University of Groningen

Establishing Content Validity of the CLEFT-Q

Tsangaris, Elena; Riff, Karen W. Y. Wong; Goodacre, Tim; Forrest, Christopher R.; Dreise, Marieke; Sykes, Jonathan; de Chalain, Tristan; Harman, Karen; O'Mahony, Aisling; Pusic, Andrea L.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Tsangaris, E., Riff, K. W. Y. W., Goodacre, T., Forrest, C. R., Dreise, M., Sykes, J., de Chalain, T., Harman, K., O'Mahony, A., Pusic, A. L., Thabane, L., Thoma, A., & Klassen, A. F. (2017). Establishing Content Validity of the CLEFT-Q: A New Patient-reported Outcome Instrument for Cleft Lip/Palate. *Plastic and Reconstructive Surgery. Global Open*, *5*(4), [1305]. https://doi.org/10.1097/GOX.0000000000001305

Establishing Content Validity of the CLEFT-Q: A New Patient-reported Outcome Instrument for Cleft Lip/Palate

Elena Tsangaris, MSc*
Karen W.Y. Wong Riff, MD, MSc, FRCSC†
Tim Goodacre, MBBS, FRCS‡
Christopher R. Forrest, MD, MSc, FRCSC†
Marieke Dreise, BA§
Jonathan Sykes, MD¶
Tristan de Chalain, MB, ChB‖
Karen Harman, MD**
Aisling O'Mahony, DDS††
Andrea L. Pusic, MD, MPH, FACS‡‡
Lehana Thabane, PhD*
Achilleas Thoma, MD, FRCSC, FACS§§
Anne F. Klassen, DPhil¶¶

Background: The CLEFT-Q is a new patient-reported outcome instrument designed to measure outcomes that matter to patients. The aim of this qualitative study was to establish content validity of the CLEFT-Q in patients who differ by age and culture.

Methods: Patients aged between 6 and 29 years were recruited from plastic surgery clinics in Canada, India, Ireland, the Philippines, the Netherlands and the United States. Healthcare providers and other experts participated in a focus group or provided individual feedback. Input was sought on all aspects of the CLEFT-Q (item wording, instructions, and response options), and to identify missing content. Patient interviews and expert feedback took place between September 2013 and September 2014.

Results: Sixty-nine patients and 44 experts participated. The first draft of the CLEFT-Q consisted of 163 items measuring 12 constructs. The first round of feedback identified 92 items that required revision. In total, 3 rounds of interviews, and the involvement of an artist to create pictures for 17 items, were needed to establish content validity. At the conclusion of cognitive interviews, the CLEFT-Q consisted of 13 scales (total 171 items) that measure appearance, health-related quality of life, and facial function. The mean Flesch-Kincaid readability statistic for items was 1.4 (0 to 5.2).

Conclusion: Cognitive interviews and expert review allowed us to identify items that required re-wording, re-conceptualizing, or to be removed, as well as any missing items. This process was useful for refining the CLEFT-Q scales for further testing. (*Plast Reconstr Surg Glob Open 2017;5:e1305; doi: 10.1097/GOX.0000000000001305; Published online 25 April 2017.*)

INTRODUCTION

Cleft lip and/or palate (CL/P) is the most common congenital craniofacial anomaly.[1–3] CL/P may negatively affect an individual's appearance, speech, or dentition, resulting in diminished health-related quality of life (HRQOL).[4,5] Treatments for patients with CL/P aim to improve facial appearance and function, as well as psychological and social health.

DOI: 10.1097/GOX.0000000000001305

Disclosure: *The authors have no financial interest to declare in relation to the content of this article. The Article Processing Charge was paid from the CIHR grant.*

Supplemental digital content is available for this article. Clickable URL citations appear in the text.

Outcomes of CL/P treatment have typically been appraised objectively using observer-reported or clinician-reported assessments.[6–8] However, because the goal of CL/P treatment is to improve a patient's physical health and quality of life, these outcomes would be difficult to measure through the use of observer- or clinician-reported outcome assessments alone. Patient-reported outcome (PRO) instruments are a unique tool that can be used to accurately evaluate a patient's perspective of their own health outcomes.[9] Understanding the patients' perspective using carefully designed PRO instruments could provide evidence-based information to inform clinical practice and future research.

