BMJ Open

Combining clinical and patient reported outcome measures for standardized speech assessment in cleft patients with the ICHOM Cleft Standard Set: What is the optimal assessment of speech in cleft lip and palate patients?

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-071571
Article Type:	Original research
Date Submitted by the Author:	03-Feb-2023
Complete List of Authors:	Ombashi, Saranda; Erasmus Medical Center, plastic and reconstructive surgery Kurniawan, Melissa; Erasmus Medical Center, plastic and reconstructive surgery Allori, Alexander; Duke University Hospital & children's health center, Plastic, Maxillofacial & Oral Surgery Sharif-Askary, Banafsheh; Duke University Hospital & children's health center, Plastic, Maxillofacial & Oral Surgery Rogers-Vizena, Carolyn; Boston Children's Hospital, Plastic and Oral Surgery Koudstaal, Maarten; Erasmus MC Sophia Children Hospital, Department of Oral and Maxillofacial Surgery Franken, Marie-Christine; Erasmus Medical Center, Ear, Nose and Throat, Speech Therapy Mink van der Molen, Aebele; University Medical Center Utrecht, Department of Plastic Surgery Mathijssen, Irene; Erasmus MC Sophia Children Hospital, Department of Plastic and Reconstructive Surgery Klassen, Anne; McMaster University, Pediatrics Versnel, Sarah; Erasmus MC Sophia Children Hospital, plastic and reconstructive surgery
Keywords:	ORAL & MAXILLOFACIAL SURGERY, Paediatric head & neck surgery < PAEDIATRIC SURGERY, Paediatric oral & maxillofacial surgery < PAEDIATRIC SURGERY, Paediatric plastic & reconstructive surgery < PAEDIATRIC SURGERY, PLASTIC & RECONSTRUCTIVE SURGERY, Quality of Life

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Combining clinical and patient reported outcome measures for standardized speech assessment in cleft patients with the ICHOM Cleft Standard Set: What is the optimal assessment of speech in cleft lip and palate patients?

S. Ombashi[1,2], M. Kurniawan[1,], A.C. Allori[3], B. Sharif-Askary[3], C.R. Rogers-Vizena[4], M.J. Koudstaal[2, 5, 6], M.C. Franken[7], A.B. Mink van der Molen[2,8] , I.M.J. Mathijssen [1,2,6], A.F. Klassen[9] , S.L. Versnel [1,2,6]

[1] Department of Plastic and Reconstructive Surgery, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands

[2] European Reference Network for rare and/or complex craniofacial anomalies and ear, nose and throat disorders

[3] Department of Plastic, Maxillofacial & Oral Surgery, Duke University Hospital & Children's Health Center, Durham, North Carolina, United States of America

[4] Department of Plastic and Oral Surgery, Boston Children's Hospital, Boston, Massachusetts, United States of America

[5] Department of Oral and Maxillofacial Surgery, Special Dental Care and Orthodontics, Erasmus University Medical Center, Rotterdam, The Netherlands

[6] Dutch Craniofacial Center, Erasmus University Medical Center, Rotterdam, The Netherlands

[7] Department of Ear, Nose and Throat, Speech Therapy, Erasmus University Medical Center, Rotterdam, The Netherlands

[8] Department of Plastic and Reconstructive Surgery, Univeristy Medical Center Utrecht, Utrecht, The Netherlands

[9] Department of Pediatrics, McMaster University, Hamilton, Canada

Numer of words: 4031

Number of tables: 2

Number of figures: 4

Number of appendices: 3

Corresponding Author: S. Ombashi, MD. Dr. Molewaterplein 40, 3015 GD Rotterdam, Erasmus University Medical Center; Tel: +316-31989928; E-mail: s.ombashi@erasmusmc.nl

ABSTRACT

Objectives

Speech problems in patients with a cleft palate are often complex and multifactorial. Finding the optimal way of monitoring these problems is challenging. The ICHOM(International Consortium of Health Outcomes Measurement) developed a set of standardized outcome measures at specific ages for patients with a cleft lip and/or palate, including measures of speech assessment. This study evaluates the type and timing of speech outcome measures currently included in this ICHOM set, and additionally discusses speech assessments in other cleft protocols and initiatives.

Methods

An international, multicenter study was set up. Outcomes of clinical measures and Patient Reported Outcome Measures(PROMs) were collected according to the ICHOM set. PROM data from a field test of the CLEFT-Q were collected to examine the value of additional moments of measurement that are used in other cleft initiatives. Analyses were done per cleft type and in different age-groups, and included univariate regression analyses, trend analyses, T-tests, correlations and floor and ceiling effects.

Results

A total of 2500 patients were included. The PROMs correlated low to moderate with clinical outcome measures. The clinical outcome measures correlated low to moderate with each other as well. In contrast, two CLEFT-Q scales correlated strongly with each other. All PROMs and the Percent Consonants Correct(PCC) showed an effect of age. In patients with an isolated cleft palate, a ceiling effect was found in the intelligibility in context scale(ICS).

Conclusion

Recommendations for an optimal speech outcome assessment in cleft patients are made, based on the evaluation of the speech coutome measures in the ICHOM set . Measurement moments of different cleft protocols and initiatives are considered in this proposition. Concerning the type of measures, adjustment of the current PCC score outcome seems appropriate. For centres with adequate resources and specific interest in research, translation and validation of an upcoming tool, the CAPS-A, is recommended.

Strengths and limitations

- **International multicenter study**
- **Cross-sectional data analyses**
- **Elaborate evaluation of multiple time-points**
- **Assessment of PROMs and clinical outcome measures**

INTRODUCTION

In cleft palate patients with or without a cleft lip (CP±L), speech development is often complex. Persistent velopharyngeal incompetence, residual fistula, adenoid atrophy, surgical intervention and hearing problems influence speech disorder severity in this population(1-5). Speech problems in patients with CP±L can have a large impact on an individual's life, as proper speech skills play an essential role in activities, social functioning and participation in society(6). Many treatment pathways are focused on speech improvement to ameliorate Quality of Life (QoL)(7). Logically, speech assessment is an important parameter in cleft care.

