RESEARCH

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

Anne F. Klassen DPhil, Karen WY Wong Riff MD MSc, Natasha M. Longmire MSc, Asteria Albert MD PhD, Gregory C. Allen MD, Mustafa Asim Aydin MD, Stephen B. Baker MD DDS, Stefan J. Cano PhD CPsychol, Andrew J. Chan MD BDS, Douglas J. Courtemanche MD MS, Marieke M. Dreise BA, Jesse A. Goldstein MD, Timothy E.E. Goodacre MB BS FRCS, Karen E. Harman MD, Montserrat Munill MD DDS, Aisling O. Mahony MA (Oral Biology) DDS, Mirta Palomares Aguilera BSc, Petra Peterson MD, Andrea L. Pusic MD MPH, Rona Slator FRCS(Plast) DPhil, Mia Stiernman MD, Elena Tsangaris PhD, Sunil S. Tholpady MD PhD, Federico Vargas MD, Christopher R. Forrest MD MSc

■ Cite as: *CMAJ* 2018 April 16;190:E455-62. doi: 10.1503/cmaj.170289

Visual abstract available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.170289/-/DC2

ABSTRACT

BACKGROUND: Patients with cleft lip and/or palate can undergo numerous procedures to improve appearance, speech, dentition and hearing. We developed a cleft-specific patient-reported outcome instrument to facilitate rigorous international measurement and benchmarking.

METHODS: Data were collected from patients aged 8–29 years with cleft lip and/or palate at 30 hospitals in 12 countries between October 2014 and November 2016. Rasch measurement theory analysis was used to refine the scales and to examine reliability and validity. Normative CLEFT-Q values were computed for age, sex and cleft type.

RESULTS: Analysis led to the refinement of an eating and drinking checklist and 12 scales measuring appearance (of the face, nose, nostrils, teeth, lips, jaws and cleft lip scar), health-related quality of life (psychological, social, school, speech distress) and speech function. All scales met the requirements of the Rasch model. Analysis to explore differential item functioning by age, sex and country provided evidence to support the use of a common scoring algorithm for each scale for international use. Lower (worse) scores on CLEFT-Q scales were associated with having a speech problem, being unhappy with facial appearance, and needing future cleft-related treatments, providing evidence of construct validity. Normative values for age, sex and cleft type showed poorer outcomes associated with older age, female sex and having a visible cleft.

INTERPRETATION: The CLEFT-Q represents a rigorously developed instrument that can be used internationally to collect and compare evidence-based outcomes data from patients aged 8–29 years of age with cleft lip and/or palate.

Cleft lip and/or palate is the most common congenital craniofacial abnormality, with 7.9 clefts per 10 000 live births annually.[1] This condition presents with an abnormal connection between the mouth and the nose that can be visible on the lip, or invisible, affecting the palate alone. Patients can undergo numerous procedures to improve appearance, speech, dentition and hearing. Within and between countries, there is substantial variation in how surgeries are carried out.[2–4] For example, the landmark Eurocleft study[2,5–8] identified that the 201 participating centres used 194 different protocols for treatment of just 1 cleft subtype.

The goal of treatment for cleft lip and/or palate is to improve appearance, speech and psychosocial function, but measurement of outcomes has used mainly objective measures.[2–8] Inclusion of the patient perspective using a cleft-specific patient-reported outcome instrument could increase understanding of

© 2018 Joule Inc. or its licensors

patient concerns. Patient-reported outcome instruments measure outcomes that matter to patients, from their perspective.[9] Such tools are increasingly used worldwide to inform patient care, as quality metrics, as audit tools and in research.[10–16] Some countries (e.g., the United Kingdom, Netherlands, Sweden) measure patient-reported outcomes at the national level to compare providers or in clinical registries.[10–14]

To address the call for a cleft-specific patient-reported outcome instrument,[17] our team developed the CLEFT-Q. Our goal was to develop a common metric that could be used internationally and that works the same across age. The optimal design for an internationally applicable instrument would be for development to take place cross-culturally from the outset. Such an instrument could allow for rigorous measurement of the burden of cleft lip and/or palate in both high-resource and low-resource settings. Ensuring an instrument works the same for patients of different ages requires a mixed-methods approach with extensive qualitative research to identify common concerns for scale development, followed by advanced statistical techniques to identify any items biased by age. Full protocol details have been published.[18] Content was developed from interviews with 138 patients from 6 countries.[19] Findings were used to develop a conceptual framework comprising 3 domains; i.e., appearance, health-related quality of life and facial function. Cognitive interviews with 69 patients from 6 countries, and feedback from an international sample of 44 experts, provided input used to refine 13 scales and establish their content validity.[20] The scales were then translated and culturally adapted into 5 languages[21,22] following international guidelines.[23,24]

The aim of this paper is to describe reliability and validity findings based on a set of statistical and graphical tests, including that of construct validity, whereby CLEFT-Q scores were hypothesized to be worse for participants who needed (v. did not need) cleft-related treatments, who were less (v. more) happy with their appearance and who had a more (v. less) severe speech problem. We also present CLEFT-Q normative values for age, sex and cleft type.

Methods

Study participants

Participants were aged 8–29 years with cleft lip, cleft palate, cleft lip and palate or cleft lip and alveolus, and able to complete the CLEFT-Q independently. Participants were recruited from craniofacial clinics in 30 hospitals located in 12 countries.