A systematic review by Eckstein et al.[4] identified 44 PRO instruments that had been used to measure quality of life or satisfaction in patients with CL/P. Although 5 questionnaires were validated in a CL/P population, none were developed with a CL/P focus.[4] The authors of this review call for the development of a new valid and reliable cleft-specific PRO instrument that could be used in patient care and clinical research to evaluate the impact of surgery and treatment on patients' quality of life.[4]

Best practice for the development of a PRO instrument is iterative and involves item generation, item reduction, and psychometric evaluation.[10,11] Phase 1 should involve the development of a conceptual framework, and the generation of items via a literature review, qualitative interviews with patients, and expert feedback.[10,11] To finalize phase 1, cognitive interviews with patients are necessary to ensure that the items forming scales are relevant to patients and are appropriately understood, to minimize error that may result from item misinterpretation during data collection.[10–12] Expert feedback in conjunction with cognitive interviews may also help to provide information about the clinical relevance of specific items.[13,14]

The CLEFT-Q is a specific and unique PRO instrument developed to measure outcomes that matter to children and young adults with CL/P. The aim of the present study was to establish content validity of the CLEFT-Q. Content validity is a measurement property that appraises whether items in a scale are comprehensive and adequately reflect the perspective of the population of interest.[13] For the present study, we used cognitive interviews with patients and expert feedback to determine the content validity of the CLEFT-Q for patients who vary by age and culture.

METHODS

Research Ethics Board Approval

The study was approved by the Research Ethics Board at participating hospitals and was conducted in accordance with policies for ethical conduct in research involving humans. All participants or their legal guardians provided written informed consent or assent for participation according to each center's policy.

The CLEFT-Q

The CLEFT-Q was developed in accordance with internationally recommended guidelines for the development of a new PRO instrument.[9,11,15,16] Findings from a systematic review were used to develop a preliminary conceptual framework.[17] Concepts within the framework were developed into questions that formed an interview guide used in a series of qualitative interviews.[18] A total of 136 interviews were conducted with 138 patients (including 2 sets of twins who were interviewed together) with CL/P from 6 countries including Canada, England, India, Kenya, the Philippines, and the United States.[18] Interviews were recorded, translated/transcribed into English, coded, and analyzed, resulting in the refinement of the CLEFT-Q conceptual framework, which included 3 domains and 12 minor themes as follows: appearance (of the face, nose, teeth, lips, jaws, and cleft lip scar), HRQOL (psychological, social, school, and speech-related distress), and facial function (speech and eating/drinking). From the coded data, we also created a comprehensive item pool. For each minor theme, the item pool was used to develop a set of items that together mapped out a scale on a continuum, from more to less of the construct. Whenever possible, items were created using positive or neutral wording that could be understood by patients as young as 6 years old. For each scale, 4 labeled response options were chosen to align with published guidelines.[19] Each CLEFT-Q scale was designed to be independently functioning. This approach aimed to reduce patient burden, as only the scales relevant to the research or clinical purpose need to be completed.

The first draft of the CLEFT-Q was translated into Assamese (India), Tagalog (the Philippines), and Dutch (the Netherlands) to facilitate cognitive interviews in multiple countries. Translations were conducted in accordance with the International Society for Pharmacoeconomics and Outcomes Research guidelines for the translation and cultural adaptation of PRO instruments.[20]

Study Design, Data Collection, and Analysis

Table 1 illustrates the cognitive interviewing approach used, which was adapted from Willis.[21] Cognitive interviewing involved a series of one-on-one semi-structured interviews with patients, using a cognitive interview guide (**see PDF, Supplemental Digital Content 1**, which displays the cognitive interview guide, *http://links.lww.com/PRSGO/A427*). The objectives for cognitive interviewing were to determine challenges with participant comprehension of the item wording, instructions, and response options, using the think aloud approach and to identify items that were thought to be duplicate, missing, irrelevant, or insensitive.[21,22] Verbal probing was used concurrently, whereby the interviewer asked specific questions about content that was found to be problematic in preceding interviews.[21,22] This combined approach made it possible to identify problems with item interpretation and response selection.[12,23] For items expressed as ambiguous or difficult to understand, upon being informed of the item's meaning, participants were encouraged to suggest revisions to item wording. Similarly, experts provided feedback on the relevance and comprehensibility of the CLEFT-Q items, instructions, and response options, and were encouraged to suggest missing content.

Table 1. Cognitive Interview Approach

Cognitive Interviewing	Steps Outlined by Willis[21]	Modified Steps for Present Study
Purpose	To gather information about the functioning of the survey questions before to finalizing for field testing	None
Sample size	Typically small and may consist of several rounds (approximately 10 interviews per round)	3 rounds of interviews with 17–27 participants in each
Recruitment	Participants obtained through a variety of recruitment strategies to produce variations in the types of individuals recruited	Participants were recruited during their clinic visits or were contacted by telephone
Interviewers	Generally a small number (1–4) of highly trained cognitive interviewers	Most interviews were conducted by E.T.; however, 2 additional interviewers were trained by E.T. to conduct interviews in India and the Netherlands
Materials presented	A cognitive interview guide consisting of the survey questionnaire along with probe questions	None
Method	Flexible administration that relies heavily on probe questions to enhance the quality of the questionnaire	None
Analysis	Qualitative analysis of responses based on interviewer notes or recording of verbatim interviews	Qualitative analysis conducted based on interviewer notes