However, no consensus has been reached regarding best diagnostic speech outcome measures and their timing in this population(5). Developing scientifically solid instruments to assess speech in an objective manner is complicated, because listener's perception of speech deficits, even by experts, may differ substantially(8). An additional challenge is systematic assessment of the patient's perspective, which is essential to include due to the impact of speech problems on the individual(9). Although widely-accepted agreement seems essential for improvement of cleft care, finding consensus is complex, especially since speech outcomes should be comparable between different languages to facilitate international collaboration.

Recently, the International Consortium for Health Outcomes Measurement (ICHOM) developed the ICHOM Standard Set for Cleft Lip and Palate (ICHOM Standard Set), with different pathways for varying cleft types(10). Based on patient and expert consensus, a minimal, accessible set of outcome measures was established to enable benchmarking between cleft centers in a systematic manner. For speech assessment, an outcome set was included with both clinical measures and Patient Reported Outcome Measure (PROMs), being the patient's and parent's perspectives.

So far, the selected standardized speech outcome measures and their timing have not been evaluated. As an increasing number of centers are implementing this set, it is important to critically evaluate and optimize this ICHOM Standard Set. Three centers, the Boston Children's Hospital (Boston, USA), Duke University Hospital (Durham, USA), and the Erasmus Medical Center (Rotterdam, The Netherlands), started clinical implementation and an international collaboration in 2015. The overarching aim of this collaboration is to share data and knowledge obtained by using the set in standard care. Additionally, they collaborate with McMaster University (Hamilton, Canada), who developed and field tested the CLEFT-Q questionnaire, of which many scales are included in the ICHOM Standard set.

The objective of this study was to evaluate the current standardized speech outcome measures of the ICHOM Standard Set for patients with CP±L. More specifically, the value of every speech outcome measure was examined, as well as the best age intervals for assessment of these outcome measures. In addition, other speech assessment tools are discussed. Finally, recommendations are made for an optimal and complete assessment of speech in patients with CP±L, that is efficient and accessible for all cleft centers.

METHODS

Patient population

Three centers (Boston Children's Hospital, Duke University Hospital, and Erasmus Medical Center,) each implemented the ICHOM Standard Set in 2015. All patients treated at these centers for a cleft palate with a cleft lip/cleft alveolus (CL(A)P), or an isolated cleft palate (CP) who were assessed according to the ICHOM Standard Set (age range 5-22 years), were included. In addition, another patient group derived from an international field test of the CLEFT-Q, by McMaster University(11). According to the age cut-off of the ICHOM Standard Set, only outcomes from CLEFT-Q scales of field test patients with a CP±L up to 22 years old were included in the current study(fig 1). Patients from the participating centers were excluded in case they could not sufficiently speak or write the language native to the center's country .

Patient and public involvement

In the development of the ICHOM Standard Set, patients were actively involved. The ICHOM Standard Set was implemented in each center as part of regular clinical care. Data was pseudonymized and collected retrospectively, and ethical approval was obtained to do so without explicit consent from each patient and parent. Results of this study may be of use to further improve the currently used ICHOM Standard Set, and therewith regular clinical care.

Outcome measures

PROMs

CLEFT-Q Scales

The CLEFT-Q is developed specifically to assess QoL from the patient's perspective in patients with a CP±L. A literature review, patient interviews and psychometric testing, established the final content of the scales, which covers several overarching domains(12-14). Speech is assessed through two scales, each covering a different domain. Both scales have 3 response options for each item (always; sometimes; never); a lower score equals a worse outcome. Completing the scales can be done online; it will take the patient several minutes.

Speaking-Related Distress (SDistress) is part of the psychosocial domain. The scale contains 10 items that relate to the psychosocial part of speech difficulties, like nervousness or frustration(14).

Speech Function (SFunction) focuses on the functional speech difficulties that patients themselves identify, for example the ability to say certain letters or words. The scale consists of 12 items that belong to the facial function domain(14).

Intelligibility in Context Scale(ICS) is a 7-item, parent-reported questionnaire to assess the intelligibility of the child on a Likert-point scale. The items assess the degree to which the speech of the patient is understood by different communication partners. ICS appeared to be a valid and reliable tool for children with speech disorders, (15, 16), but not specifically designed nor validated for patients with CP±L(17). It is available in several languages, and normative data exists for English-speakers(18, 19).

Clinical outcome measures

Percent Consonants Correct(PCC) is developed to detect speech sound errors. PCC scores are calculated by using a standard, crossecitonally translated set of words that include all speech problems children with CP±L often tend to have.

In case of any problems, their severity can be categorized: PCC scores of 85-100% indicate mild to no problems; scores of 65-84.9% indicate mild-moderate problems; scores of 50-64.9% indicate moderate-severe problems; and scores <50% indicate severe problems(20). PCC is suitable for usage in patients with CP±L when assessed by well-trained clinicians(8).

Velopharyngeal Competence rating(VPC) discriminates between three categories: 'competent', 'marginally incompetent' and 'incompetent'. The outcome is determined by the speech therapist based on the PCC test and spontaneous speech. In case of any clinical evidence of minor problems regarding the competence, VPC was categorized as 'marginally incompetent'. When clinically significant problems were detected, suggesting surgical management and/or speech therapy, VPC was categorized as 'incompetent'. Prior studies found VPC to be suitable as a first clinical choice for the assessment of velopharyngeal dysfunction and is recommended for both clinical follow-up and research(20).

Data collection

All participating centers obtained ethical approval for the current study from their local ethics committees. Data was collected restrospectively over a 6 year period (2015-2020). According to the ICHOM protocol, both CLEFT-Q Speech scales were assessed at ages 12 and 22 years(fig. 1).
Both PCC and VPC were scored at ages 5, 12 and 22 years, and ICS at ages 5 and 12 years.
The field test cross-sectionally collected data from patients with a cleft across 12 different countries with different income-statuses(11). As 8 years is the minimum age to complete the CLEFT-Q, both CLEFT-Q Speech scales from field test patients with CP±L from 8-22 years old were included(fig. 1).
Income status of the country according to the World Bank Classification was made within the field test. Data from the ICHOM centers were all categorized as deriving from high-income countries.

Baseline characteristics that were collected included gender, type of cleft, and age at the time of assessment.

Data analysis

Data was analysed in R-studio, a free software environment for statistical computing and graphics(21). Psychometric validation of the SDistress and SFunction confirmed suitability to use a 0 to 100 scale deriving from the sum scores for analysis(11).

