ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

Rasch Analysis of Patient- and Parent-Reported Outcome Measures in the International Consortium for Health Outcomes Measurement (ICHOM) Standard Set for Cleft Lip and Palate

Inge Apon, MD, Nikki van Leeuwen, DPhil, Alexander C. Allori, MD, MPH, Carolyn R. Rogers-Vizena, MD, Maarten J. Koudstaal, MD, DMD, PhD, Eppo B. Wolvius, MD, DMD, PhD, Stefan J. Cano, PhD, Anne F. Klassen, DPhil, Sarah L. Versnel, MD, PhD

ABSTRACT

Objectives: The aim of this study was to evaluate the psychometric performance of the patient- and parent-reported measures in the International Consortium for Health Outcomes Measurement (ICHOM) Standard Set for Cleft Care, and to identify ways of improving concept coverage.

Methods: Data from 714 patients with cleft lip and/or palate, aged 8 to 9, 10 to 12.5, and 22 years were collected between November 2015 and April 2019 at Erasmus University Medical Center, Boston Children's Hospital, Duke Children's Hospital, and from participating sites in the CLEFT-Q Phase 3 study. The Standard Set includes 9 CLEFT-Q scales, the Nasal Obstruction Symptom Evaluation (NOSE) questionnaire, the Child Oral Health Impact Profile–Oral Symptoms Scale (COHIP-OSS), and the Intelligibility in Context Scale (ICS). Targeting, item-fit statistics, thresholds for item responses, and measurement precision (PSI) were analyzed using Rasch measurement theory.

Results: The proportion of the sample to score within each instruments range of measurement varied from 69% (ICS) to 92% (CLEFT-Q teeth and COHIP-OSS). Specific problems with individual items within the NOSE and COHIP-OSS questionnaires were noted, such as poor item fit to the Rasch model and disordered thresholds (6 of 10). Reliability measured with PSI was above 0.82 for the ICS and all but one CLEFT-Q scale (speech distress). PSIs were lowest for the COHIP-OSS (0.43) and NOSE questionnaire (0.35).

Conclusion: The patient- and parent-reported components within the facial appearance, psychosocial function, and speech domains are valid measures; however, the facial function and oral health domains are not sufficiently covered by the CLEFT-Q eating and drinking, NOSE, and COHIP-OSS, and these questionnaires may not be accurate enough to stratify cleft-related outcomes.

Keywords: cleft lip and palate, ICHOM, patient-reported outcomes, psychometric performance, Rasch measurement theory.

VALUE HEALTH. 2020; ■(■):■–■

Introduction

Cleft lip and/or palate (CL/P) is the most prevalent congenital craniofacial anomaly affecting approximately 7.94 per 10 000 live births worldwide.[1,2] This complex disorder can negatively influence an individuals' appearance and psychosocial well-being, and cause functional disabilities such as problems with feeding, dentition, hearing, and speech.[3,4] Patients may need to undergo many surgical and nonsurgical procedures from infancy through young adulthood to improve physical and psychosocial function and well-being. To date, almost every cleft center has its own treatment protocol based on various literature and own experiences, resulting in differences in outcomes and quality of care.[5,6] Research into the psychosocial consequences of different treatment strategies for CL/P has been conducted without a uniform strategy.[7]

Traditionally, the success or failure of a cleft treatment has been evaluated and interpreted by clinicians[6,8,9]; however, clinician-reported outcomes fail to encompass the perspective of patients and their parents or caregivers, especially with regard to quality of life. In 2016, the Cleft Lip and Palate Working Group of the International Consortium for Health Outcomes Measurement (ICHOM) proposed a Standard Set of cleft-specific outcome measures for the comprehensive appraisal of cleft care. This set has been implemented over the past few years in several centers worldwide.[10-13] It stresses the importance of the patient's perspective by incorporating parent- and patient-reported outcome measures (PROMs). Specifically, the set includes 9

CLEFT-Q scales,[14-16] the Child Oral Health Impact Profile–Oral Symptoms Scale (COHIP-OSS),[17] the Nasal Obstruction Symptom Evaluation (NOSE) questionnaire,[18] and the parent-reported Intelligibility in Context Scale (ICS).[19] These instruments were chosen to cover core concepts of facial appearance, psychosocial function, speech, facial function (including eating/drinking and breathing), and oral health. Each of these conceptual domains should be assessed using clinically relevant, reliable, and valid scales to properly inform clinical decision making and to facilitate future comparative effectiveness research and quality-improvement projects.

In encouraging the adoption of any standardized outcomes-assessment framework, it is essential to verify that each of the included measures is robust enough to accurately and reliably appraise the corresponding conceptual construct or outcome domain. To that end, the aim of this study was to evaluate the psychometric performance of the patient- and parent-reported outcome measures in the ICHOM Standard Set for Cleft Care, such that we might gain insight into potential gaps of concept coverage.

Methods

Study Setting

De-identified CL/P outcome data were collected prospectively in clinical practice between November 2015 and April 2019 at Erasmus University Medical Center, Duke Children's Hospital, Boston Children's Hospital, and at international centers participating in the CLEFT-Q phase 3 study (Canada, United States, United Kingdom) led by researchers of McMaster University. The aim of the phase 3 study was to measure change in outcomes following 4 specific cleft-related procedures (alveolar bone grafting, secondary cleft lip revision, jaw surgery, and rhinoplasty). Research ethics approvals were obtained at the Institutional Review Board of each center.

