D761) is applicable to all linear scars caused by surgery or trauma. This version includes 9 items, 7 of which are the same as included in the generic version, and 2 extra items referring to the widening of scar edges and marks resulting from surgical closure. All scales were designed by a scientific communication specialist (the online scientist).

DISCUSSION

To date, the POSAS is the only instrument that allows for the assessment of both the patient and professional perspective on scar quality. This study describes the Delphi process leading to the development of the Observer Scales of the POSAS 3.0. To our knowledge, this is the first study to develop an outcome measurement instrument using the Delphi process in the field of scar assessment. We aimed to create a universally accepted outcome measurement instrument by consulting the potential end-user community of the new POSAS version with regard to its content.

When comparing the content of the Observer Scale of the POSAS 3.0 to the POSAS 2.0,[13] there are some substantial differences. First and foremost, a few extra items have been added to the POSAS 3.0 version. The single-item surface area that had been included in the POSAS 2.0, has now been further specified into 2 different items, adherence to underlying tissues and tension of the scar and adjacent structures, as a result of scar contracture. In addition, scar width and marks resulting from wound closure are new items included only in the linear scar version of the POSAS3.0. Second, a few items have been rephrased in the new POSAS version. For example, the previous item *relief* was rephrased into *surface texture*, and

the item *thickness* into *surface level*. Also, the item *pliability*, included in the POSAS 2.0, has been replaced by the item *firmness*. Third, the answering options have been revised in the POSAS 3.0. The previous version included a 10-point scale with end labels only, whereas the POSAS 3.0 has a verbal 5-point scale. Throughout this Delphi study, the type and number of answering options were important points of discussion. Unfortunately, the available literature on rating scales did not provide clear-cut answers regarding these issues.[14–16] The literature indicates that providing verbal anchors to scale points has been shown to improve the reliability and validity of the resulting data.[17] Furthermore, the performance of the 10-point scale in the POSAS 2.0 has been evaluated in 3 Rasch analysis studies.[18–20] These analyses showed that observers were able to discriminate between 9 of 10 of the answering options in burn scars[18] but were unable to do this for linear and keloid scars.[19,20] In one of these studies, a 5-point scale was suggested for linear scars.[19] In the third round of this study, panelists preferred the 5-point scale with verbal anchors, which is supported by the aforementioned research. A similar scale is used in the Patient Scale of the POSAS 3.0. Finally, the POSAS 2.0 consisted of 1 generic version for all scar types, whereas an extra version specifically for linear scars has been developed in the current study. Initially, we aimed to make a third version for keloid scars, but throughout this study, this was not indicated. Keloids are pathologically growing fibrotic lesions that develop and progress differently to other types of scars and require different treatment strategies.[21] Clinically, they can grow continuously and invade the adjacent healthy skin.[22] The pathological growth is a unique feature of keloids, which is impossible to quantify without the use of measurement devices. One of the overarching principles of the POSAS project has been to develop instruments on the basis of clinical features which can be used in any setting, without recourse to instrumentation. This was the reason to refrain from developing a keloid-specific POSAS version.

One of the strengths of the study was an extensive systematic review conducted in preparation of this study.[1] This provided a complete overview of all the characteristics of scar quality reported in the literature, enabling the assembly of a broad range of scar characteristics and terms in the first Delphi questionnaire. Another strength of this study was the emphasis on the narrative of the panelists' arguments. This provided a deeper understanding of their opinions, which improved our proposals. Rather than only relying on the percentage consensus, we focused on the rationale underpinning their decisions, and when these yielded new insights, they were included in the next round as alternative proposals. In addition, panelists were provided the opportunity to make comments or raise concerns regarding the study procedure in general. For example, panelists gave feedback on the second Delphi questionnaire, reflecting to us that the questionnaire was too wordy. This motivated us to make the third questionnaire more concise. Another strength of this study was our internationally diverse panel. We have included experts working with scars from 23 different countries spanning 7 different continents. The biggest group of panelists originated from Australia, the Netherlands, and the United Kingdom. These were also the countries where most of the members of the steering committee were based. Only 10% of the panelists originated from nonwestern regions (eg, Africa, India, [South]East-Asia, and South America). As a result, we may have missed some of the perspectives of a large group of other ethnicities. During the Delphi process, we extensively discussed how specific items (ie, vascularity and pigmentation) should be scored in patients with different skin types. As a result, we decided to rate these items by comparing the scarred skin to nonscarred skin in patients. In future research, it may be interesting to evaluate if the item *vascularity* is scored differently in patients with different skin types (ie, differential item functioning analysis).