For analysis of the ICS questionnaire, the average score of the seven items was used. VPC was used as an ordinal variable, whereas PCC scores were expressed as proportions.

All participating ICHOM centers are high-income countries, whereas part of the field test data was collected in upper middle and lower middle income countries. To prevent possible influence of income-status on the outcomes, univariate regression analyses were used to examine differences in outcome scores of the SDistress and SFunction before further analyses. Data was categorized according to the income status of the country where the data had been collected.

In order to examine the added value of each PROM and clinical outcome measure in regard with the other measures, correlations were examined between the PROMs; between the clinical outcome measures; and between the PROMs and the clinical outcome measures. Pearson correlations were used, and outcomes were analysed per cleft type. Correlations were considered strong in case $r>0.7$; moderate between $r=0.5$ and $R= 0.7$; and weak in case $r<0.5$.

Analyses within and between different age-groups were done to explore whether the current outcome measures are assessed at the optimal age-points, and whether additional measurement moments are indicated either for PROMs or clinical outcome measures. Therefore, not only time-points of the ICHOM protocol were included in analyses. As CLEFT-Q outcome scores of all ages between 8-22 years were included from the field test data, time-points used by other large initiatives as Eurocleft, Scandcleft and

Americleft were considered as well (22-24). Doing so, the following age-groups were set up: 5-7 years; 8-9 years; 10-13 years; 14-16 years; 17-19 years, 20-22 years(fig. 1).

Per age-group, possible differences between scale scores were examined with independent T-tests. Bonferroni correction was applied for multiple testing. Trend analyses were performed to identify potential problems in specific age-groups.

Floor- and ceiling effects were examined to identify the suitability of the outcome measures in our population. A floor or ceiling effect is seen when a considerable amount of the outcome scores are either scored the best (in this case a maximum score, thus a ceiling effect), or the worse (in this case a minimum score, thus a floor effect). Both effects result in a truncated distribution of the outcomes on either side of the scale (25, 26). Minimum and maximum score outcomes of all PROMs and the PCC were evaluated. A percentage of 20% or more of the patients scoring the minimum or maximum outcome score was considered as a ceiling effect. In VPC, the outcome distributions were examined.

RESULTS

Characteristics of the included participants

A total of 2500 patients were included in the study; 1723 derived from the field test, and 777 from the ICHOM centers (table 1).

Table 1: Demographics and phenotypes

Descriptives total

	Field test (n = 1723)	ICHOM (n = 777)	Overall (n = 2500)
Sex			
Male	981 (56.9%)	444 (57.1%)	1425 (57.0%)
Female	742 (43.1)	333 (42.9%)	1075 (43.0%)
Cleft Type			
CP	517 (30.0%)	301 (38.7%)	818 (32.7%)
CL(A)P	1206 (70.0%)	476 (61.3%)	1682 (67.3%)
Income classifications			
High Income	1364 (79.2%)	777 (100%)	2141 (85.6%)
Up Middle income	199 (11.5%)	0 (0%)	199 (8.0%)
Low Middle Income	160 (9.3%)	0 (0%)	160 (6.4%)

There were slightly more males than females, and relatively more patients with a CL(A)P than with a CP. Significant differences between countries with a High and Upper-/Lower-Middle income status of the field test were found and results are shown in appendix 1.

Therefore, further analyses were done only with the patient population of the field test deriving from countries with a high income status, like all participating ICHOM Cleft centers (n = 2141). The subgroup characteristics are included in appendix 2.

Associations between the outcome measures

Correlations between all outcome measures (clinical and PROMs), in both cleft types (CP and CL(A)P), appeared significant (P<0.05), except for the correlation between PCC and SDistress (p = 0.285)(fig. 2).

Correlation PROMs: The SDistress and SFunction showed a strong correlation in patients with CP (r = 0.76) and a moderate correlation in patients with CL(A)P (r = 0.68). The ICS and SFunction correlated strongly (r = 0.73) in patients with CP, and moderately (r = 0.64) in patients with CL(A)P; whereas the ICS and the SDistress correlated moderately in patients with CP (r = 0.52) and weak in patients with CL(A)P (r = 0.47).

Correlation clinical outcomes: VPC and PCC were (negatively) moderately correlated in both cleft types (r = -0.62 and r = -0.67 in patients with CP and CL(A)P respectively).

Correlation PROMs and clinical outcomes: Moderate correlations were found between the PCC and ICS in patients with CP (r = 0.64), and in patients with CL(A)P (r= 0.5). VPC and ICS had a (negative) weak correlation in patients with CP (r = -0.49), and CL(A)P (r = -0.43). The SDistress and SFunction were weakly correlated with VPC and PCC (negatively) in both cleft types (fig. 2).

Comparing outcome measures between age-groups

SDistress and SFunction: showed the highest mean outcome (i.e. the most favourable ratings) in the age-group of 14-16 years old. From thereon, a slightly downward trend is seen (fig. 3). In both CLEFT-Q Speech scales the lowest mean outcome scores were found in the age-group of 8-9 years old, which was significantly different in comparison with the other age-groups in patients with CL(A)P (p<0.05, table 2).

Table 2 Mean outcomes per age-group, per cleft type

		Age categories	SDistress	SFunction	ICS	PCC
Cleft Palate		5 – 7			4.08 (0.73)	72.49 (31.14)
		8 – 9	70.44 (20.64)	66.12 (22.12)		
		10 – 13	74.84 (19.91)	72.56 (21.52)	4.45 (0.62)[1]	92.02 (11.48)[1]

	14 – 16	75.10 (22.77)	74.66 (24.32)		
	17 – 19	70.59 (16.92)	73.00 (19.02)		
	20 – 22	70.27 (21.93)	71.69 (24.98)		90.92 (15.33)[1]
	5 – 7			4.03 (0.50)	46.06 (25.59)
Cleft Lip Palate	8 – 9	65.32 (21.61)	64.75 (20.16)		
	10 – 13	73.16 (20.62)[1]	69.57 (20.96)	4.32 (0.55)[1]	76.25 (22.63)[1]
	14 – 16	72.81 (19.66)[1]	74.22 (19.14)[1]		
	17 – 19	71.82 (19.67)[1]	71.41 (20.79)[1]		
	20 – 22	71.49 (18.00)	72.26 (18.40)[1]		86.83 (16.13)[1,2]

1 $p < 0.05$ compared to 5-7 or 8-9 age
2 $P < 0.05$ compared to 10 – 13 age

ICS : Both patient groups showed a significant difference between the two age-groups 5 and 12 years, ICS was significantly lower at 5 years than at 12 years in both cleft types (tab. 2).