This table was adapted from Willis GB. Analysis of the cognitive interview in questionnaire design: understanding qualitative research. Toronto: Oxford University Press; 2015.
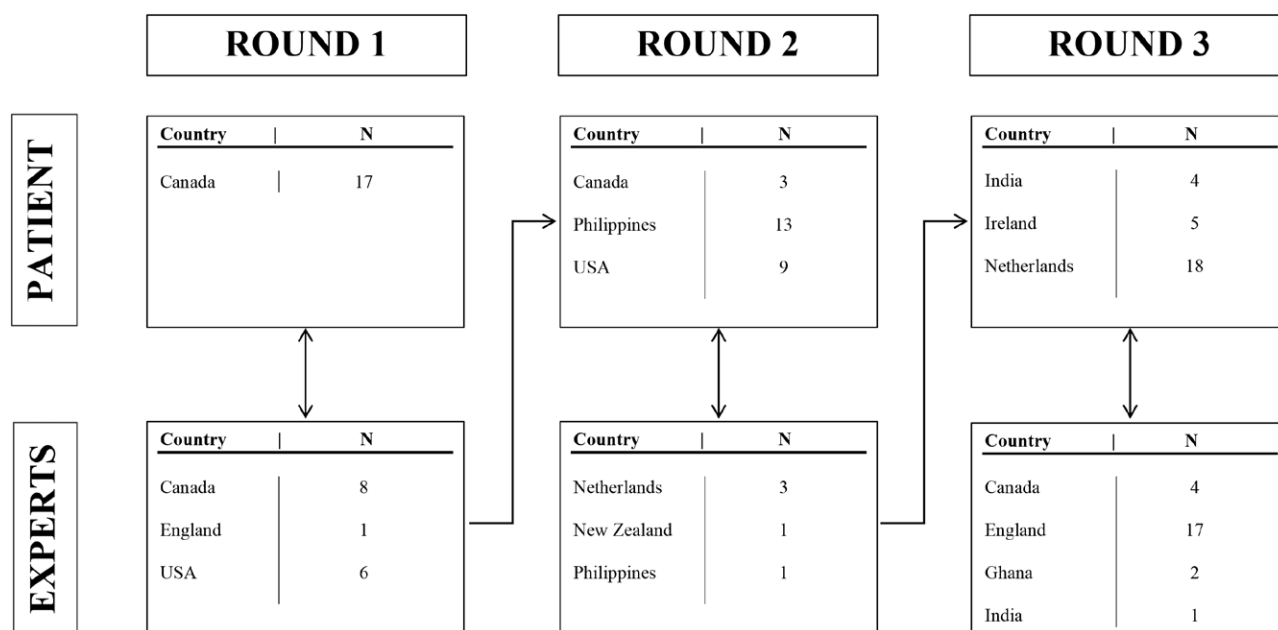
Fig. 1. Sequence of cognitive interviews with patients and expert feedback.

Cognitive interviews and expert feedback were conducted in rounds. Figure 1 highlights the sequence of cognitive interviews and expert feedback by country and round. Round 1 interviews involved patients from McMaster Children's Hospital (Hamilton) and the Hospital for Sick Children (Toronto) in Canada; round 2 interviews involved patients from Canada, UC Davis Health System (Sacramento, Calif.) in the United States, and the Operation Restore Hope New Zealand medical mission trip to the Philippines; and round 3 interviews involved patients from St. James Hospital in Dublin, Ireland, Operation Smile Comprehensive Cleft Care Center in Guwahati, India, and the University of Groningen in the Netherlands. Expert feedback was interspersed throughout each round and varied by country and expertise.

To obtain expert feedback, in round 1, focus groups were held in Canada and the United States (facilitated by E.T. and K.W.R., respectively). Each focus group involved health-care providers, whose clinical focus was CL/P, meeting together to review and discuss the CLEFT-Q content. During each round, the CLEFT-Q was distributed to a range of experts who provided written feedback on the instrument. In the third round, the CLEFT-Q was circulated for feedback to 16 members of the National Cleft Psychologists Special Interest Group in England. This group discussed its content at a meeting and provided written feedback as a group.