Although most panelists had mixed caseloads, burn scars remained a primary reported focus of the panelists. As a result, agreement rates may have been higher on items that are important to the burn community. However, this potential bias was reduced by asking panelists in the second round of the Delphi to rate the importance of items for different scar types, and by developing 2 different versions. Although we gained some interesting data on keloids and acne scars, more research is necessary to determine how these scars can be best evaluated. In this Delphi study, we had an initial response rate of 46% among invited panelists (n = 268). Of the panelists that participated in the first round (n = 124), 55 (44%) completed all 3 rounds. A Delphi process is a qualitative method in which the emphasis is on the quality of the provided arguments. Therefore, the quality of the total number of responses is more important than the response rate. When comparing our study to other international Delphi studies among healthcare professionals, we had a relatively high number of panelists that participated in all rounds (n = 55).[23–26]

The rising importance of e-health and telemedicine promotes the use of remote scar measurements by both the patient and professional.[27] Multiple panelists suggested to incorporate example pictures for the standardization of item scores. At the moment, the POSAS is not suitable for remote measurements, as it contains tactile items that cannot be rated without touching the scar. However, future plans are being made for the development of a POSAS version appropriate for remote scar assessment, as well as the development of a Patient Scale of the POSAS version for children.

Currently, the reliability of the Observer Scale of the POSAS 3.0 is being established. More information on how to obtain and use the POSAS 3.0 scales can be found on the website www.posas.org.

CONCLUSIONS

This is the first study using the Delphi process to develop an outcome measurement instrument in the field of scar assessment. Panelists from all over the world reached a consensus on the content of 2 versions of the Observer Scale: the generic version and the linear scar version. By involving the potential end-user community of the new POSAS version in its development, it is expected that POSAS 3.0 will be associated with high content validity.

For a comprehensive scar assessment in patients with all scar types, the Observer Scale should be used in combination with the Patient Scales of the POSAS 3.0.

Michelle E. Carrière, MD
Department of Epidemiology and Data Science
Amsterdam UMC, Vrije Universiteit Amsterdam
Amsterdam, The Netherlands
E-mail: m.carriere@amsterdamumc.nl

APPENDIX 1: COLLABORATORS

Members of the POSAS Group: David G. Greenhalgh; Peter Moortgat†; Jun Wu‡; Rei Ogawa§; Naiem Moiemen‖; Adam J. Singer***

*From the *Davis Medical Center, Shriners Children's Northern California, University of California, Sacramento, CA; †Organization for Burns, Scar-aftercare and Research (OSCARE), Antwerp, Belgium; ‡Sun Yat-Sen University, Guangzhou, China; §Nippon Medical School, Tokyo, Japan; ‖Queen Elisabeth Hospital, University Hospitals Birmingham Foundation Trust, Birmingham, United Kingdom; and **Renaissance School of Medicine at Stony Brook University, New York, NY.*

DISCLOSURES

Van Zuijlen is the developer of the previous versions of the POSAS. This study was funded by the Dutch Burn Association (grant no. 16.106). The chair of Paul van Zuijlen is funded by the Dutch Burns Foundation. Mokkink and de Vet receive royalties payment for the textbook Measurement in Medicine A Practical Guide. Although the POSAS is free of use for health care professionals and patients, commercial parties are asked for a payment when the POSAS is being used for activities with commercial intention. The payments are managed by the Dutch Burn Foundation. The other authors have no financial interest to declare in relation to the content of this article.

ACKNOWLEDGMENTS

The authors would like to thank all panelists who participated in this Delphi study. The following panelists who participated through all rounds and agreed to be acknowledged in this article are listed.

R.B. Ahuja, C. Barron, N. Brewin, Chen, T.W. Chiu, N. Clayton, L. de Cuyper, H. DeJong, J. P.W. Don Griot, D. Edgar, S. Egg, D. Eisen, K. Everaars, I. Goutos, A. Guy, M.J. Hallam, K. Hietanen, M. Hjellestad, C. Huang, F. Issa, A.C. Issler-Fisher, A. Katsu, B. Kipping, M. Koziej, N.I. Lahdenperä, O. Lapid, S. Liu, T. Mahoney, A. McKittrick, J. Meirte, E. Middelkoop, B. Nedelec, F.B. Niessen, M.K. Nieuwenhuis, R.G. Owen, M. Özer, J. Paloneva, I. Parry, J.M. Passer, S. Rea, J. Repo, V. Schepel, S. Scholten, N. Scuderi, M. Simons, J. Sobb, K. Stiles, A. Thynne, T. Travis, M. van Baar, A.L. van de Kar, E. van den Kerckhove, C.M. van der Horst, C.H. van der Vlies, M.B.A. van der Wal, S. Vercelli, P.D. Verhaegen, N. Wilson-Jones, S. Wolf, J. Wong, R. Wong She, F. Wood, J. Zuccaro.