PCC: Observing the trends in the clinical outcome measures, an upward trend regarding PCC score was seen (fig. 3). In both cleft types, PCC scores differed significantly between the age-groups(tab. 2).

VPC: In the age-group of 5 years, 25.6% of the patients with CP and 60.6% of the patients with CL(A)P were scored as incompetent. In 22 year-olds, this percentage was 11.1% in patients with CP and 16.7% in patients with CL(A)P.

No floor effects were found in any of the PROMs. In patients with CP, the ICS showed a ceiling effect (29.0%, n = 169). No ceiling effects were observed in patients with CL(A)P. An overview of all maximum scores and the VPC score distribution is shown in appendix 3.

DISCUSSION

Evaluation of the value of the current ICHOM speech outcome measures

All correlations between PROMs were moderate, except for the strong correlation of the SFunction with both the SDistress and the ICS in patients with a CP. The fact that the correlation between the SFunction and SDistress is stronger in patients with CP than in patients with CL(A)P, suggests that the visibly different appearance in patients with CL(A)P plays a significant role in SDistress as well; in a social context, looking differently may cause additional or more distress besides having speech problems. This is supported by our finding that the ICS correlated moderately with SFunction, but weakly with SDistress in the CL(A)P group. Parent reported speech intelligibility correlated higher to children's self report of their speech function, than it did to the speech distress the children themselves experience. In the latter, distress about appearance could be included. This finding suggests that the ICS can give an indication of 'patient-reported' SFunction in young children who cannot complete a PROM themselves yet (7 years and younger).

The PROMs showed weak correlations with the clinical reported outcomes; except for the moderate correlation that was seen between the ICS and the PCC in both patient groups. Based on these findings, PROMs appear to be of added value, as they provide different information than captured with the clinical outcome measures included in the Standard Set. They add a unique dimension to speech outcome measurement – a subjective dimension related to the patient's experiences with everyday speaking situations. While clinical measures objectively appraise the quality of speech, they will probably be insufficient to adequately capture the more nuanced social, emotional, and psychological aspects of SDistress and SFunction. With this additional self-report and parental information, clinicians

can more comprehensively explore the patients' problems concerning speech in order to find out whether additional treatment or guidance is indicated.

Evaluation of the impact of age of assessment on measurement outcomes

In both CLEFT-Q Speech scales, the age-group of 8-9 years enholds the worst scores. Speech improvement due to speech therapy or late closure of the hard palate (in certain protocols around the age of 9 years when alveolar bone grafting is performed), might explain the higher, better scores in the age-groups of 10-13 and 14-16 years. In age-groups 17 and up however, CLEFT-Q scores appeared to decline whereas PCC scores improved. This finding suggests that (almost) adult patients with CP±L develop feelings of insecurity concerning their speech, although their speech sound production remains good, or even improves. This is in line with speech therapists' experiences in the outpatient clinic, where patients were seen in person at the age of 22, but not at age of 17-19. Quite often, when discussing outcomes of the CLEFT-Q scales as well as the PCC with the patient, (s)he reacted surprised when told that no (cleft-related) problems were present in their speech. Taking the lower CLEFT-Q scores in 8-9-, 17-19- and 20–22-year-olds that were found in the field test into consideration, additional assessment of a PROM at the age-groups of 8-9 (youngest age at which this PROM can be assessed) and 17-19 years old should be considered for implementation in the ICHOM Standard Set. Therewith, monitoring patients more closely will be enabled, and any concerns of patients with CP±L regarding their speech can be discussed timely.

The two CLEFT-Q speech scales showed to capture overlapping information as they strongly correlate in patients with CP. Questions deriving from the SDistress are not measurable in any other manner, whereas SFunction from the patient's perspective might be less of added value for a PROM questionnaire. Therefore, implementation of the CLEFT-Q SDistress scale in patients with both cleft types is recommended at the age-groups of 8-9 years and 17-19 years. (fig. 4).

A ceiling effect in **ICS** outcomes of patients with CP, without clear differences between average scores in patients with CP and CL(A)P, suggests that the group with CP contains a diverse population and severity of the speech problems vary widely. Furthermore, since ICS is not specifically developed for a population with CP±L, it is debatable whether this tool captures the information necessary to point out all relevant speech problems in the patient group.

However, exclusion of ICS could mean that a large part of the speech problems in the population with CP would remain undetected. Assessment at 5 and 12 years in patients with both cleft types, which is the current timing in the ICHOM Standard Set, appears therefore appropriate despite the ceiling effect.

Although **VPC** scores were relatively favourable in patients with CP, no changes regarding the implementation of the VPC scores are recommended as the outcomes showed to vary. VPC can serve as a suitable screening tool and outcomes are easily gathered by the observation of a clinician. Hence, patient-burden is low and the tool efficiently detects any velopharyngeal problems.

PCC scores that were found indicated speech sound problems especially in the younger age-groups of the patients with CL(A)P. 22-year-olds with both cleft types showed mild speech sound problems in general. Therefore time points as currently implemented in the ICHOM Standard Set appear adequate. In contrast, the suitability of PCC-assessment in a cleft set focusing on standardized outcome measures is still debatable, as inter- and intra-center reliability have not been investigated thoroughly in all participating centers so far(8). Future research should include an examination of scoring and interpreting PCC scores in different centers and/or different countries.

Future considerations regarding alternative speech outcome measures

In order to establish an optimal cleft set for speech assessment, other standardized outcome measures should be considered. Based on clinical experience with ICHOM Standard Set, possible suggestions for additional outcome measures are discussed here.

Regarding **PROMs** for speech assessment in patients with CP±L, the CLEFT-Q scales seem to be the most suitable PROMs available. Their comprehensive psychometric examination and cross-cultural character make them accessible for all cleft centers that seek an efficient minimal cleft set that comprises all important speech parameters(11-13). The standardized approach for translation and validation of the CLEFT-Q questionnaire enables accessibility of the PROM even for centers that still need to translate the CLEFT-Q into their native language(27, 28). Another cleft-specific PROM is the Cleft Hearing and Speech Questionnaire (CHASQ). Whereas the psychometric properties of the CLEFT-Q were examined throughout Rasch measurement theory, classical test theory was used for the CHASQ(29). A recent cross-sectional questionnaire study that compared the CLEFT-Q with the CHASQ, found that the majority of the patients with CP±L preferred the CLEFT-Q(29). Therefore, implementation of the CHASQ speech does not seem to be of added value in the current cleft set.