Patient interview and expert feedback data were entered into a Microsoft Excel (2016) worksheet for analysis. A reparative approach to data analysis was employed, which involved examining findings and revising items

concurrently.[21,22,24] Changes were made to the CLEFT-Q after each round, where the evidence collected was brought to the research team to assess the compiled results and make revisions before moving on to the next round of item testing. Willis[22] recommended conducting 2–4 sets of interviews with 5–15 people in each. Data were analyzed using the "text summary" approach, which involved summarizing notes from cognitive interviews and expert feedback, to identify consistent themes.[22]

In each round of revisions, considerations were necessary to ensure that the CLEFT-Q items had the lowest possible Flesch–Kincaid (F–K) readability level. The F–K readability score indicates the grade-reading level of an item. Readability cutoffs were determined in accordance with the reading comprehension literature.[25–27]

Sampling

Eligible participants were patients aged 6 years and older with CL/P, who could read and understand any of the target languages. Age eligibility was based on research reporting that children as young as 5 years of age are able to self-report on age-appropriate questionnaires, and children as young as 8 years of age are able to self-report on wellbeing, psychosocial health, and health-promoting behaviors.[28,29]

In Canada, Ireland, India, the Netherlands, and the United States, a member of the health-care team approached patients in clinic or by telephone and invited them to consider participation in the study. Interviews were set up either by a health care or research team member and were conducted in the hospital, at the patients' home, or over the phone, depending on the site or patient preference. In the Philippines, interviews took place during a week-long Operation Restore Hope surgical mission trip to Batangas, Philippines. The intake nurse informed patients of the study, and patients willing to participate were interviewed before or after their surgery.

Experts in the field of CL/P were recruited through our team's professional network. Feedback was obtained face-to face during focus groups or through individual feedback by e-mail.

RESULTS

Translation and data collection for the CLEFT-Q cognitive interviews and expert feedback took place between September 2013 and September 2014. Sixty-nine patients participated in our cognitive interviews. Mean age of participants at the time of recruitment was 13.2 years (range, 6–26 years old), with 58% of the sample aged 13 years or younger. More females (52%) participated, and most patients had CL/P (77%; Table 2). Forty-four experts provided individual feedback, of which 13 participated in 1 of 2 focus groups. Most experts were psychologists (48%) and were from England (41%; Table 3).

During the cognitive interviews and expert feedback, items that were identified to be hard to understand, or considered irrelevant, were either revised to improve comprehensibility or dropped. Items that were dropped were mainly because the item (1) represented a difficult

Table 2. Demographic Characteristics of the Patient Who Participated in the Cognitive Interviews

Demographic Variables	No. Patients, Count (%), N = 69
Country	
Canada	20 (29.0)
India	4 (5.8)
Ireland	5 (7.3)
The Philippines	13 (18.8)
The Netherlands	18 (26.1)
The United States	9 (13.0)
Sex	
Male	33 (47.8)
Female	36 (52.2)
Age (y)	
6–9	16 (23.2)
10–13	24 (34.8)
14–17	13 (18.8)
18–21	11 (16.0)
22–29	5 (7.2)
Type of cleft	
Cleft lip only	9 (13.0)
Cleft palate only	6 (8.7)
Cleft lip and palate	53 (76.8)
Cleft lip and alveolus	1 (1.5)

concept in which the meaning could not be clarified, (2) was not considered to be important by patients, or (3) was considered clinically irrelevant by experts.

We conducted 3 rounds of revisions involving patients and experts. At the start of round 1, the CLEFT-Q consisted of 163 items in 12 scales. Feedback from 17 patients and 15 experts (Fig. 1) led to 25 items that remained the same, 92 revised, 46 deleted, and 48 added. Most of the revisions in this round were needed to ensure that item wording worked with the new response options (n = 58 items), as feedback led to us change the response options for all 7 appearance scales, ie, from "Strongly Disagree, Disagree, Agree, Strongly Agree" to "Like a Lot, Like a Little, Dislike a Little, Dislike a Lot." This change was required because patients and experts expressed some difficulty applying the response options to the items. Thirty-four items within the HRQOL and facial function scales also required revision. By completion of round 1, the CLEFT-Q consisted of 165 items.

A total of 25 patients and 5 experts provided feedback in round 2 (Fig. 1). In this round, the vast majority of items (total 140, 85%) remained the same, 23 items were revised, 2 were dropped, and 6 were added. Items were revised mainly to clarify their meaning or to add an example.