Patient Population

All patients with orofacial clefts were eligible for data collection. They were all treated by a multidisciplinary cleft team. Cleft phenotypic categories were specified as follows: cleft lip (CL); cleft palate (CP); cleft lip and alveolus (CLA); and cleft lip and palate (CLAP). Outcomes were measured at time points defined by patient's age: T8 (range 8-9), T12 (range 10-12.5), and T22 (22 years or end of treatment, whichever is soonest).[10] Outcome data were collected electronically via home-based computer, an iPad at clinics, or paper and pencil and stored with REDCap[20,21] or Gemstracker,[11] dependent on the site's preferences (Supplemental Materials S1). All scales were administered in the native language of the country where each institution is located using approved translations of the instruments.

Patient-Reported Outcome Measures

The outcome measures assessed in this study include 9 patient-reported CLEFT-Q scales, the patient-reported COHIP-OSS and NOSE questionnaire, and the parent-reported ICS.

The CLEFT-Q is a rigorously developed, cleft-specific instrument focusing on 3 major domains: appearance, facial function, and health-related quality of life.[14-16] Each major domain was further broken down conceptually into subdomains, based on thematic content analysis of extensive focus groups and semi-structured interviews.[16] The CLEFT-Q face, jaws, teeth, psychological, school, social, speech function, and speech distress scales and the CLEFT-Q eating and drinking checklist were adopted as part of the ICHOM Standard Set. For the assessment of oral health, the Child Oral Health Impact Profile–Oral Symptoms Scale (COHIP-OSS) was included. The COHIP-OSS is a subscale of the larger COHIP, which was developed to measure various outcomes on oral health in school-aged children with different oral conditions, including CL/P.[17,22]

For assessing the quality of life related to nasal breathing, the Nasal Obstructive Symptom Evaluation (NOSE) questionnaire was adopted.[18,23] This questionnaire was developed to evaluate breathing outcomes of rhinoplasty and/or septoplasty treatment in adults.[24]

For speech, the Intelligibility in Context Scale (ICS) was developed to discriminate children with speech difficulties.[19] Because parents and family play an important role in representing the young patient with cleft, they were invited to complete the ICS by rating the degree of their children's intelligibility when speaking to various communication partners.

More information on the scales, including the core concepts measured, timing for completion, and example questions can be found in Table 1.

Statistical Analysis

Descriptive analyses were performed using SPSS software (IBM SPSS Statistics for Windows, Version 25.0, released 2017, IBM Corp). To provide insights into the performances of the PROMs, we applied Rasch measurement theory using RUMM 2030 software (RUMM version 2030, 1997-2020, RUMM Laboratory Pty Ltd) to our dataset with polytomous response options. Rasch analysis is a method that examines the extent to which the patient's responses match the predictions of the responses from the mathematical, logistic Rasch model. The difference between the expected and observed responses indicate the degree of rigorous measurement.[25-29] Within RUMM, we used the Partial-Credit Model, as this places no constraints on the threshold parameters. For this study, the following 4 keystones of Rasch measurement theory were assessed:

Targeting

The extent to which the distribution of the responses of the sample matches the range that can be measured by a specific scale is called targeting. Targeting is evaluated both graphically as with the percentage of the sample to score within the scale's range. When the sample is normally distributed and matches the construct as defined by the sample, a high percentage will be reached. A lower percentage corresponds with more mismatch and suggests that some patients' real ability cannot be determined with the scale.

Item-fit statistics

To evaluate whether responses are consistent with the expectations of the Rasch model, 3 fit indicators were examined: the χ^2 values (item-trait interaction), the log residuals (item-person interaction), and the item characteristic curves. The ideal fit residuals are between -2.5 and $+2.5$ with χ^2 values nonsignificant after Bonferroni adjustment. Inconsistent use of response options or multidimensionality can contribute to individual item misfit.

Thresholds for item response options

The thresholds between the response options of the scales were examined to determine whether they were used in an orderly fashion. Disordered thresholds can occur as a consequence of unclear definitions, too many response options, or underutilization of an option.[27]

Table 1. Overview of the patient- and parent-reported outcome measures in the ICHOM Standard Set for Cleft Care.