Besides the used **VPC** measure, a more elaborate variant exists, namely the VPC-Summary (VPC-Sum). This includes assessment of hypernasality, passive VPI symptoms and the transcriptions of active non-oral consonant errors(30). VPC-sum can either be reported as a score between 0 and 6, or as a dichotomized outcome (velopharyngeal competence or incompetence)(30).

Calculation of the VPC-Sum is based on single-words, whereas VPC-rate is based on observation of spontaneous speech (31). VPC-Sum would be an interesting measure due to its efficiency, although it may not be achievable to implement VPC-Sum in all centers in the near future as only five different

languages are currently available(31). Other alternatives such as nasopharyngoscopy or MRI are invasive, expensive and enlarge the patient burden(32), and therefore not easy accessibility for all centers.

The currently implemented **PCC** lacks any categorization of consonant errors. The **Eurocleft speech group** created a research protocol with a phonetic framework, which was used in six centers and five different languages(33). It also included consonant production, but assessed on sentence level instead of single words. It categorized into 3 groups (correct; almost correct; incorrect). Further division into 21 error categories that were sampled in five groups was done in case of incorrect consonants (Nasal airflow; glottal realisations; alveolar deviations; sibilant deviations; other)(33). Moreover, general speech quality was assessed concerning hyper- and hyponasality, and voice quality(34). Expert rating of these outcomes requires periodic training of sufficient interrater reliability. However, it might be too detailed for implementation in an efficient, clinically oriented cleft set. Therefore, we suggest to further categorize the PCC score, although not as detailed as in the Eurocleft studies. Based on clinical experience with the ICHOM Standard Set, it is recommended that speech pathologists report whether any cleft related, phonological, or phonetic problems are detected.

Another clinical outcome measure, the **Great Ormond Street Speech Assessment 1998** (GOS.SP.ASS'98), provides a comprehensive view of all speech associated features for patients with CP±L(35, 36). Its suitability for inter-center comparison would make it interesting for the ICHOM Standard Set(10); however, it is too detailed for clinical audit(37). In succession the **Cleft Audit Protocol for Speech Augmented** (CAPS-A) was developed for cleft-related problems, and could be an alternative for PCC(38). Seen its rigorous psychometric assessment, it fits well into a set that seeks standardized outcome measures. The Americleft Speech Project found that an acceptable inter- and intrarater reliability can be

achieved(37, 39). Furthermore, it is suitable for assessment in 5-years-olds, which enables detection of speech problems at an earlier age(40). However, the CAPS-A is limited in types of statistical analyses due to the scaling type used (equal appearing interval)(41). A more practical challenge concerning implementing the CAPS-A would be the required training of all involved speech therapists, and the amount of time the assessment takes (15 min.)(38). Moreover, the CAPS-A is developed and applicable for English-speaking countries, necessitating translation and validation in other languages(37). The CAPS-A is not ideal for centers interested in a minimal and efficient cleft set. However, centers with experience and resources are highly recommended to implement this tool in order to promote further international standardization of elaborate speech assessment in patients with CP±L(fig. 4). Implementation of the CAPS-A would also enable the use of the recently developed and validated CAPS-A-VPC-Sum score to reliably measure velopharyngeal function(42). Our suggestion for centers that consider the implementation of the CAPS-A, is to assess it at the ages of 5-7, 10-13 and 20-22 years in order to enable long-term follow-up.

Limitations of the study

Data was analyzed cross-sectional. Longitudinal analyses to explore development of speech and for benchmarking will be possible in the future, since data collection continues.

CONCLUSION

From the current study, it can be concluded that the current ICHOM Standard Set is informative and efficient. PROMs were shown to be of added value, and the CLEFT-Q appeared to be the most suitable PROM. Therefore, continuation of collecting the current outcome measures and time points is

recommended. Furthermore, a minor extension is suggested: In addition to the current time-points of assessment, it is recommended to implement the CLEFT-Q SDistress scale at the age of 8-9 and 17-19 as well. Further adjustments of the set could comprise an additional categorization of the PCC score, based on the framework of Eurocleft and adjusted for clinical usage.

ACKNOWLEDGMENTS

We would like to thank all children and caregives that participated in the study, that contributed by filling out the questionnaires and allowing the research team to use their data for the current study. Furthermore, we would like to thank the speech pathologists from all our participating centers, that contributed tremendously by collecting the data consequently and by advising the research team on important clinical considerations

Contributor statement

Saranda **Ombashi**: Data collection in Dutch Center; correspondence to other centers; data analyses; writing and adjusted the manuscript based on advice of the other authors.*

Melissa S.I.C. **Kurniawan**: Data analyses, co-writer on the methods and results section of the manuscript.*

Alexander C. **Allori:** Data collection at Duke University Hospital, advisor of project plan and manuscript**.***

B. **Sharif-Askary:** Data collection at Duke University Hospital, advisor of project plan and manuscript.*

Carolyn R. **Rogers:** Data collection at Harvard University Hospital, advisor of project plan and manuscript.*

Maarten J. **Koudstaal**: Writer of the Ethical Board Approval, Supervision of data analyses. Advisor and supervisor on project plan and manuscript.*

M.C. **Franken:** Data collection at Erasmus Medical Cewnter, advisor of project plan and manuscript, with special focus on the clinical content.*

A.B. **Mink van der Molen**: advisor of project plan and manuscript.*

Irene M.J. **Mathijssen:** Overall supervision of project plan and manuscript.*

Anne F. **Klassen:** Data collection Field Test Mc Master University, advisor on project plan and manuscript.*

Sarah L. **Versnel:** Project leader, correspondence with all other centers. Supervision on project plan, data analyses and manuscript.*

*All authors approved the final version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved

Competing interests:

The CLEFT-Q is owned by McMaster University and the Hospital for Sick Children. Prof. Anne Klassen is a co-developer of the CLEFT-Q and would receive a share of any licence revenues as royalties based on her institutions' inventor sharing policy if the CLEFT-Q was used in a for profit study.