In round 3, 169 items were tested in 27 patients, with 24 experts providing feedback (Fig. 1). At this stage, very minor changes were required. A total of 137 items remained the same, 28 were revised, 4 were dropped, and 6 were added. Items that patients had some difficulty with at this stage tended to be specific appearance items (total 17, 61%). To ensure that the appearance items were as easy as possible to understand, we had an artist create 17 images that illustrated specific parts of the face [eg, … how the tip of your nose looks (the very end of your nose)?], or how specific parts of the face look during movement [eg, ... how much you can move your lips (like to whistle or kiss?)]. We also included a picture of nos-

Table 3. Demographic Characteristics of the Experts Who Provided Feedback

Speciality	Canada	England	Ghana	India	The Netherlands	New Zealand	The Philippines	The United States	Total Count (%)
Audiologist	1	0	0	0	0	0	0	0	1 (2.3)
Cleft care coordinator	1	0	0	0	0	0	0	0	1 (2.3)
Developmental pediatrician	1	0	0	0	0	0	0	0	1 (2.3)
Nurse	0	0	0	0	1	0	0	0	1 (2.3)
Orthodontist	1	0	0	0	0	0	0	0	1 (2.3)
Pediatric otorhinolaryngologist	1	0	0	0	0	0	0	0	1 (2.3)
Plastic surgeon	2	0	1	1	1	1	1	1	8 (18.1)
Psychologist	0	16	0	0	1	0	0	4	21 (47.7)
Psychometrician	0	1	0	0	0	0	0	0	1 (2.3)
Speech-language pathologist	3	1	1	0	0	0	0	0	5 (11.3)
Social worker	2	0	0	0	0	0	0	1	3 (6.8)
Total count (%)	12 (27.2)	18 (40.9)	2 (4.6)	1 (2.3)	3 (6.8)	1 (2.3)	1 (2.3)	6 (13.6)	N = 44

trils, jaws, and a cleft lip scar to ensure that participants knew exactly which part of the face the scales referred to. At the completion of the cognitive interviews and expert review, the CLEFT-Q comprised 171 items within 13 scales that reflected the original conceptual framework (Table 4).

Instructions, Response Options, and Readability

Although minor changes were needed to finalize the instructions for the CLEFT-Q scales, the 3 rounds provided time to explore different sets of response options for each scale. At the end of the process, the 7 scales that measure appearance ask respondents to answer each item thinking of how their face (or specific area of their face) looks now. Respondents are then asked to answer for each item "how much do you like …" using the 4 response options listed above. All other CLEFT-Q scales include a series of statements, with instructions asking respondents to answer each item in relation to the past week, and in terms of frequency: "Never," "Sometimes," "Often," and "Always."

Mean F–K readability for the 171 items was 1.4 (range, 0–5.2) for scales (Table 4). The readability scores were below the fifth-grade reading level, with the exception of 2 items in the psychological scale: *"I feel okay about myself"* and *"I feel confident,"* which scored slightly higher (F–K readability score = 5.2).

Table 5 provides an example of how a CLEFT-Q scale (Cleft Lip Scar) was modified after each round. This scale differed from all other appearance scales, as the original items had either negative content or used negative phrasing. Experts who participated in round 1 suggested that we reconceptualize the scale to create a version with items that used positive or neutral content and phrasing. From the original set of 10 items, 8 items were dropped, 2 were revised and retained, and 9 new items were included from the original item pool. Further changes made in relation to feedback from subsequent rounds involved minor wording changes to improve clarity.

DISCUSSION

Cognitive interviews with patients and expert feedback were used to establish content validity of the CLEFT-Q for patients who varied by age and culture. This psychometric property was achieved by obtaining feedback from a large international sample of patients, who helped us refine instructions and a set of items for each scale, and to choose appropriate response options. Content validity was also established through feedback from an international sample of experts in CL/P who provided insights about the suitability and perceived difficulty of items. Input received in round 1 led to substantial revisions of the first draft of the CLEFT-Q. Once we made these revisions, subsequent rounds largely involved the "fine tuning" of items

Table 4. Characteristics of the Pilot-test Version of the CLEFT-Q Scales

Concept Measured	No. Items	Cleft Type for Completion of Scales	F–K, Range (Mean)
Appearance scales			
Cleft lip scar	13	CL/P, CLA, CL	0–2.4 (0.9)
Face	16	Any cleft diagnosis	0–3.6 (0.6)
Jaws	12	Any cleft diagnosis	0–2.8 (0.8)
Lips	17	Any cleft diagnosis	0–2.3 (0.4)
Nose	13	Any cleft diagnosis	0–2.5 (0.7)
Nostrils	9	Any cleft diagnosis	0.5–2.4 (1.4)
Teeth	12	Any cleft diagnosis	0–2.4 (0.6)
HRQOL scales			
Psychological function	14	Any cleft diagnosis	0–5.2 (1.9)
School function	11	Any cleft diagnosis	0–3.7 (1.8)
Social function	16	Any cleft diagnosis	0–3.7 (2.3)
Speech-related distress	12	CL/P, CP, CLA	0.5–3.8 (2.3)
Facial function scales			
Eating/drinking	10	Any cleft diagnosis	0–3.9 (1.3)
Speech	16	CL/P, CP, CLA	0–5.0 (2.8)

CLA, cleft lip and alveolus; CL, cleft lip only; CP, cleft palate only.