Concept	Scale	Number of items	Cleft phenotypes assessed	Time points (years)	Examples of questions	Response options
Facial appearance	CLEFT-Q face (patient-reported)	9	CL, CP, CLA, CLAP	8, 12, 22	How much do you like … – … how your face looks when you look your best? – … how your face looks when you smile?	Not at all, a little bit, quite a bit, very much
	CLEFT-Q teeth (patient-reported)	8	CL, CP, CLA, CLAP	8, 12, 22	How much do you like … – … the size of your teeth? – … how straight your teeth look?	Not at all, a little bit, quite a bit, very much
	CLEFT-Q jaws (patient-reported)	7	CL, CP, CLA, CLAP	12, 22	How much do you like … – … the size of your jaws? – … how your jaws look from the side?	Not at all, a little bit, quite a bit, very much
Psychosocial function	CLEFT-Q psychological (patient-reported)	10	CL, CP, CLA, CLAP	12	How do you feel? – I am happy with my life. – I feel confident.	Never, sometimes, often, always
	CLEFT-Q social (patient-reported)	10	CL, CP, CLA, CLAP	8, 22	How is your social life? – I have fun with friends. – I feel like I fit in.	Never, sometimes, often, always
	CLEFT-Q school (patient-reported)	10	CL, CP, CLA, CLAP	12	How is your school life? – I like seeing my friends at school. – I feel safe at school (not bullied).	Never, sometimes, often, always
Speech	CLEFT-Q speech distress (patient-reported)	10	CP, CLAP	12, 22	How do you feel about speaking? – I get teased about my speech. – I get upset when I need to repeat myself.	Always, sometimes, never
	CLEFT-Q speech function (patient-reported)	12	CP, CLAP	12, 22	How is your speech? – It is hard for my family to understand my speech. – I need to concentrate to speak well.	Always, sometimes, never
	Intelligibility in Context Scale (parent-reported)	7	CP, CLAP	12	Think about your child's speech intelligibility over the past month and identify the degree of understanding. – Do you understand your child? – Do immediate members of your family understand your child?	Never, rarely, sometimes, usually, always
Facial function	CLEFT-Q eating and drinking (patient-reported)	9	CP, CLA, CLAP	8, 12, 22	How is your eating and drinking? – Food falls out of my mouth when I eat. – Food or drinks go up my nose.	Always, often, sometimes, never
	NOSE (patient-reported)	5	CL, CP, CLA, CLAP	8, 12	How much of a problem were the following conditions for you? – Nasal blockage or obstruction. – Trouble breathing through my nose.	No problem, mild, moderate, fairly bad, severe problem

continued on next page

Table 1. Continued

Concept	Scale	Number of items	Cleft phenotypes assessed	Time points (years)	Examples of questions	Response options
Oral health	COHIP-OSS (patient-reported)	5	CP, CLA, CLAP	8, 12	In the past 3 months, have you … – … had pain in your teeth? – … had bleeding gums?	Never, almost never, sometimes, fairly often, almost all of the time

The measured core concepts including measurement instruments, number of items per scale, phenotypic groups, age groups for scale completion, and examples of questions with their response options are presented.
CL indicates cleft lip; CLA, cleft lip and alveolus; CLAP, cleft lip and palate; COHIP-OSS, Child Oral Health Impact Profile–Oral Symptoms Scale; CP, cleft palate; ICHOM, International Consortium for Health Outcomes Measurement; NOSE, Nasal Obstruction Symptom Evaluation.

Measurement precision

For each scale, the estimated measurement precision is given by the person separation index (PSI). Extreme values were withdrawn from the analyses. A higher PSI indicates higher reliability and a better discrimination among patients with different outcomes. A PSI of 0.7 is the lowest level of acceptability and is able to differentiate 3 groups.[30]

Although Rasch measurement theory may also be applied toward the exploration of differences in item functioning between centers or countries, we did not address differential item functioning (DIF) in this study. For the majority of the scales included in the Standard Set, DIF has been investigated before; DIF was examined in the CLEFT-Q international field-test study with 2434 patients from 10 countries. In 23 of the 110 CLEFT-Q items, DIF was identified by country, but was shown to have negligible impact on scoring.[14]

Results

A total of 714 unique patients with CL/P completed at least one of the scales (as appropriate based on cleft phenotype and age), resulting in 748 assessments available for analysis. In total, 60% (n = 425) of patients were found to have CLAP, and 55% (n = 391) of patients were male. Further demographics are presented in Table 2.

Results of the Rasch analyses are presented in Table 3. With regard to ***targeting***, the highest percentages of participants to score within the scales' measurement ranges were the CLEFT-Q teeth and the COHIP-OSS (both 92%). The CLEFT-Q jaws scale and the ICS were the least targeted (70% and 69%, respectively). This is depicted in Figure 1, where an example is given of the person-item threshold distribution for the ICS showing that the instrument's items did not cover the ability of persons at the higher end of the continuum.

Examination of ***item-fit statistics*** showed log residuals outside the ±2.5 for 13 of the 102 items for the entire set, from which 6 of these items had a significant χ^2 value. These items were all a marginal source of misfit with minor influence on the validity of the scale. None of the items in the set failed all 3 criteria for fit. In Table 4, an example of model fit evaluation with item-characteristic curves is given for the CLEFT-Q face scale for 2 items.

For ***thresholds for item response options***, 14 of the 102 items had disordered thresholds, including all 5 items of the COHIP-OSS. Figure 2 illustrates this phenomenon of disordered thresholds with a characteristics probability curve of one item of the COHIP-OSS. The figure shows that the middle response options are never the most likely to be selected by this population in this specific clinical setting. The NOSE questionnaire and ICS both showed similar results for one disordered item ("trouble sleeping" and "understood by parents," respectively). Rescoring the NOSE questionnaire and the COHIP-OSS by combining the middle scores resulted in better threshold ordering. The CLEFT-Q eating and drinking checklist showed 7 disordered items.

For ***measurement precision***, PSI values ranged from 0.82 to 0.88 for the CLEFT-Q scales, except for the speech distress scale (0.61) and eating and drinking (0.49). The analysis of the ICS revealed high reliability with a PSI value of 0.86. In contrast, the reliability scores for the NOSE and COHIP-OSS questionnaires were 0.35 and 0.43, respectively. This finding suggests that these scales were therefore not able to discriminate between patients with different qualities of nasal breathing and oral health.