Data collection

Participants were recruited between October 2014 and November 2016. Data were collected either face to face or by mail using electronic devices (tablets) or paper and pencil (booklets) based on each site's logistics and preference. The appearance scales asked respondents how much (not at all, a little, quite a bit, very much) they like the appearance of their face, nose, nostrils, teeth, lips, jaws and cleft lip scar, and to answer thinking of how they look now. Health-related quality of life (psychological, school, social, speech distress) and facial function (speech, and eating and drinking) scales ask respondents how often (never, sometimes, often, always) a set of statements apply to them, and to answer based on the past week. Participants completed only relevant scales (e.g., jaws, for participants aged 12–29 yr; cleft lip scar, for anyone with cleft lip; and school, for participants aged 8–18 yr and attending school). Site staff completed a clinical form that asked about cleft type, speech and current or future treatment needs. We entered data into REDCap,[25] hosted at McMaster University (Canada).

Statistical analysis

We used SPSS (IBM SPSS Statistics, Version 24, IBM Corp) and RUMM2030[26] software for data analysis. In Rasch analysis (described in detail elsewhere[27]), individuals are placed along a scale based on the probability that they will answer a set of items in a certain manner.[28]

We based decisions about items on a set of statistical and graphical tests.[27,29,30] We amended the sample to 500 for tests of fit statistics.[27] We conducted the analysis using the entire sample, except for the 2 speech scales (participants without a speech problem were excluded) and eating and drinking (participants with cleft lip only were excluded).

Thresholds for item response options

We examined thresholds between response options (e.g., "not at all" and "a little") to determine if a scale's response categories were ordered, meaning that a "1" on a 4-point scale must sit lower in the continuum than a "2," and so on.[31] On detecting disordered thresholds, we recoded the scores.

Item fit statistics

Three fit indicators were examined: log residuals (item–person interaction), X^2 values (item–trait interaction) and item characteristic curves.[22] Ideal fit residuals are between –2.5 and +2.5 with X^2 values nonsignificant after Bonferroni adjustment. We interpreted fit statistics together and in relation to clinical usefulness.

Dependency

We inspected residual correlations between pairs of items to identify any that were greater than 0.30, as high residual correlations can inflate reliability. For any residual correlations greater than 0.30, we carried out a subtest to determine impact on reliability.[29]

Targeting

We examined item locations to determine whether they were evenly spread over a reasonable range that matched the range of the construct reported by the sample.[27]

Differential item functioning

We used this test to determine if individuals in subgroups (i.e., country, sex, age [8–10, 11–13, 14–17, 18–29 yr]) responded differently to items despite the same measured trait level. We chose random samples to create equal-sized subgroups, with a proviso that there had to be a minimum of 200 to provide 50 for 4 class intervals.[30] The number of countries in the analysis by scale was as follows: jaws = 2; cleft lip scar and school = 3; teeth and psychological = 5; face, lips, nose, nostrils and social = 6. For age, we

divided the sample into 4 groups to ensure there were at least 50 participants per class interval in the differential item functioning analysis. We conducted the analyses with and without adjusting the overall sample to 500.[27] Items with X^2 values significant after Bonferroni adjustment were split on the variable that evidenced differential item functioning, and the new and original person locations were correlated to examine any impact on scoring.[27]

Person separation index

This statistics measures error associated with the measurement of people in a sample. Higher values show greater reliability.[32]

In addition to Rasch analyses, we computed Cronbach α.[33] To examine construct validity,[34] Rasch logit scores were transformed into 0 (worse) to 100 (best). We hypothesized that participants who may need (v. not need) cleft-related treatments, or were receiving (v. not receiving) psychological therapy to cope with cleft lip and/or palate, would report lower scores on the scales designed to evaluate these treatments. We asked participants: "How happy are you about how you look (your appearance)?" Four options (very unhappy, a little unhappy, a little happy, very happy) were provided. We hypothesized that unhappy participants (v. happy) would report lower scores on appearance and health-related quality of life scales. Third, we hypothesized that participants with a mild or moderate-to-severe speech problem (v. no problem) would report lower scores on speech and scales. Mean CLEFT-Q scores by age, sex and cleft type provided normative values.

As we found data for subgroups to be normally distributed, we used parametric (*t* test or one-way analysis of variance [ANOVA]) tests to examine mean differences. We used p values < 0.05 to identify statistical significance.

Ethics approval

Ethics board approval was obtained at 30 hospitals in 12 countries; the first board to grant approval was the Hamilton Integrated Research Ethics Board in Canada.

Results

Table 1 shows sample characteristics. The proportion of participants with a syndrome and/or other craniofacial anomaly was 10.9%.

Participants (n = 2434) in our convenience sample were recruited with different approaches: 23 sites used face-to-face recruitment and 7 sites included a mailout. Nonparticipants were tracked by 13 sites: 8 (using face-to-face) recruited 1096 patients and had 144 decline, 88.4% response rate; and 5 (using mailout) recruited 422 patients and had 872 decline, 32.6% response rate.