Table 5. Example Modifications to the Cleft Lip Scar Scale after Each Round of Cognitive Interviews and Expert Feedback

Round 1 Items	Decision	Round 2 Items	Decision	Round 3 Items	Decision	Final Items for Pilot-testing
My cleft scar looks big.	Revise	… the size of your cleft scar?	Retain	… the size of your cleft scar?	Revise	… the size of your cleft lip scar?
My cleft scar looks odd when I smile.	Revise	… how your cleft scar looks when you smile?	Retain	… how your cleft scar looks when you smile?	Revise	… how your cleft lip scar looks when you smile?
I do not like how my cleft scar looks.	Drop	—	—	—	—	—
My cleft scar is easy for people to see.	Drop	—	—	—	—	—
My cleft scar looks uneven.	Drop	—	—	—	—	—
My cleft scar looks abnormal.	Drop	—	—	—	—	—
My cleft scar feels tight.	Drop	—	—	—	—	—
My cleft scar looks ugly.	Drop	—	—	—	—	—
My cleft scar looks weird.	Drop	—	—	—	—	—
My cleft scar looks wide.	Drop	—	—	—	—	—
—	—	—	—	—	Add	… the width of your cleft lip scar?
—	Add	… the shape of your cleft scar?	Retain	… the shape of your cleft scar?	Revise	… the shape of your cleft lip scar?
—	Add	… the color of your cleft scar?	Retain	… the color of your cleft scar?	Revise	… the color of your cleft lip scar?
—	Add	… how your cleft scar looks when you laugh?	Retain	… how your cleft scar looks when you laugh?	Drop	—
—	Add	… how your cleft scar looks in photos?	Revise	… how your cleft scar is in photos?	Revise	… how your cleft lip scar looks in photos?
—	Add	… how your cleft scar looks in the mirror?	Retain	… how your cleft scar looks in the mirror?	Revise	… how your cleft lip scar looks in the mirror?
—	Add	… how your cleft scar looks from far away?	Retain	… how your cleft scar looks from far away?	Revise	… how your cleft lip scar looks from far away?
—	Add	… how your cleft scar looks up close?	Retain	… how your cleft scar looks up close?	Revise	… how your cleft lip scar looks up close?
—	Add	… how much your cleft scar has faded?	Retain	… how much your cleft scar has faded?	Revise	… how much your cleft lip scar has faded over time?
—	Add	… how your cleft scar feels when you touch it?	Retain	… how your cleft scar feels when you touch it?	Revise	… how your cleft lip scar feels when you touch it (smooth or bumpy)?
—	—	—	—	—	Add	… how much the color of your cleft lip scar matches your skin color?

to improve comprehension as much as possible. After round 3, the CLEFT-Q was determined to be ready for pilot testing, in which the results were used to further refine the CLEFT-Q items and response options.

Assessment of outcomes in CL/P care has primarily relied on objective evaluations by health-care providers, with very little patient input. Engaging patients in the assessment of treatment outcomes may provide an important perspective in research that measures the impact of CL/P care globally. In 2000 and 2001, the World Health Organization's international consensus meetings on craniofacial anomalies called for outcome measures that capture issues that "… matter to ordinary people rather than sophisticated surrogate measures that may have little relevance in everyday life."[30] More recently, Mossey et al.[31] restated the World Health Organization recommendation in a seminal paper, where they stressed the need for standardized PRO instruments. Specifically, a need was expressed for "… psychological and QOL measures and economic outcomes."[31]

Neither our team's literature reviews[18,32] nor the updated review by Eckstein et al.[4] identified a CL/P-specific PRO instrument. Five PRO instruments, including the Youth Quality of Life–Facial Differences,[33] Pediatric Voice-Related Quality of Life,[34] Cleft Audit Protocol for Speech–Augmented,[35] Child Oral Health Impact Profile,[36,37] and Child Oral Health Quality of Life,[38–40] were validated in a population of patients with CL/P. These instruments capture a range of issues that are important to patients with CL/P, including oral health (Child Oral Health Quality of Life),[36,37] speech-related issues (Pediatric Voice-Related Quality of Life),[34] and craniofacial specific quality of life concerns (Youth Quality of Life–Facial Differences).[33] However, an important limitation is that 4 of these PRO instruments did not include qualitative patient input in their initial development (exception Youth Quality of Life–Facial Differences). The exclusion of qualitative patient input in the initial development of the various instruments may explain the absence of important cleft-specific concepts. To our knowledge, the CLEFT-Q is the first self-report CL/P-specific PRO instrument developed according to published guidelines for PRO instrument development.[18] The CLEFT-Q covers cleft-specific issues from the perspective of patients who varied by age and culture.[18]