The other authors have no disclosures.

Funding

This project received no funding. However, S. Ombashi works as a PhD-student for an European Union-funded Network, the 'Eureopean Reference Network for Craniofacial anomalies and Ear, Nose and Throat disorders'. The results and findings from this project are of help in further European alignment concerning standardized outcome measures in cleft care.

Data sharing statement

A data sharing agreement was set up and signed by all participating centers that shared data for this project.

Ethics approval statement

All participating centers obtained ethical approval for the current study from their local ethics committees.

References

1. Kummer AW. Types and causes of velopharyngeal dysfunction. Semin Speech Lang. 2011;32(2):150-8.
2. Knight J, Cassell CH, Meyer RE, Strauss RP. Academic outcomes of children with isolated orofacial clefts compared with children without a major birth defect. Cleft Palate Craniofac J. 2015;52(3):259-68.
3. Van Lierde KM, Dhaeseleer E, Luyten A, Van De Woestijne K, Vermeersch H, Roche N. Parent and child ratings of satisfaction with speech and facial appearance in Flemish pre-pubescent boys and girls with unilateral cleft lip and palate. Int J Oral Maxillofac Surg. 2012;41(2):192-9.
4. Peterson-Falzone SJ. Speech outcomes in adolescents with cleft lip and palate. Cleft Palate Craniofac J. 1995;32(2):125-8.
5. Baek RM, Kim BK, Jeong JH, Ahn T, Park M, Han J. The effect of age at surgery and compensatory articulation on speech outcome in submucous cleft palate patients treated with double-opposing Z-plasty: A 10-year experience. J Plast Reconstr Aesthet Surg. 2017;70(5):646-52.
6. Markham C, van Laar D, Gibbard D, Dean T. Children with speech, language and communication needs: their perceptions of their quality of life. Int J Lang Commun Disord. 2009;44(5):748-68.
7. Nicola K, Watter P. Health-related quality of life from the perspective of children with severe specific language impairment. Health Qual Life Outcomes. 2015;13:127-.
8. Sell D, Sweeney T. Percent Consonant Correct as an Outcome Measure for Cleft Speech in an Intervention Study. Folia Phoniatr Logop. 2019:1-9.
9. Klassen AF, Tsangaris E, Forrest CR, Wong KW, Pusic AL, Cano SJ, et al. Quality of life of children treated for cleft lip and/or palate: a systematic review. J Plast Reconstr Aesthet Surg. 2012;65(5):547-57.
10. Allori AC, Kelley T, Meara JG, Albert A, Bonanthaya K, Chapman K, et al. A Standard Set of Outcome Measures for the Comprehensive Appraisal of Cleft Care. Cleft Palate Craniofac J. 2017;54(5):540-54.
11. Klassen AF, Riff KWW, Longmire NM, Albert A, Allen GC, Aydin MA, 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-E62.
12. Wong Riff KW, Tsangaris E, Goodacre T, Forrest CR, Pusic AL, Cano SJ, 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.
13. Tsangaris E, Wong Riff KWY, Goodacre T, Forrest CR, Dreise M, Sykes J, 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.
14. Wong Riff KWY, Tsangaris E, Goodacre TEE, Forrest CR, Lawson J, Pusic AL, 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;55(3):442-50.

15. McLeod S, Crowe K, Shahaeian A. Intelligibility in Context Scale: Normative and Validation Data for English-Speaking Preschoolers. Lang Speech Hear Serv Sch. 2015;46(3):266-76.
16. McLeod S. Intelligibility in Context Scale: cross-linguistic use, validity, and reliability. Speech, Language and Hearing. 2020;23(1):9-16.
17. 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):648-56.
18. McLeod S, Harrison LJ, McCormack J. Skattning av förståelighet i kontext: Svenska [Intelligibility in Context Scale: Swedish] (T. Lagerberg, Trans.) Bathurst, NSW, Australia: Charles Sturt University. 2012.
19. McLeod S, Harrison LJ, McCormack J. Schaal voor verstaanbaarheid in de context: Nederlands [Intelligibility in Context Scale: Dutch] (A. van Doornik, Trans). Bathurst, NSW, Australia: Charles Sturt University. 2013.
20. Lohmander A, Hagberg E, Persson C, Willadsen E, Lundeborg I, Davies J, et al. Validity of auditory perceptual assessment of velopharyngeal function and dysfunction – the VPC-Sum and the VPC-Rate. Clinical Linguistics & Phonetics. 2017;31(7-9):589-97.
21. The R Project for Statistical Computing: The R Foundation; 2021 [Available from: https://www.r-project.org/.
22. Semb G, Enemark H, Friede H, Paulin G, Lilja J, Rautio J, et al. A Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 1. Planning and management. J Plast Surg Hand Surg. 2017;51(1):2-13.
23. Long RE, Jr., Hathaway R, Daskalogiannakis J, Mercado A, Russell K, Cohen M, 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(3):239-43.
24. Shaw WC, Semb G, Nelson P, Brattstrom V, Molsted K, Prahl-Andersen B, et al. The Eurocleft project 1996-2000: overview. J Craniomaxillofac Surg. 2001;29(3):131-40; discussion 41-2.
25. Lim CR, Harris K, Dawson J, Beard DJ, Fitzpatrick R, Price AJ. Floor and ceiling effects in the OHS: an analysis of the NHS PROMs data set. BMJ Open. 2015;5(7):e007765-e.
26. Stucki G, Liang MH, Stucki S, Katz JN, Lew RA. Application of statistical graphics to facilitate selection of health status measures for clinical practice and evaluative research. Clin Rheumatol. 1999;18(2):101-5.
27. Tsangaris E, Riff K, Vargas F, Aguilera MP, Alarcon MM, Cazalla AA, et al. Translation and cultural adaptation of the CLEFT-Q for use in Colombia, Chile, and Spain. Health Qual Life Outcomes. 2017;15(1):228.
28. Tsangaris E, Wong Riff KWY, Dreise M, Stiernman M, Kaur MN, Piplani B, et al. Translation and cultural adaptation of the CLEFT-Q into Arabic, Dutch, Hindi, Swedish, and Turkish. European Journal of Plastic Surgery. 2018;41(6):643-52.
29. Stiernman M, Klintö K, Persson M, Becker M. Comparison of Corresponding Scores From the Cleft Hearing Appearance and Speech Questionnaire (CHASQ) and CLEFT-Q in Swedish Patients With Cleft Lip and/or Palate. The Cleft Palate-Craniofacial Journal. 2020:105566562096412.