Table 2. Patient characteristics.

Characteristics	Number of patients (%) Total N = 714
Cleft type	
Cleft lip only	51 (7)
Cleft palate only	165 (23)
Cleft lip and alveolus	73 (10)
Cleft lip and palate	425 (60)
Sex	
Male	391 (55)
Female	323 (45)
Sample	
Erasmus University Medical Center	362 (51)
Duke Children's Hospital	105 (15)
Boston Children's Hospital	95 (13)
CLEFT-Q phase 3 study	152 (21)

Time points	Number of measurements (%) Total N = 748
8 years (range 7-9)	379 (51)
12 years (range 10-13)	244 (32)
22 years (range 20-24)	125 (17)

Table 3. Scale performance statistics determined with Rasch measurement theory.

Scale	Sample size	Targeting (% within range)	Items outside ± 2.5	Number of χ^2 significant *P* values	Number of disordered thresholds	Person separation index
CLEFT-Q face	695	86	3	1	0	0.86
CLEFT-Q teeth	665	92	2	1	0	0.86
CLEFT-Q jaws	322	70	0	0	0	0.84
CLEFT-Q psychological	399	77	0	0	0	0.88
CLEFT-Q social	508	81	2	1	0	0.83
CLEFT-Q school	355	81	2	1	0	0.82
CLEFT-Q speech distress	257	76	0	0	0	0.61
CLEFT-Q speech function	274	81	1	0	0	0.83
Intelligibility in Context Scale	210	69	1	0	1	0.86
CLEFT-Q eating and drinking	501	74	1	1	7	0.49
NOSE	454	72	1	1	1	0.35
COHIP-OSS	426	92	0	0	5	0.43

COHIP-OSS indicates Child Oral Health Impact Profile–Oral Symptoms Scale; NOSE, Nasal Obstruction Symptom Evaluation.

Discussion

The ICHOM Cleft Lip and Palate Working Group acknowledged the importance of the patient perspective of health and included 12 patient- and parent-reported outcome scales in the ICHOM Standard Set for Cleft Care. These patient- and parent-reported instruments cover the core concepts of facial appearance, psychosocial function, speech, facial function (including eating/drinking and breathing), and oral health. The instruments were selected based on multiple criteria, including prior published evidence of instrument validation, clinical significance, practicality in implementation, availability, and translation into multiple languages. Although the instruments were previously subject to some degree of validity testing, they have not yet undergone robust psychometric evaluation after implementation in real-world clinical practice. Our study provides the first independent

Figure 1. Intelligibility in Context Scale person-item threshold distribution. This figure shows the targeting between the items, shown by the histogram in the lower half, and the patient sample, represented by the histogram in the upper half. At the lower end of the continuum the items are not covered by persons (*arrow 1*), whereas at +5 logit (*arrow 2*) and at the higher end of the continuum the persons are not covered by the items (*arrow 3*). This scale would benefit from including items that are more difficult.

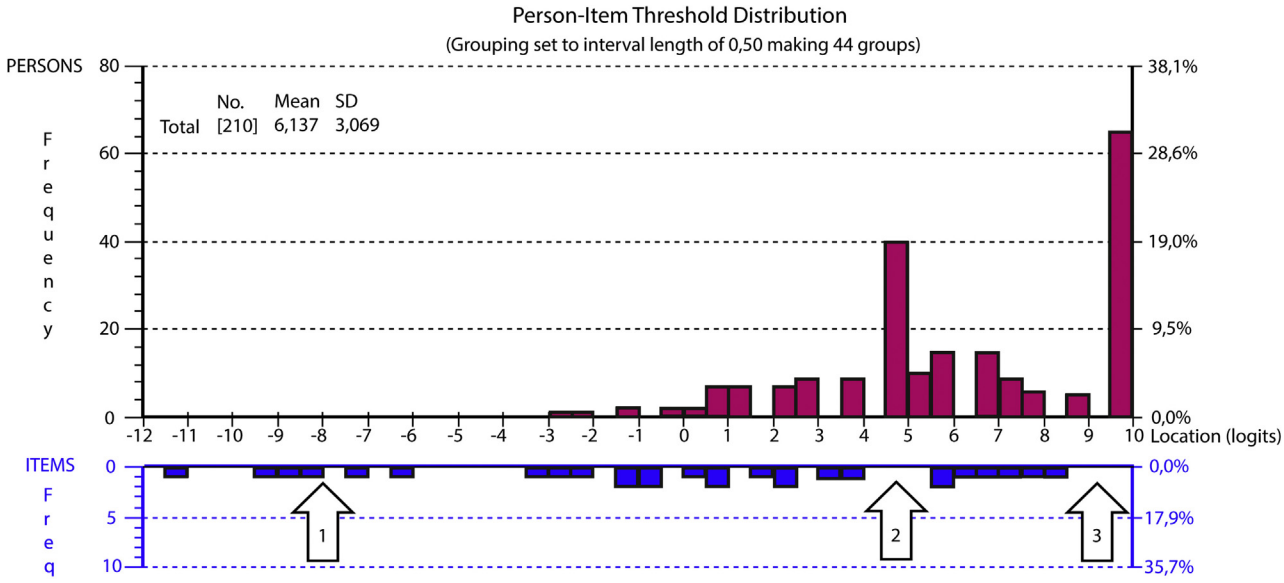
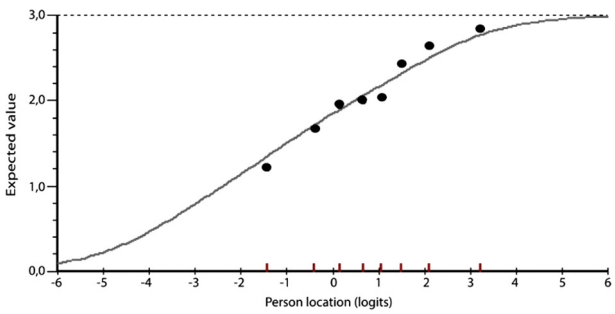
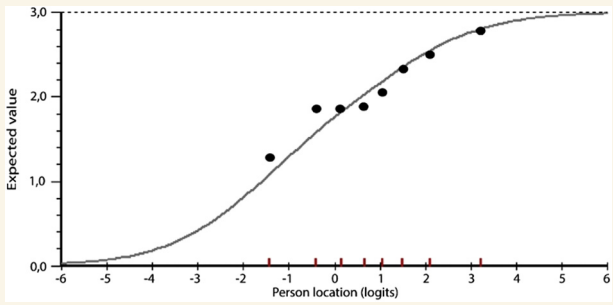