Psychometric findings

Rasch analysis provided evidence of reliability and validity for 12 of 13 scales. The eating and drinking scale proved an exception. Eight items had disordered thresholds. After we rescored across the 2 middle response options, the item fit statistics were acceptable, but overall model fit was significant (p = 0.02) and the person separation index was low (< 0.60). Used as a problem checklist rather than a scale, the proportion of participants to report a

Table 1: Characteristics of participants in the CLEFT-Q study

Characteristic	No. of participants (%) n = 2434
Country	
Canada	624 (25.6)
United States	362 (14.9)
England	339 (14.0)
India	232 (9.5)
Colombia	210 (8.6)
Netherlands	206 (8.5)
Ireland	100 (4.1)
Sweden	100 (4.1)
Spain	93 (3.8)
Chile	89 (3.7)
Turkey	54 (2.2)
Australia	25 (1.0)
Language	
English	1450 (59.6)
Spanish	392 (16.1)
Hindi	232 (9.5)
Dutch	206 (8.5)
Swedish	100 (4.1)
Turkish	54 (2.2)
Age, yr	
8–9	426 (17.5)
10–11	411 (16.9)
12–13	372 (15.3)
14–15	385 (15.8)
16–17	293 (12.0)
18–20	300 (12.3)
≥ 21	246 (10.1)
Missing	1 (0.1)
Sex	
Male	1351 (55.5)
Female	1081 (44.4)
Missing	2 (0.1)
Student	
Yes	2149 (88.3)
No	274 (11.3)
Missing	11 (0.4)
Type of schooling	
School regular	1834 (85.3)
School home	35 (1.6)
College or university	248 (11.5)
Other	24 (1.1)
Missing	8 (0.4)
Cleft type	
Cleft lip and palate	1399 (57.5)
Cleft palate	568 (23.3)
Cleft lip	263 (10.8)
Cleft lip and alveolus	204 (8.4)
Current speech problem	
No	1271 (52.2)
Mild	808 (33.2)
Moderate	270 (11.1)
Severe	32 (1.3)

problem was as follows: biting into some foods (42%), some foods hard to chew (37.4%), food or drink going up the nose (36.6%), having to eat slowly (36.1%), having to take small bites (32.7%), food getting stuck in hole in the mouth (28.1%), having to avoid certain foods (25.8%), trouble drinking through a straw (18.6%), and food falling out of the mouth (14%).

Rasch analysis reduced the CLEFT-Q by 35 items to 110. Thresholds were disordered for 4 items in speech distress and