30. Hammarström IL, Nyberg J, Alaluusua S, Rautio J, Neovius E, Berggren A, et al. Scandcleft Project Trial 2-Comparison of Speech Outcome in 1- and 2-Stage Palatal Closure in 5-Year-Olds With UCLP. Cleft Palate Craniofac J. 2020;57(4):458-69.
31. Lohmander A, Willadsen E, Persson C, Henningsson G, Bowden M, Hutters B. Methodology for speech assessment in the Scandcleft project--an international randomized clinical trial on palatal surgery: experiences from a pilot study. Cleft Palate Craniofac J. 2009;46(4):347-62.
32. Shprintzen RJ, Marrinan E. Velopharyngeal insufficiency: diagnosis and management. Curr Opin Otolaryngol Head Neck Surg. 2009;17(4):302-7.
33. Brondsted K, Grunwell P, Henningsson G, Jansonius K, Karling J, Meijer M, et al. A phonetic framework for the cross-linguistic analysis of cleft palate speech. Clinical Linguistics & Phonetics. 1994;8(2):109-25.
34. Grunwell P, Brondsted K, Henningsson G, Jansonius K, Karling J, Meijer M, et al. A six-centre international study of the outcome of treatment in patients with clefts of the lip and palate: the results of a cross-linguistic investigation of cleft palate speech. Scand J Plast Reconstr Surg Hand Surg. 2000;34(3):219-29.
35. Sell D, Harding A, Grunwell P. A screening assessment of cleft palate speech (Great Ormond Street Speech Assessment). Eur J Disord Commun. 1994;29(1):1-15.
36. Sell D, Harding A, Grunwell P. GOS.SP.ASS.'98: an assessment for speech disorders associated with cleft palate and/or velopharyngeal dysfunction (revised). Int J Lang Commun Disord. 1999;34(1):17-33.
37. John A, Sell D, Sweeney T, Harding-Bell A, Williams A. The cleft audit protocol for speech-augmented: A validated and reliable measure for auditing cleft speech. The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association. 2006;43(3):272-88.
38. Ahl R, Harding-Bell A. Comparing Methodologies in a Series of Speech Outcome Studies: Challenges and Lessons Learned. The Cleft Palate-Craniofacial Journal. 2017;55(1):35-44.
39. Chapman KL, Baylis A, Trost-Cardamone J, Cordero KN, Dixon A, Dobbelsteyn C, et al. The Americleft Speech Project: A Training and Reliability Study. Cleft Palate Craniofac J. 2016;53(1):93-108.
40. Britton L, Albery L, Bowden M, Harding-Bell A, Phippen G, Sell D. A cross-sectional cohort study of speech in five-year-olds with cleft palate ± lip to support development of national audit standards: benchmarking speech standards in the United Kingdom. Cleft Palate Craniofac J. 2014;51(4):431-51.
41. Baylis A, Chapman K, Whitehill TL, The Americleft Speech G. Validity and Reliability of Visual Analog Scaling for Assessment of Hypernasality and Audible Nasal Emission in Children With Repaired Cleft Palate. Cleft Palate Craniofac J. 2015;52(6):660-70.
42. Pereira VJ, Tuomainen J, Lee KYS, Tong MCF, Sell DA. A perceptual outcome measure of velopharyngeal function based on the Cleft Audit Protocol for Speech-Augmented (CAPS-A VPC-Sum): Validation through a speech osteotomy study. Int J Lang Commun Disord. 2021;56(4):754-67.
43. Smarius BJA, Haverkamp S, de Wilde H, van Wijck-Warnaar A, Mink van der Molen AB, Breugem CC. Incidence of cleft-related speech problems in children with an isolated cleft lip. Clin Oral Investig. 2020.

44. Vallino LD, Zuker R, Napoli JA. A study of speech, language, hearing, and dentition in children with cleft lip only. Cleft Palate Craniofac J. 2008;45(5):485-94.
45. Teele DW, Klein JO, Chase C, Menyuk P, Rosner BA. Otitis media in infancy and intellectual ability, school achievement, speech, and language at age 7 years. Greater Boston Otitis Media Study Group. J Infect Dis. 1990;162(3):685-94.
46. Ruegg TA, Cooper ME, Leslie EJ, Ford MD, Wehby GL, Deleyiannis FWB, et al. Ear Infection in Isolated Cleft Lip: Etiological Implications. The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association. 2017;54(2):189-92.

FIGURE LEGENDS

Figure 1. Speech outcomes of ICHOM Cleft Lip and Palate and field test data

PROMs filled out by the patients are colored in purple, and grey if filled out by the parent(s). Clinical outcome measures are colored in yellow.

*Data derived from the field test only in these cohorts, as there is no measurement moment included in ICHOM Cleft Lip and palate in these age-groups.

Figure 2. Correlations in patients with CP and CL(A)P

All correlations in both cleft types appeared significant (P<0.05), except for the correlation between the PCC and CLEFT-Q SDistress in patients with CL(A)P (p = 0.285). Note that VPC is inversely scored (higher numbers correspond to worse outcomes), thus accounting for the negative correlations with the other scales.