Table 4. Examination of item fit of 2 CLEFT-Q face scale items.

Item	Fit residual	χ^2	Item-characteristic curve	Interpretation
1	−2.640	22.34		Marginal overdiscrimination: the observed scores form a steeper curve than the expected scores. This item ("how your face looks when you look your best") is very similar to another item ("when you are ready to go out") and might therefore become redundant; however, this finding is not significant.
3	2.861*	25.25		Marginal underdiscrimination: the observed scores form a flatter curve than the expected scores. In clinical practice, a lot of patients consider this item ("how much do you like the shape of your face") as a more objective item in contrast with the other more subjective questions in the scale about "how you look"; however, the deviation is very mild and is not considered clinically relevant.

The observed values are represented by black dots and the expected values by the curve. High negative fit residuals are associated with redundancy or dependency of items and high positive fit residuals with multidimensionality.
*Significant *P* value.

evaluation of the psychometric performance of these instruments as used within the context of the ICHOM Standard Set for Cleft Care.

The Rasch analysis showed that the scales relating to the concepts of facial appearance, speech function, and psychosocial function worked properly with high reliability parameters.

Scales That Lacked Adequate Resolution at the Higher End of the Continuum

The CLEFT-Q speech distress scale, which was incorporated in the set for the evaluation of 12-year-old children and young-adult patients with CP or CLAP phenotypes, showed a slightly lower PSI value than the other CLEFT-Q scales. This is most likely caused by some mistargeting, because a lot of these patients have already completed intensive speech therapy and do not experience speech problems anymore. As a result, reliability of the scale is somewhat compromised without influencing the other psychometric characteristics.

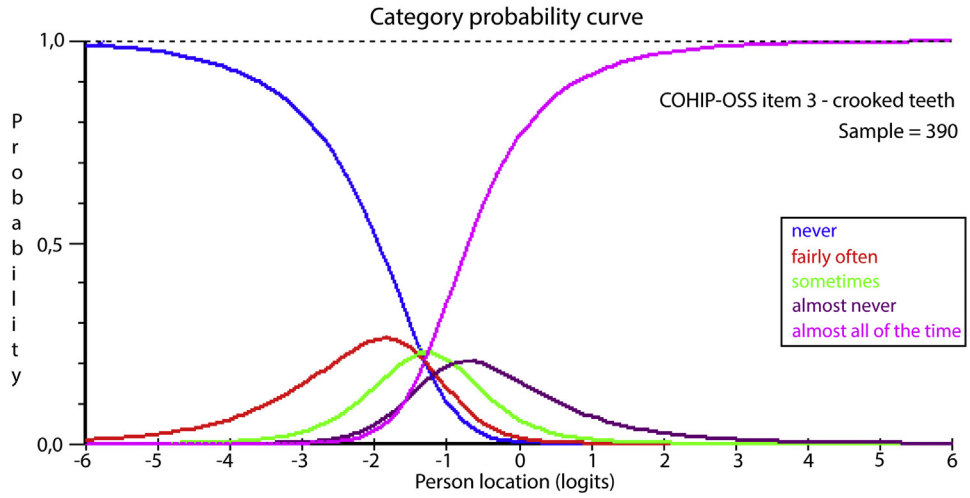
The 7-item ICS is included as a parent-reported outcome measure. It has previously been tested and validated in preschool aged children without cognitive or developmental disorders and has shown to be effective in discriminating children with speech difficulties.[19] In our study, the majority of patients scored high. As a result, a large group of patients is located at the upper extreme of the continuum, and these patients were not targeted by the scale items. This miscalibration of the scale range has the effect of impairing the possibility of accurately determining the patient's intelligibility in context, or of being sensitive to change after speech-related interventions such as revision palatoplasty, pharyngoplasty, or speech therapy. To improve the ICS, more items are needed at the higher end of the continuum.