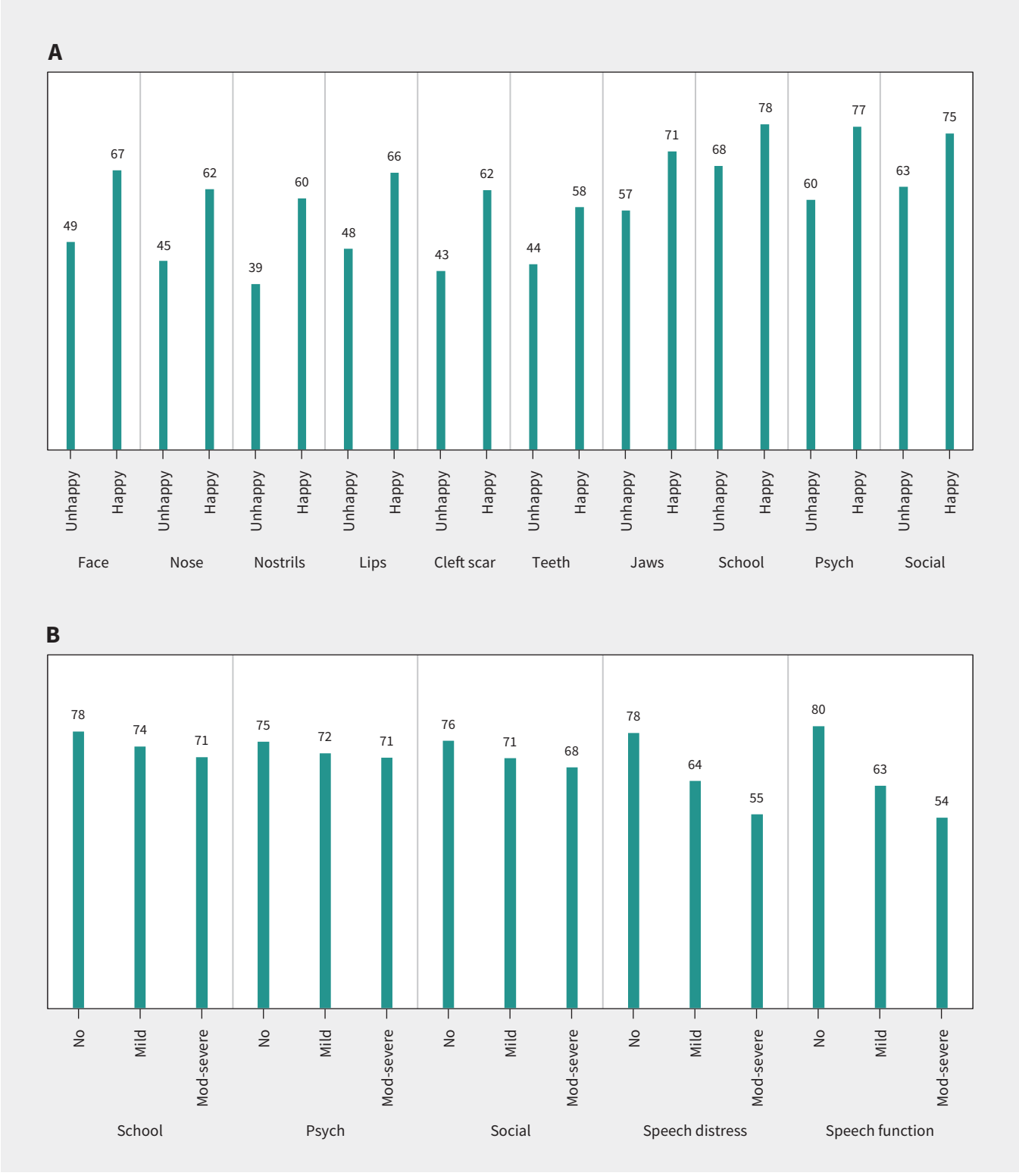


Figure 1: A) Mean scores for CLEFT-Q appearance and health-related quality of life scales based on answer to question: "How happy are you about how you look (your appearance)?" B) Mean scores for CLEFT-Q speech and health-related quality of life scales by current speech problem. Note: Mod = moderate, psych = psychological.

2 items in speech function. When we rescored all items across the 2 middle response options, both scales had ordered thresholds. Subsequent Rasch analysis used the rescored data.

All 110 items had nonsignificant X^2 p values after Bonferroni adjustment (Appendix 1, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.170289/-/DC1). Item fit was within ± 2.5 for 58/110 items. The item residual correlations were greater than 0.30 for 2 pairs of items (r = 0.43 speech distress; r = 0.32 speech function). Subtests showed that the correlated items had marginal impact on scale reliability (< 0.01 drop in the person separation index).

Appendix 2 (available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.170289/-/DC1) shows the distribution of person measurement and item locations for 3 scales to illustrate targeting. These figures show that most participants scored inside the range for which the scales provide measurement. Participants who scored outside the range of the scales tended to be those who "very much" liked how their face looks overall (Figures 1A and 1C), or did not have a speech problem (Figures 1B and 1D).

Appendix 1 shows findings for differential item functioning, which was evident for 23 items by country, 25 items by age group and 10 items for sex. When we amended the sample to 500, differential item functioning was evident for 9 items by country and 4 items by age group. When we split items that were significant in the unadjusted analysis by the relevant patient characteristic, Pearson correlations between the original and split person locations showed that differential item functioning had a negligible impact, with all correlations ≥ 0.99.

All 12 scales satisfied the requirements of the Rasch model, with nonsignificant X^2 values (Appendix 3, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.170289/-/DC1). Person separation index values were ≥ 0.85 for 10/12 scales (with extremes) and 9/12 (without extremes). Cronbach α values ranged from 0.89 to 0.96. Scale-level missing data ranged from 9% to 2.2%. The Flesch–Kincaid readability statistics[35] for CLEFT-Q items were 0 to 5.2, with 115/119 items at grade 3 reading level or lower.

The number of participants judged to need cleft-related treatments was as follows: orthodontic (n = 860, 35.3%), rhinoplasty (n = 743, 30.5%), speech therapy or surgery (n = 491, 20.2%), orthognathic surgery (n = 401, 16.5%), lip revision (n = 380, 15.6%) and alveolar bone graft (n = 253, 10.4%). The mean scores for the CLEFT-Q scales designed to evaluate outcomes for these treatments were lower for participants who needed (v. did not need) treatment (p < 0.001 on all independent samples t tests). The mean differences in scores were as follows: orthodontic: teeth scale –7.9 (1.0; lower [worse] for those who needed treatment); rhinoplasty: nose scale –11.7 (1.0), nostrils scale –15.9 (1.3); speech therapy or surgery: speech distress scale –14.5 (1.1); speech function scale –15.0 (1.1); orthognathic surgery: jaws scale –15.3 (1.8); lip scar