Most studies developing a new PRO instrument do not typically use as large a sample as we did, either for the initial development (n = 138) or cognitive interview (n = 69) stages. Our team decided it was crucial to ensure that the content of the CLEFT-Q and the final wording of each item resonated with young children and young adults with different types of clefts and from multiple countries, including low- and middle-income countries. Therefore, an important strength of our study is the inclusion of a large, heterogeneous, international sample of patients who took part in the cognitive interviews to refine the scales. Additionally, consistent cognitive interview methodology procedures were used throughout the process. This consistency was maintained by having 1 experienced qualitative researcher conduct all English interviews and oversee 22 interviews conducted in non-English languages.

Although we specifically set the inclusion criteria for age to be at the lower range of what is possible for self-report (6 years old),[28] a limitation of our study is that only 2 children were 6 or 7 years of age. Although both patients were able to read, and accurately interpret and respond to items in the CLEFT-Q scales, further research is required to determine if the CLEFT-Q (or specific scales) can be completed by children under the age of 8 years. In addition, the use of interviewer notes rather than tape recordings may have resulted in incomplete data because the interviewer may not have been as comprehensive in note taking. Therefore, this approach to data collection may have resulted in key comments from participants being missed.[21] However, given the number of participants in our study, and the numerous countries involved, we anticipate that this potential problem is unlikely to be clinically important.

CONCLUSIONS

No changes were required to the CLEFT-Q conceptual framework. Cognitive interviews and expert review allowed us to identify items that required rewording, reconceptualizing, or to be removed, as well as any missing items. This process was useful for refining the CLEFT-Q scales. The CLEFT-Q has now been field tested in an international study that involved 30 hospitals in 12 countries. In addition, a subset of the CLEFT-Q scales has been included in the International Consortium for Health Outcomes Measurement standard set.[41]

Anne F. Klassen, DPhil
Department of Pediatrics
McMaster University
Room HSC 3N27
1280 Main Street West
Hamilton, Ontario L8S 4K1, Canada
E-mail: aklass@mcmaster.ca

ACKNOWLEDGMENTS

REFERENCES

1. Tollefson TT, Senders CW, Sykes JM. Changing perspectives in cleft lip and palate: from acrylic to allele. *Arch Facial Plast Surg.* 2008;10:395–400.
2. Murray JC, Daack-Hirsch S, Buetow KH, et al. Clinical and epidemiologic studies of cleft lip and palate in the Philippines. *Cleft Palate Craniofac J.* 1997;34:7–10.
3. Das SK, Runnels RS Jr, Smith JC, et al. Epidemiology of cleft lip and cleft palate in Mississippi. *South Med J.* 1995;88:437–442.
4. Eckstein DA, Wu RL, Akinbiyi T, et al. Measuring quality of life in cleft lip and palate patients: currently available patient-reported outcomes measures. *Plast Reconstr Surg.* 2011;128:518e–526e.
5. Agbenorku P. Orofacial clefts: a worldwide review of the problem. Available at https://www.hindawi.com/journals/isrn/2013/348465. Accessed June 10, 2016.

6. Semb G, Brattström V, Mølsted K, et al. The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 1: introduction and treatment experience. *Cleft Palate Craniofac J.* 2005;42:64–68.

7. Long RE Jr, Hathaway R, Daskalogiannakis J, et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 1. Principles and study design. *Cleft Palate Craniofac J.* 2011;48:239–243.

8. Heliövaara A, Küseler A, Skaare P, et al. Scandcleft randomized trials of primary surgery for unilateral cleft lip and palate: 6. Dental arch relationships at 5 years. *J Plast Surg Hand Surg* 2016;25:1–6.

9. US Food and Drug Administration Guidance for Industry. *Patient-reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.* Rockville, Md.: Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research; 2009.

10. Cano SJ, Klassen A, Pusic AL. The science behind quality-of-life measurement: a primer for plastic surgeons. *Plast Reconstr Surg.* 2009;123:98e–106e.

11. 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 product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding. *Value Health.* 2011;14:978–988.

12. García AA. Cognitive interviews to test and refine questionnaires. *Public Health Nurs.* 2011;28:444–450.

13. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. *Qual Life Res.* 2009;18:1263–1278.

14. Rothrock NE, Kaiser KA, Cella D. Developing a valid patient-reported outcome measure. *Clin Pharmacol Ther.* 2011;90:737–742.

15. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002;11:193–205.

16. 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 product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument. *Value Health.* 2011;14:967–977.