Imbalanced Scales That Performed More Like Checklists

Facial function is covered by the CLEFT-Q eating and drinking checklist and the NOSE questionnaire. The developers of the CLEFT-Q previously reported that the reliability of the eating and drinking checklist was low (PSI < 0.60).[14] Our present study confirms these findings: most items in this questionnaire had disordered thresholds, which is why the creators of the CLEFT-Q emphasize the use of the term "checklist" rather than "scale."

Additionally, the NOSE questionnaire asks the patient how much of a problem some specific symptoms were for the patient over the past month, for example, "nasal blockage" or "trouble breathing through my nose."[18,24] This is the first evaluation of the psychometric properties of the NOSE questionnaire in children with CL/P and revealed disordered thresholds for the item "trouble sleeping." Prior assessments in adults corroborate that this item contributed least in terms of measuring the construct of the scale.[24] Anecdotally, cleft clinicians at Erasmus University Medical Center experienced that the phrasing of the NOSE questions was too difficult to understand for children of this young age; parents were often asked to explain what "obstruction of the nose" means or whether they have "trouble sleeping." According to the category probability curves and item-threshold distribution, most children with CL/P experienced no problems breathing through their nose and thus respond at the end of the scale. Experiencing no problems might be incorrect in these patients because they do not know otherwise in view of their congenital nature. A small number of children with severe

Figure 2. Category probability curve for item 3 (crooked teeth) of the COHIP-OSS showing disordered thresholds. The *x* axis represents the construct with increasing severity to the right. The *y* axis shows the probability of choosing the response categories. The middle categories were never the most likely to be selected. COHIP-OSS indicates Child Oral Health Impact Profile – Oral Symptoms Scale.



problems will score on the other end, whereas the middle options are not sensitive enough to measure small differences between patients. This finding was underlined by a very low PSI indicating no more than 2 groups can be discriminated with this questionnaire. A similar situation can be seen in a recent application of a modified NOSE questionnaire to investigate the prevalence of nasal obstruction symptoms in children with CL/P.[31] Modifications included a longer recall period of 12 months and questions and answers being rephrased from "problems" to "concerns." For the analysis of frequencies of NOSE scores and differences between cleft phenotypes, response categories were merged from 5 to 3. With these response options, differences in nasal obstruction severity between unilateral and bilateral CLAP patients were found. This shows that with a small modification of the NOSE questionnaire, discriminative value can be slightly increased to enhance clinical utility. Although this instrument might be useful as a screening tool or symptom checklist in clinical practice, we believe that the NOSE questionnaire in its current form is not sufficient as a pediatric PROM scale and suboptimal for the assessment of the young patient with cleft. In the same manner that the CLEFT-Q eating and drinking checklist is called a checklist rather than a scale, we would encourage that people refer to NOSE as a checklist rather than as a validated scale, as used in the pediatric cleft population.

This phenomenon of performing as a symptom screening tool, rather than a robust scale, also applies to the use of the COHIP-OSS for the assessment of oral health. This instrument measures the patient's view on oral health symptoms and was originally validated in a very heterogeneous sample of patients with diverse conditions affecting oral health, including patients with CL/P.[17,22] In our analysis of 8- and 12-year-old children with CL/P, the COHIP-OSS scale demonstrated low reliability, and all category thresholds were disordered. Most of the children responded at one end of the scale, reporting they "never" had any of the symptoms, except for the item "crooked teeth," which is most often scored as "almost all of the time." The latter can be explained by the fact that 8-year-old children are in mixed dentition, and orthodontic treatment is awaiting. The middle response options of the COHIP-OSS were hardly used. Our findings suggest that either there are too many irrelevant response options, or the 5 options are not distinctive enough. Although this scale has been tested and validated in school-aged children with different types of clefts, our study confirmed the necessity to test and validate measurement instruments when used in different populations and under altered circumstances, because measurement characteristics can differ.[32]

To Keep or to Discard? That Is the Question

With regard to the use of PROM data for future comparative effectiveness research, it is important to minimize measurement error on outcomes. Therefore, it should be taken into consideration whether the use of poorly validated, not well-understood instruments for children with CL/P, is sufficient enough for measuring the respective outcome domains. In a truly valid scale, all items should measure the same construct, resulting in a sum score that informs patients and healthcare professionals on the overall well-being of the patient regarding the specific construct measured by the scale. The final sum score of a scale can then be used for comparative effectiveness research. When a scale measures subtly different constructs, resulting in a checklist, every single item may be appraised as an independent entity with a separate score, but no overall sum score should be calculated. A checklist can still be relevant for clinical decision making, because individual elements can be intervened upon; however, because of its multidimensionality it is less suitable for outcome comparisons, such as comparing treatment techniques, protocols, or centers, as sum scores are not interpretable.[33,34]

An attempt to improve the performances of the CLEFT-Q eating and drinking checklist, COHIP-OSS questionnaire, and NOSE questionnaire by adding items or changing response options could be an option. On the other hand, it may be better to search for (or develop) a different scale that truly fits the concept. If the intended usage of these questionnaires is more akin to a screening tool than a diagnostic tool, then adding a quantitative measurement (eg, nasometry measurement for the assessment of the nasal airway) for the corroboration of poorly scoring children could be considered. If the intended usage of these questionnaires is for outcome comparisons, a conservative option is to remove these 3 checklists from the set. This will reduce burden on patients and will allow the clinicians to focus on the most useful PROMs.