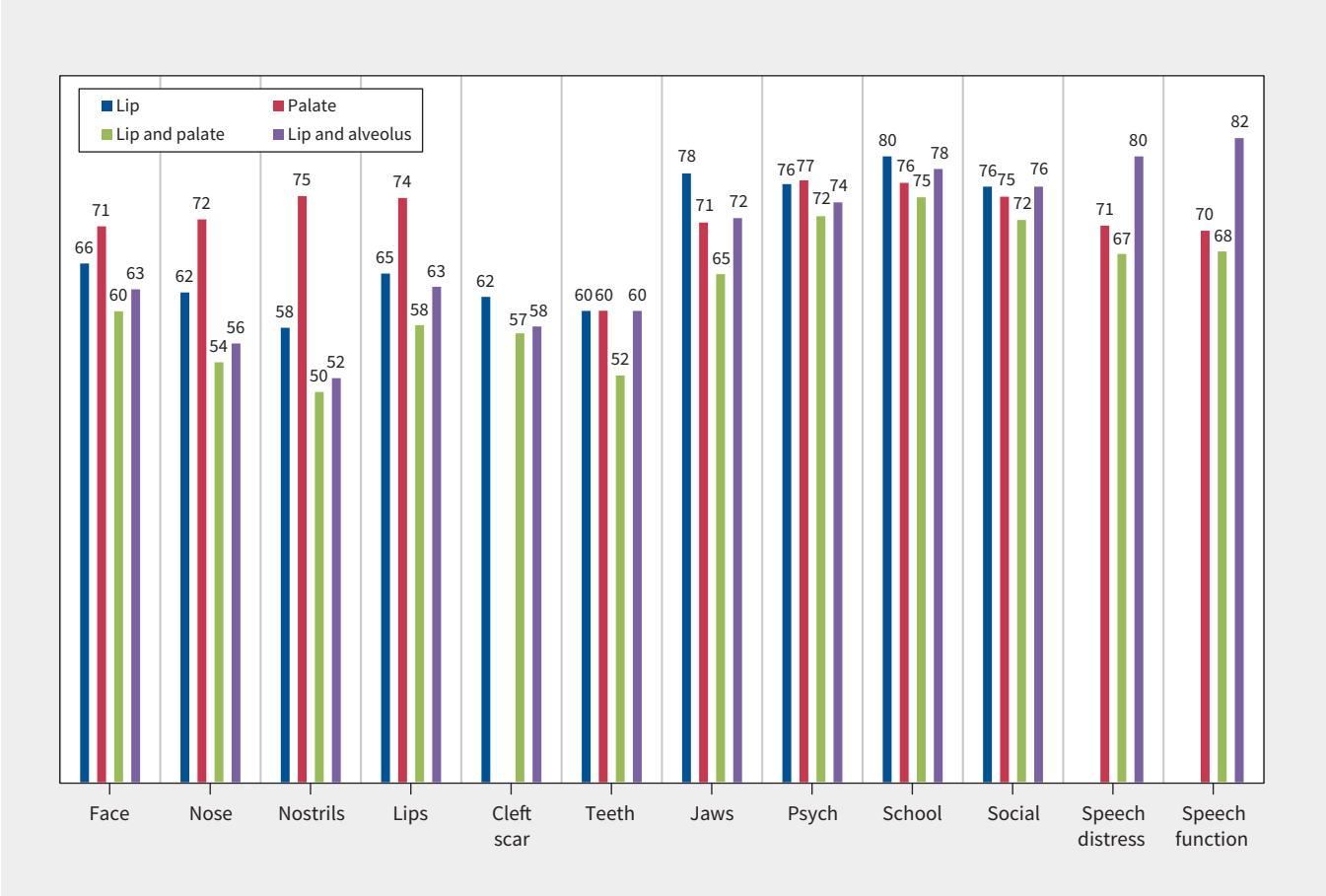


Figure 2: Mean scores for CLEFT-Q by cleft type. Note: Psych = psychological.

revision: lips scale −11.9 (1.3), cleft lip scar scale −10.4 (1.7); and alveolar bone graft: teeth scale −7.2 (1.6). Participants (n = 276) who were receiving psychological therapy (v. no therapy) reported lower scores on the school (−4.3 [1.6], p < 0.01) and social (−4.0 [1.2], p < 0.01) scales.

Participants who were unhappy (n = 381) versus happy (n = 1663) with how they looked reported lower scores (p < 0.001 on independent samples t tests) on all appearance and health-related quality of life scales (Figure 1A). Mean scores for health-related quality of life and speech scales (Figure 1B) were highest for the group without a speech problem and lowest for the group with a moderate or severe speech problem (p < 0.001 on ANOVA).

Normative values

Appendix 4 (available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.170289/-/DC1) shows the normative values. By age group, mean scores for 10/12 scales (exceptions were teeth and school) differed significantly. For most appearance scales, psychological and social scores were higher for younger participants and lower for older participants. By sex, differences in mean scores were significant for lips, cleft scar, jaws, psychological, social and speech function; male participants reported better scores on 5 of the 6 scales. By cleft type (Figure 2), the difference in mean scores was significant for 11/12 scales (the exception was cleft lip scar). Participants with an invisible cleft reported higher (better) scores for face, nose, nostrils and lips scales, and participants with no palate involvement reported higher scores for teeth and jaws. On all scales, the cleft lip and palate group had the lowest mean scores.

Interpretation

The CLEFT-Q represents a rigorously developed patient-reported outcome instrument for patients aged 8 to 29 years. The psychometric analyses showed that 12/13 scales evidenced reliability and validity and measured a clinical hierarchy for each concept. Importantly, the age, sex and language differential item functioning analysis provided evidence for the use of a common scoring algorithm for each scale. The eating and drinking scale represented an exception. For this set of items, there was no clinical hierarchy and the items instead form a problem checklist. Findings of a poor functioning scale highlight the usefulness of modern psychometric methods for diagnosing problems within scales. As patient-reported outcome instruments are increasingly used worldwide to inform clinical practice, comparative effectiveness research and quality improvement initiatives,[10–16] it is crucial that only clinically meaningful and scientifically sound instruments are used.

Comparisons with other studies

Research shows that the overall impact of cleft lip and/or palate on psychosocial adjustment is low.[36] Our study provides findings about factors associated with poorer outcomes including cleft-related treatments, age, sex and cleft type. The finding of lower scores in those who need cleft-related surgery is in line with a study of 1200 youth with cleft lip and/or palate, which showed that surgical candidates reported lower health-related quality of life than nonsurgical candidates.[37] Our findings are also in agreement with Ranganathan and colleagues,[38] who found lower health-related quality of life in 71 children who wanted cleft revision surgery.