17. Klassen AF, Tsangaris E, Forrest CR, et al. Quality of life of children treated for cleft lip and/or palate: a systematic review. *J Plast Reconstr Aesthet Surg.* 2012;65:547–557.

18. Wong Riff KW, Tsangaris E, Goodacre T, et al. International multiphase mixed methods study protocol to develop a cross-cultural patient-reported outcome instrument for children and young adults with cleft lip and/or palate (CLEFT-Q). *BMJ Open.* 2017;7:e015467.

19. Khadka J, Gothwal VK, McAlinden C, et al. The importance of rating scales in measuring patient-reported outcomes. *Health Qual Life Outcomes.* 2012;10:80.

20. Wild D, Grove A, Martin M, et al.; ISPOR Task Force for Translation and Cultural Adaptation. 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;8:94–104.

21. Willis GB. *Analysis of the Cognitive Interview in Questionnaire Design: Understanding Qualitative Research.* Toronto: Oxford University Press; 2015.

22. Willis GB. *Cognitive Interviewing: A Tool for Improving Questionnaire Design.* New York: Sage Publications; 2005.

23. Murtagh FE, Addington-Hall JM, Higginson IJ. The value of cognitive interviewing techniques in palliative care research. *Palliat Med.* 2007;21:87–93.

24. Miller K. Cognitive interviewing. In Madans J, Miller K. Maitland A, Willis G, eds. *Question Evaluation Methods.* New Jersey: Wiley; 2011:51–75.

25. Doak CC, Doak LG, Root JH. *Teaching Patients with Low Literacy Skills.* Philadelphia, PA: J.B. Lippincott; 1996.

26. Weiss BD, Coyne C. Communicating with patients who cannot read. *N Engl J Med.* 1997;337:272–274.

27. Prout HT, Chizik R. Readability of child and adolescent self-report measures. *J Consult Clin Psychol.* 1988;56:152–154.

28. Varni JW, Limbers CA, Burwinkle TM. How young can children reliably and validly self-report their health-related quality of life?: An analysis of 8,591 children across age subgroups with the PedsQL 4.0 Generic Core Scales. *Health Qual Life Outcomes.* 2007;5:1.

29. Bevans KB, Riley AW, Moon J, et al. Conceptual and methodological advances in child-reported outcomes measurement. *Expert Rev Pharmacoecon Outcomes Res.* 2010;10:385–396.

30. World Health Organization. *Global Strategies to Reduce the Healthcare Burden of Craniofacial Anomalies: Report of WHO Meetings on International Collaborative Research on Craniofacial Anomalies.* Geneva: World Health Organization; 2000.

31. Mossey PA, Little J, Munger RG, et al. Cleft lip and palate. *Lancet.* 2009;374:1773–1785.

32. Klassen A, Stotland MA, Skarsgard ED, Pusic AL. Clinical research in pediatric plastic surgery and systematic review of quality of life questionnaires. *Clin Plast Surg.* 2008;35:251–267.

33. Patrick DL, Topolski TD, Edwards TC, et al. Measuring the quality of life of youth with facial differences. *Cleft Palate Craniofac J.* 2007;44:538–547.

34. Boseley ME, Cunningham MJ, Volk MS, et al. Validation of the pediatric voice-related quality-of-life survey. *Arch Otolaryngol Head Neck Surg.* 2006;132:717–720.

35. John A, Sell D, Sweeney T, et al. The cleft audit protocol for speech-augmented: a validated and reliable measure for auditing cleft speech. *Cleft Palate Craniofac J.* 2006;43:272–288.

36. Broder HL, McGrath C, Cisneros GJ. Questionnaire development: face validity and item impact testing of the Child Oral Health Impact Profile. *Community Dent Oral Epidemiol.* 2007;35(Suppl 1):8–19.

37. Broder HL, Wilson-Genderson M. Reliability and convergent and discriminant validity of the Child Oral Health Impact Profile (COHIP Child's version). *Community Dent Oral Epidemiol.* 2007;35(Suppl 1):20–31.

38. Jokovic A, Locker D, Guyatt G. Short forms of the Child Perceptions Questionnaire for 11-14-year-old children (CPQ11-14): development and initial evaluation. *Health Qual Life Outcomes.* 2006;4:4.

39. Jokovic A, Locker D, Stephens M, et al. Validity and reliability of a questionnaire for measuring child oral-health-related quality of life. *J Dent Res.* 2002;81:459–463.

40. Jokovic A, Locker D, Stephens M, et al. Measuring parental perceptions of child oral health-related quality of life. *J Public Health Dent.* 2003;63:67–72.

41. International Consortium for Health Outcomes Measurement (ICHOM). Cleft lip and palate. Available at http://www.ichom.org/medical-conditions/cleft-lip-palate. Accessed June 10, 2016.