Strengths and Limitations

Because the ICHOM Standard Set is meant to be measured worldwide, a strength of this study is the international cohort of patients with CL/P resulting in a reflection of the cleft population that is eligible for completing the ICHOM Standard Set; however, a limitation of our study is that low-income countries were not represented in this cohort. Additionally, as a result of the clinical transition phase of implementing the set, some 7-year-old children were asked to complete one or more of the outcome questionnaires, resulting in a slightly broader age range than advised by the ICHOM Reference Guide (age range 8-9).[10] The ages of eligibility were set at 8, because it is known that children as young as 8 years are able to report on well-being and psychosocial health[35,36]; however, given the small number of 7-year-old children included in this study and the large total sample size, we do not expect to find different results. Furthermore, we believe that including these patients in our sample gives a good reflection of daily clinical practice.

Conclusions

To improve patient-centered care and to facilitate future comparative effectiveness research and quality-improvement endeavors, it is important to include clinically meaningful and scientifically sound measurement instruments in an outcome set. This study found that most of the patient- and parent-reported components recommended by the ICHOM Standard Set for Cleft Care are valid tools for assessing cleft-specific outcomes. Importantly, the CLEFT-Q eating and drinking checklist, the COHIP-OSS, and NOSE questionnaire were not found to be robust enough for outcomes comparisons, and instead work like a checklist rather than a measurement scale. As a result, the concepts of facial function (including eating/drinking and breathing) and oral health are not sufficiently covered by the PROMs included in the ICHOM Standard Set for Cleft Care.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https: https://doi.org/10.1016/j.jval.2020.10.019.

Article and Author Information

Accepted for Publication: October 9, 2020

Published Online: xxxx

doi: https://doi.org/10.1016/j.jval.2020.10.019

Author Affiliations: Department of Oral and Maxillofacial Surgery, Erasmus University Medical Center, Rotterdam, The Netherlands (Apon, Koudstaal, Wolvius); Department of Public Health, Erasmus University Medical Center, Rotterdam, The Netherlands (Leeuwen); Department of Plastic, Maxillofacial and Oral Surgery, Duke Children's Hospital, Durham, North Carolina, USA (Allori); Department of Plastic and Oral Surgery, Boston Children's Hospital, Boston, Massachusetts, USA (Rogers-Vizena); Department of Craniofacial Surgery, Karolinska University Hospital, Stockholm, Sweden (Koudstaal); Modus Outcomes, Letchworth Garden City, United Kingdom (Cano); Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada (Klassen); Department of Plastic and Reconstructive Surgery, Erasmus University Medical Center, Rotterdam, The Netherlands (Versnel).

Address correspondence to: Inge Apon, MD, Department of Oral and Maxillofacial Surgery, Erasmus University Medical Center, NA-building, Room 2401, Postbus 2040, 3000 CA Rotterdam, The Netherlands. Email: i.apon@erasmusmc.nl

Author Contributions: *Concept and design*: Apon, Allori, Wolvius, Klassen, Versnel.

Acquisition of data: Apon, Allori, Rogers-Vizena, Koudstaal, Klassen, Versnel.

Analysis and interpretation of data: Apon, Rogers-Vizena, Wolvius, Cano, Klassen, Versnel.

Drafting of the manuscript: Apon, van Leeuwen, Allori, Rogers-Vizena, Cano, Klassen, Versnel.

Critical revision of the paper for important intellectual content: van Leeuwen, Allori, Rogers-Vizena, Koudstaal, Wolvius, Cano, Klassen, Versnel.

Statistical analysis: Apon, Cano, Klassen.

Provision of study materials or patients: Koudstaal, Klassen, Versnel.

Obtaining funding: Koudstaal, Versnel.

Administrative, technical, or logistic support: van Leeuwen.

Supervision: van Leeuwen, Koudstaal, Cano, Klassen, Versnel.

Conflict of Interest Disclosures: Dr Cano is co-owner of Modus Outcomes and reported having a patent BODY-Q copyright will receive a share of any license revenues as royalties based on the inventor sharing policy. Dr Klassen reported receiving grants from the Canadian Institutes of Health Research during the conduct of the study, and personal fees from Allergan outside the submitted work. In addition, Dr Klassen reported having a patent CLEFT-Q and copyright is pending. No other disclosures were reported.

Funding/Support: Dr Klassen received a financial grant of the Canadian Institutes of Health Research to support this research.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgement: We are grateful for all the help provided by Jet de Gier, Mona Haj, Anna Miroshnychenko, Charlene Rae, Karl Sanchez, Linda van der Sluijs, and Mariska van Veen-van der Hoek. We are thankful for the support provided by the Canadian Institutes of Health Research.