Limitations

Patients with syndromes and other craniofacial anomalies were underrepresented in our sample,[39] and patients younger than 8 years were excluded. Guidelines suggest 8 years of age is the lower limit for valid and reliable collection of patient-reported outcomes data.[40,41] To broaden the criteria to include these underrepresented groups would have required parent-report, which was outside the scope of our study. We did not use parents as proxies for patients who were not able to self-report, as they could have created error in the field-test data set.

The response rates were 88.4% for the face-to-face recruitment of nonparticipants in the study, and 32.6% for the mailout recruitment. The characteristics of the nonparticipants are not known.

Although the REDCap survey allowed participants to skip items, resulting in 2.2%–9.0% scale-wise missing data, an advantage of Rasch analysis is the ability to score scales with missing data. Our study did not examine test–retest reliability or responsiveness, nor how scales correlate with other instruments. These topics are the focus of the next phase of our research.

Finally, our large sample means that small differences in mean scores were statistically significant. Future research to understand the clinical meaning of CLEFT-Q scores is needed.

Conclusion

Recently, generic instruments such as PROMIS, KIDSCREEN and the Pediatric Quality of Life Inventory (PedsQL) were used to measure outcomes in cleft lip and/or palate.[42–45] Because cleft-related treatments can substantially change appearance and speech, using scales that lack content validity for cleft lip and/or palate could make it appear that interventions do not improve outcomes, when in fact the scales might not be asking the right questions. The CLEFT-Q provides clinicians and researchers with a rigorously developed instrument they can use to measure clefts and the impact of treatment. The CLEFT-Q is made available free of charge to all nonprofit users once they have signed a licensing agreement.

The science of psychometrics and the sophistication of electronic health records systems have advanced such that it is now possible for providers of health care to collect patient-reported data over the Internet or on a smart device with data used for multiple purposes.[46] Integration of the CLEFT-Q into electronic health records for use in patient care, research and global benchmarking, such as the International Consortium for Health Outcomes Measurement's cleft standard set initiative,[47] could transform health services and outcomes for patients with cleft lip and/or palate around the world.

References

1. Tanaka SA, Mahabir RC, Jupiter DC, et al. Updating the epidemiology of cleft lip with or without cleft palate. *Plast Reconstr Surg* 2012;129:511e-18e.

2. Semb G, Brattstrom V, Molsted 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;42:64-8.

3. Long RE Jr, Hathaway R, Daskalogiannakis J, et al. The Americleft study: an intercenter study of treatment outcomes for patients with unilateral cleft lip and palate part 1. Principles and study design. *Cleft Palate Craniofac J* 2011;48:239-43.

4. Heliövaara A, Küseler A, Skaare P, et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 6. Dental arch relationships at 5 years. *J Plast Surg Hand Surg* 2017;51:52-7.

5. Brattström V, Molsted K, Prahl-Andersen B, et al. The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 2: craniofacial form and nasolabial appearance. *Cleft Palate Craniofac J* 2005;42:69-77.

6. Mølsted K, Brattstrom V, Prahl-Andersen B, et al. The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 3: dental arch relationships. *Cleft Palate Craniofac J* 2005;42:78-82.

7. Semb G, Brattstrom V, Molsted K, et al. The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 4: relationship among treatment outcome, patient/parent satisfaction, and the burden of care. *Cleft Palate Craniofac J* 2005;42:83-92.

8. Shaw WC, Brattstrom V, Molsted K, et al. The Eurocleft study: intercenter study of treatment outcome of patients with complete cleft lip and palate. Part 5: discussion and conclusions. *Cleft Palate Craniofac J* 2005;42:93-8.

9. *Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims*. Silver Spring (MD): US Food and Drug Administration; 2009. Available: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf (accessed 2017 Nov. 16).

10. Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;346:f167.

11. Nelson EC, Eftimovska E, Lind C, et al. PROMs in practice. *BMJ* 2015;350:g7818.

12. Williams K, Sansoni J, Morris D, et al. *Patient-reported outcome measures: literature review*. Sydney: Australian Commission on Safety and Quality in Health Care; 2016.

13. Varagunam M, Hutchings A, Neuburger J, et al. Impact on hospital performance of introducing routine patient reported outcome measures in surgery. *J Health Serv Res Policy* 2014;19:77-84.

14. Basch E. Patient-reported outcomes — harnessing patients' voices to improve clinical care. *N Engl J Med* 2017;376:105-8.

15. Haverman L, Engelen V, van Rossum MA, et al. Monitoring health-related quality of life in paediatric practice: development of an innovative web-based application. *BMC Pediatr* 2011;11:3.

16. Greenhalgh J, Dalkin S, Gooding K, et al. Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. Southampton (UK): NIHR Journals Library; 2017.

17. 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-26e.

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

19. Wong Riff KWY, Tsangaris E, Goodacre TEE, et al. What matters to patients with cleft lip and/or palate: an international qualitative study informing the development of the CLEFT-Q. *Cleft Palate Craniofac J.* 2018;3:442-50.

20. Tsangaris E, Wong Riff K, Goodacre T, et al. Establishing content validity of the CLEFT-Q: a new patient-reported outcome measure for cleft lip and/or palate. *Plast Reconstr Surg Glob Open* 2017;5:e1305.