REFERENCES

1. Carrière ME, Kwa KAA, de Haas LEM, et al. Systematic review on the content of outcome measurement instruments on scar quality. *Plast Reconstr Surg Glob Open*. 2019;7:e2424.
2. Carrière ME, Tyack Z, Westerman MJ, et al. From qualitative data to a measurement instrument: a clarification and elaboration of choices within the development of the Patient Scale of the Patient and Observer Scar Assessment Scale (POSAS) 3.0. *Burns*. 2022;49:1541–1556.
3. Durani P, McGrouther DA, Ferguson MW. Current scales for assessing human scarring: a review. *J Plast Reconstr Aesthet Surg*. 2009;62:713–720.
4. Tyack Z, Simons M, Spinks A, et al. A systematic review of the quality of burn scar rating scales for clinical and research use. *Burns*. 2012;38:6–18.
5. Vercelli S, Ferriero G, Sartorio F, et al. Clinimetric properties and clinical utility in rehabilitation of postsurgical scar rating scales: a systematic review. *Int J Rehabil Res*. 2015;38:279–286.
6. van Zuijlen PPM, Mokkink LB, Hoogewerf CJ, et al. The official update of the POSAS: an invitation to share experiences to improve the POSAS in "Project POSAS 3.0". *Burns*. 2017;43:893–894.
7. Carrière ME, Mokkink LB, Tyack Z, et al. Development of the Patient Scale of the Patient and Observer Scar Assessment Scale (POSAS) 3.0: a qualitative study. *Qual Life Res*. 2023;32:583–592.
8. Powell C. The Delphi technique: myths and realities. *J Adv Nurs*. 2003;41:376–382.
9. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. *Inf Manage*. 2004;42:15–29.
10. Linstone HA, Turoff M. *The Delphi Method: Techniques and Applications*. Addison-Wesley Publishing Company, Advanced Book Program; 2002.
11. Steurer J. The Delphi method: an efficient procedure to generate knowledge. *Skeletal Radiol*. 2011;40:959–961.
12. Mokkink LB, Boers M, van der Vleuten CPM, et al. COSMIN risk of bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Med Res Methodol*. 2020;20:293.
13. van de Kar AL, Corion LU, Smeulders MJ, et al. Reliable and feasible evaluation of linear scars by the Patient and Observer Scar Assessment Scale. *Plast Reconstr Surg*. 2005;116:514–522.
14. Beckstead JW. On measurements and their quality. Paper 4: verbal anchors and the number of response options in rating scales. *Int J Nurs Stud*. 2014;51:807–814.
15. DeCastellarnau A. A classification of response scale characteristics that affect data quality: a literature review. *Qual Quant*. 2018;52:1523–1559.
16. Steinberg L, Rogers A. Changing the scale: the effect of modifying response scale labels on the measurement of personality and affect. *Multivariate Behav Res*. 2022;57:79–93.
17. Boote AS. Reliability testing of psychographic scales. *J Advert Res*. 1981;21:53–60.
18. van der Wal MB, Tuinebreijer WE, Bloemen MC, et al. Rasch analysis of the Patient and Observer Scar Assessment Scale (POSAS) in burn scars. *Qual Life Res*. 2012;21:13–23.
19. Franchignoni F, Giordano A, Vercelli S, et al. Rasch analysis of the Patient and Observer Scar Assessment Scale in linear scars: suggestions for a Patient and Observer Scar Assessment Scale v2.1. *Plast Reconstr Surg*. 2019;144:1073e–1079e.
20. van der Wal MB, Tuinebreijer WE, Lundgren-Nilsson A, et al. Differential item functioning in the Observer Scale of the POSAS for different scar types. *Qual Life Res*. 2014;23:2037–2045.
21. Ogawa R. The most current algorithms for the treatment and prevention of hypertrophic scars and keloids: a 2020 update of the algorithms published 10 years ago. *Plast Reconstr Surg*. 2022;149:79e–94e.
22. Huang C, Ogawa R. Systemic factors that shape cutaneous pathological scarring. *FASEB J*. 2020;34:13171–13184.

23. Couzner L, Day S, Draper B, et al. What do health professionals need to know about young onset dementia? An international Delphi consensus study. *BMC Health Serv Res*. 2022;22:14.
24. Requejo-Salinas N, Lewis J, Michener LA, et al. International physical therapists consensus on clinical descriptors for diagnosing rotator cuff related shoulder pain: a Delphi study. *Braz J Phys Ther*. 2022;26:100395.
25. Jordans IPM, Verberkt C, De Leeuw RA, et al. Definition and sonographic reporting system for cesarean scar pregnancy in early gestation: modified Delphi method. *Ultrasound Obstet Gynecol*. 2022;59:437–449.
26. Rongo R, Ekberg E, Nilsson IM, et al. Diagnostic criteria for temporomandibular disorders (DC/TMD) for children and adolescents: an international Delphi study—part 1—development of axis I. *J Oral Rehabil*. 2021;48:836–845.
27. Wilson LS, Maeder AJ. Recent directions in telemedicine: review of trends in research and practice. *Healthc Inform Res*. 2015;21:213–222.