REFERENCES

1. Tanaka SA, Mahabir RC, Jupiter DC, Menezes JM. Updating the epidemiology of cleft lip with or without cleft palate. *Plast Reconstr Surg*. 2012;129(3):511e–518e.
2. International Perinatal Database of Typical Oral Clefts Working Group. Prevalence at birth of cleft lip with or without cleft palate: data from the International Perinatal Database of Typical Oral Clefts. *Cleft Palate Craniofac J*. 2011;48(1):66–81.
3. Fadeyibi IO, Coker OA, Zacchariah MP, Fasawe A, Ademiluyi SA. Psychosocial effects of cleft lip and palate on Nigerians: the Ikeja-Lagos experience. *J Plast Surg Hand Surg*. 2012;46(1):13–18.
4. Kirschner RE, LaRossa D. Cleft lip and palate. *Otolaryngol Clin North Am*. 2000;33(6):1191–1215. v-vi.
5. Bearn D, Mildinhall S, Murphy T, et al. Cleft lip and palate care in the United Kingdom – the Clinical Standards Advisory Group (CSAG) Study. Part 4: outcome comparisons, training, and conclusions. *Cleft Palate Craniofac J*. 2001;38(1):38–43.
6. Shaw WC, Semb G, Nelson P, et al. The Eurocleft project 1996-2000: overview. *J Craniomaxillofac Surg*. 2001;29(3):131–140. discussion 141-132.
7. Stock NM, Feragen KB. Psychological adjustment to cleft lip and/or palate: a narrative review of the literature. *Psychol Health*. 2016;31(7):777–813.
8. Russell K, Long Jr RE, Hathaway R, et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 5. General discussion and conclusions. *Cleft Palate Craniofac J*. 2011;48(3):265–270.
9. Rautio J, Andersen M, Bolund S, et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 2. Surgical results. *J Plast Surg Hand Surg*. 2017;51(1):14–20.
10. International Consortium of Health Outcomes Measurement (ICHOM). Data collection reference guide. https://ichom.org/files/medical-conditions/cleft-lip-palate/cleft-lip-palate-reference-guide.pdf.
11. Arora J, Haj M. Implementing ICHOM's Standard Sets of Outcomes: Cleft Lip and Palate at Erasmus University Medical Centre in the Netherlands. London: International Consortium for Health Outcomes Measurement (ICHOM), December 2016. www.ichom.org.
12. Porter ME. A strategy for health care reform – toward a value-based system. *N Engl J Med*. 2009;361(2):109–112.
13. Allori AC, Kelley T, Meara JG, et al. A standard set of outcome measures for the comprehensive appraisal of cleft care. *Cleft Palate Craniofac J*. 2017;54(5):540–554.
14. Klassen AF, Wong Riff KW, Longmire NM, et al. Psychometric findings and normative values for the CLEFT-Q based on 2434 children and young adult patients with cleft lip and/or palate from 12 countries. *CMAJ*. 2018;190(15):E455–E462.

15. Tsangaris E, Wong Riff KWY, Goodacre T, et al. Establishing content validity of the CLEFT-Q: a new patient-reported outcome instrument for cleft lip/palate. *Plast Reconstr Surg Glob Open*. 2017;5(4):e1305.
16. 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(1):e015467.
17. 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.
18. Stewart MG, Witsell DL, Smith TL, Weaver EM, Yueh B, Hannley MT. Development and validation of the Nasal Obstruction Symptom Evaluation (NOSE) scale. *Otolaryngol Head Neck Surg*. 2004;130(2):157–163.
19. McLeod S, Harrison LJ, McCormack J. The intelligibility in Context Scale: validity and reliability of a subjective rating measure. *J Speech Lang Hear Res*. 2012;55(2):648–656.
20. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208.
21. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–381.
22. Broder HL, Wilson-Genderson M, Sischo L. Reliability and validity testing for the Child Oral Health Impact Profile-Reduced (COHIP-SF 19). *J Public Health Dent*. 2012;72(4):302–312.
23. Zhang RS, Lin LO, Hoppe IC, et al. Nasal obstruction in children with cleft lip and palate: results of a cross-sectional study utilizing the NOSE scale. *Cleft Palate Craniofac J*. 2019;56(2):177–186.
24. van Zijl F, Timman R, Datema FR. Adaptation and validation of the Dutch version of the nasal obstruction symptom evaluation (NOSE) scale. *Eur Arch Otorhinolaryngol*. 2017;274(6):2469–2476.
25. Hobart J. Improving the evaluation of therapeutic intervention in MS: the role of new psychometric methods. *Health Technol Assess*. 2009;13(iii, ix-x):1–177.
26. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Vol. 1 of Studies in Mathematical Psychology. Copenhagen: Danmarks Paedagogiske Institut; 1960.
27. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol*. 2007;46(Pt 1):1–18.
28. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. *Int J Nurs Stud*. 2009;46(3):380–393.
29. Cano SJ, Hobart JC. The problem with health measurement. *Patient Prefer Adherence*. 2011;5:279–290.
30. Fisher Jr W. Reliability, separation, strata statistics. *Rasch Measurement Transactions*. 1992;6(3):238.
31. Sobol DL, Allori AC, Carlson AR, et al. Nasal airway dysfunction in children with cleft lip and cleft palate: results of a cross-sectional population-based study, with anatomical and surgical considerations. *Plast Reconstr Surg*. 2016;138(6):1275–1285.
32. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27(5):1159–1170.
33. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–1157.
34. Carle AC, Weech-Maldonado R. Validly interpreting patients' reports: using bifactor and multidimensional models to determine whether surveys and scales measure one or more constructs. *Med Care*. 2012;50(9 Suppl 2):S42–S48.
35. 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.
36. Bevans KB, Riley AW, Moon J, Forrest CB. Conceptual and methodological advances in child-reported outcomes measurement. *Expert Rev Pharmacoecon Outcomes Res*. 2010;10(4):385–396.