21. Tsangaris E, Wong Riff K, Forrest C, et al. Methods used for linguistic validation of the CLEFT-Q in 10 languages. *Qual Life Res* 2016;25:1-196.

22. Tsangaris E, Riff KWYW, Vargas F, et al. Translation and cultural adaptation of the CLEFT-Q for use in Colombia, Chile, and Spain. *Health Qual Life Outcomes* 2017;15:228. doi: 10.1186/s12955-017-0805-7.

23. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health* 2005;8:94-104.

24. Wild D, Eremenco S, Mear I, et al. Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices TaskForce report. *Value Health* 2009;12:430-40.

25. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap) — a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377-81.

26. RUMM Laboratory. Welcome to RUMM. Current version RUMM 2030. Available: http://www.rummlab.com.au/ (accessed 2017 July 27).

27. Hobart J, Cano S. Improving the evaluation of therapeutic intervention in MS: the role of new psychometric methods. *Health Technol Assess* 2009;13: iii, ix-x, 1-177.

28. Rasch G. *Probabilistic models for some intelligence and attainment tests*. Vol. 1 of studies in mathematical psychology. Copenhagen: Danmarks Paedagogiske Institut; 1960.

29. Andrich D. *Rasch models for measurement*. Sage university papers series quantitative applications in the social sciences, Vol. 07-068. Thousand Oaks (CA): Sage; 1988.

30. Andrich D, Luo G, Sheridan BE. *Interpreting RUMM2030*. Perth (Australia): RUMM Laboratory; 2016.

31. Andrich D. Rating scales and Rasch measurement. *Expert Rev Pharmacoecon Outcomes Res* 2011;11:571-85.

32. Andrich D. An index of person separation in latent trait theory, the traditional KR-20 index and the Guttman scale response pattern. *Educ Res Perspect* 1982;9:95-104.

33. Fayers PM, Hand DJ, Bjordal K, et al. Causal indicators in quality of life research. *Qual Life Res* 1997;6:393-406.

34. Strauss ME, Smith GT. Construct validity: advances in theory and methodology. *Annu Rev Clin Psychol* 2009;5:1-25.

35. Kincaid JP, Fishburne RP, Rogers RL, et al. *Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel.* Research Branch Report 8–75. Chief of Naval Technical Training: Naval Air Station Memphis; 1975.

36. Stock NM, Feragan KB, Psychological adjustment to cleft lip and/or palate: a narrative review of the literature. *Psychol Health* 2016;31:777-813.

37. Broder HL, Wilson-Genderson M, Sischo L. Oral health-related quality of life in youth receiving cleft-related surgery: self-report and proxy ratings. *Qual Life Res* 2017;26:859-67.

38. Ranganathan K, Shapiro D, Carlozzi NE, et al. The feasibility and validity of PROMIS: a novel measure of quality of life among children with cleft lip and palate. *Plast Reconstr Surg* 2016;138:675e-81e.

39. Jones MC. Etiology of facial clefts: prospective evaluation of 428 patients. *Cleft Palate J* 1988;25:16-20.

40. Bevans KB, Riley AW, Moon J, et al. Conceptual and methodological advances in child-reported outcome measures. *Expert Rev Pharmacoecon Outcomes Res* 2010;10:385-96.

41. Matza LS, Patrick DL, Riley AW, et al. Pediatric patient-reported outcome instruments for research to support medical product labelling: report of the ISPOR PRO Good Research Practices for the Assessment of Children and Adolescents Task Force. *Value Health* 2013;16:461-79.

42. Sundell AL, Tornhage C, Marcusson A. A comparison of health-related quality of life in 5- and 10-year old Swedish children with and without cleft lip and/or palate. *Int J Paediatr Dent* 2017;27:238-46.

43. Ranganathan K, Shapiro D, Aliu O, et al. Health-related quality of life and the desire for revision surgery among children with cleft lip and palate. *J Craniofac Surg* 2016;27:1689-93.

44. Cook AK, Kerins CA, Heppner CE. Dental impacts on health-related quality of life of children with orofacial clefts. *Pediatr Dent* 2016;38:218-23.

45. Broder HL, Norman RG, Sischo L, et al. Evaluation of the similarities and differences in response patterns to the Pediatric Quality of Life Inventory and the Child Oral Health Impact Scores among youth with cleft. *Qual Life Res* 2014;23:339-47.

46. Wu AW, Kharrazi H, Boulware LE, et al. Measure once, cut twice — adding patient-reported outcome measures to the electronic health record for comparative effectiveness research. *J Clin Epidemiol* 2013;66:S12-20.

47. Arora J, Haj M. *Implementing ICHOM's standard sets of outcomes: cleft lip and palate at Erasmus University Medical Centre in the Netherlands.* London and Cambridge (MA): International Consortium for Health Outcome Measures; 2016. Available: http://www.ichom.org/files/ICHOM_Erasmus_MC_Case_Study.pdf (accessed 2017 July 27).

Competing interests: Anne Klassen and Karen Wong Riff are co-developers of the CLEFT-Q and, as such, could potentially receive a share of any licence revenues as royalties based on their institutions' inventor sharing policy if the CLEFT-Q is used in a for profit study.

This article has been peer reviewed.

Affiliations: Department of Pediatrics (Klassen), McMaster University, Hamilton, Ont.; Division of Plastic and Reconstructive Surgery (Wong Riff), Department of Surgery, Hospital for Sick Children, University of Toronto, Toronto, Ont.; Department of Pediatrics (Longmire), McMaster University, Hamilton, Ont.; Department of Pediatric Surgery (Albert), Hospital Sant Joan de Déu, Universitat de Barcelona, Spain; Department of Otolaryngology – Head & Neck Surgery (Allen), University of Colorado School of Medicine, Aurora, Colo.; Department of Plastic Reconstructive and Aesthetic Surgery (Aydin), School of Medicine, Suleyman Demirel University, Isparta Turkey; Inova Children's Hospital (Baker), Falls Church, Va.; Department of Plastic Surgery (Baker), MedStar Georgetown University Hospital, Washington, DC; Modus Outcomes (Cano), Letchworth Garden City, UK; Department of Plastic Surgery (Chan), Flinders Medical Centre, Bedford Park, South Australia; Division of Plastic Surgery, Department of Surgery (Courtemanche), University of British Columbia; BC Children's Hospital, Vancouver, BC; Department of Plastic Surgery (Dreise), University Medical Centre Groningen, University of Groningen, The Netherlands; Department of Plastic Surgery (Goldstein), University of Pittsburgh Medical Center, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, Pa.; Spires Cleft Centre (Goodacre), Oxford University Hospitals NHS Foundation Trust, Oxford, UK; Department of Pediatrics (Harman), McMaster University, Ron Joyce Children's Health Centre, Hamilton, Ont.; Maxillofacial Surgery Department (Munill), Hospital Vall d'Hebron, Barcelona, Spain; National Maxillofacial Unit (Mahony), St James's Hospital, Dublin, Ireland; Fundación Gantz, departamento de Fonoaudiología (Aguilera), El Lazo, Pudahuel, Chile; Department of Reconstructive Plastic Surgery (Peterson), Karolinska University Hospital, Stockholm, Sweden; Plastic and Reconstructive Surgery (Pusic), Memorial Sloan-Kettering Cancer Center, New York, NY; West Midlands Cleft Service (Slator), Birmingham Children's Hospital, Birmingham, UK; Department of Clinical Sciences in Malmö (Stiernman), Lund University, Lund, Sweden; Department of Health Research Methods, Evidence, and Impact (Tsangaris), McMaster University, Hamilton, Ont.; Indiana University School of Medicine (Tholpady), R.L. Roudebush Veterans Affairs Medical Center, Indianapolis, Ind.; Division of Plastic Surgery Universidad el Bosque (Vargas), Fundación Operación Sonrisa Colombia, Bogotá, Colombia; The Centre for Craniofacial Care and Research (Forrest), The Hospital for Sick Children, Toronto, Ont.

Contributors: Anne Klassen, Karen Wong Riff, Stefan Cano, Timothy Goodacre, Andrea Pusic and Christopher Forrest conceived and designed the study. Anne Klassen, Asteria Albert, Gregory Allen, Mustafa Aydin, Stephen Baker, Andrew Chan, Douglas Courtemanche, Marieke Dreise, Jesse Goldstein, Timothy Goodacre, Karen Harman, Montserrat Munill, Aisling Mahony, Mirta Palomares Aguilera, Petra Peterson, Rona Slator, Mia Stiernman, Elena Tsangaris, Sunil Tholpady and Federico Vargas were involved in the acquisition of data from their sites. All listed authors revised the article critically for important intellectual content and have approved the final version to be published. All authors agreed to be accountable for all aspects of the work.

Funding: The research described in this paper was supported by a grant from the Canadian Institutes of Health Research. The authors have no financial interest to declare in relation to the content of this article. The article processing charge was paid from the Canadian Institutes of Health Research grant.

Acknowledgements: The authors are grateful for the operating grant received from the Canadian Institutes of Health Research. Andrea Pusic received support through the NIH/NCI Cancer Center Support Grant P30 CA008748. The authors thank S. Eastwood from McMaster Children's Hospital; R. Courtemanche from BC Children's Hospital; G. Gribble from Dell Children's Hospital; L. Camison from Pittsburgh Children's Hospital; C. Greenlee from Colorado Children's Hospital; L. Hindenburg from Inova Children's Hospital; P. Terrell from Riley Hospital for Children; C. Hernandez from Texas Children's Hospital; R. Fitzgerald from Alder Hey Children's Hospital; N. Fox, K. Cramer, and H. Gerrish from Broomfield Hospital; J. Shearer from Great Ormond Street Hospital; S. Diment from Salisbury District Hospital; D. Clarke from John Radcliffe Hospital; L. Duggan from St. James's Hospital; C.C. Breugem from University Medical Centre Utrecht; J.P.W. Don Griot from VU Medical Center; H.B.C. van't Hoff from Medical Center Leeuwarden; A.L. Keegstra from University Medical Center Groningen; M.J. Koudstaal from Erasmus Medical Center; and S. Singh from GS Memorial Hospital for their cooperation in field testing the CLEFT-Q.

Accepted: Nov. 28, 2017

Correspondence to: Anne Klassen, aklass@mcmaster.ca