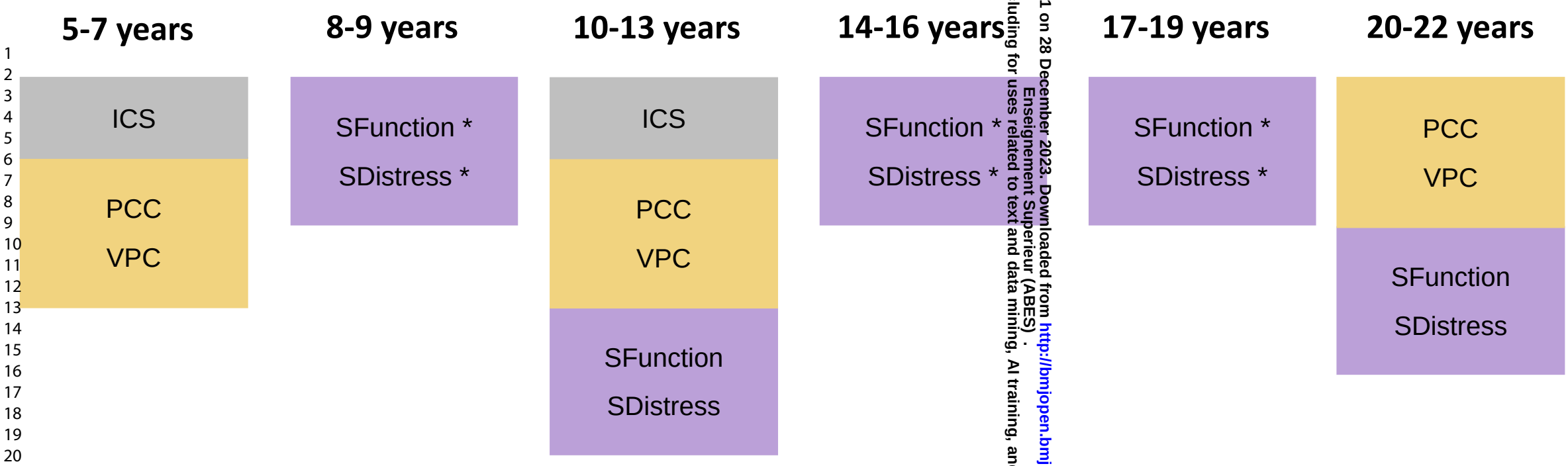
Figure 3: cross-sectional trend analyses of the age-groups

Analyses are presented per outcome measure, per cleft type.

Figure 4 Overview of the new proposed ICHOM Cleft Lip and Palate set concerning speech assessment

Newly made recommendations are coloured in pink

*Suggestion for centres that have adequate resources to implement and are interested in research with speech oucomes.



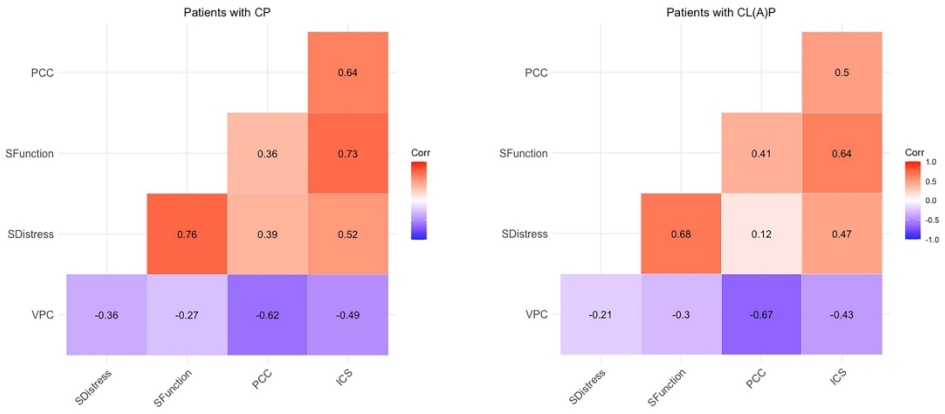


Figure 2. Correlations in patients with CP and CL(A)P. All correlations in both cleft types appeared significant (P<0.05), except for the correlation between the PCC and CLEFT-Q SDistress in patients with CL(A)P (p = 0.285). Note that VPC is inversely scored (higher numbers correspond to worse outcomes), thus accounting for the negative correlations with the other scales.

270x119mm (300 x 300 DPI)

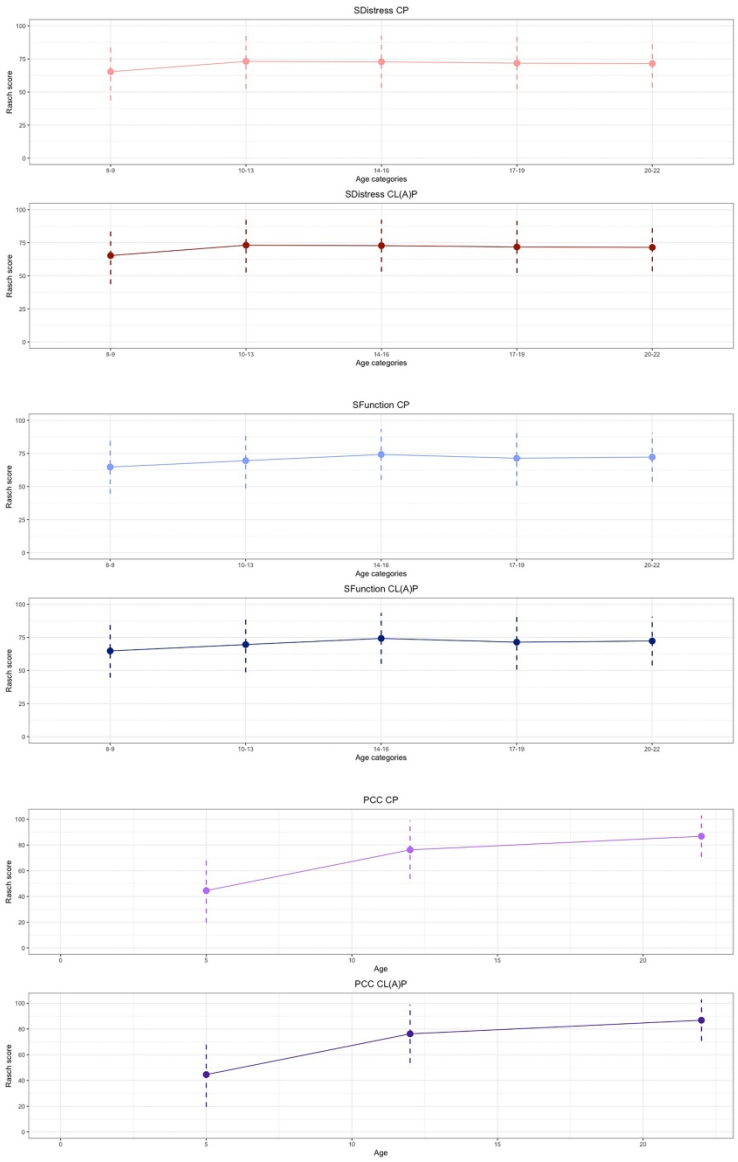
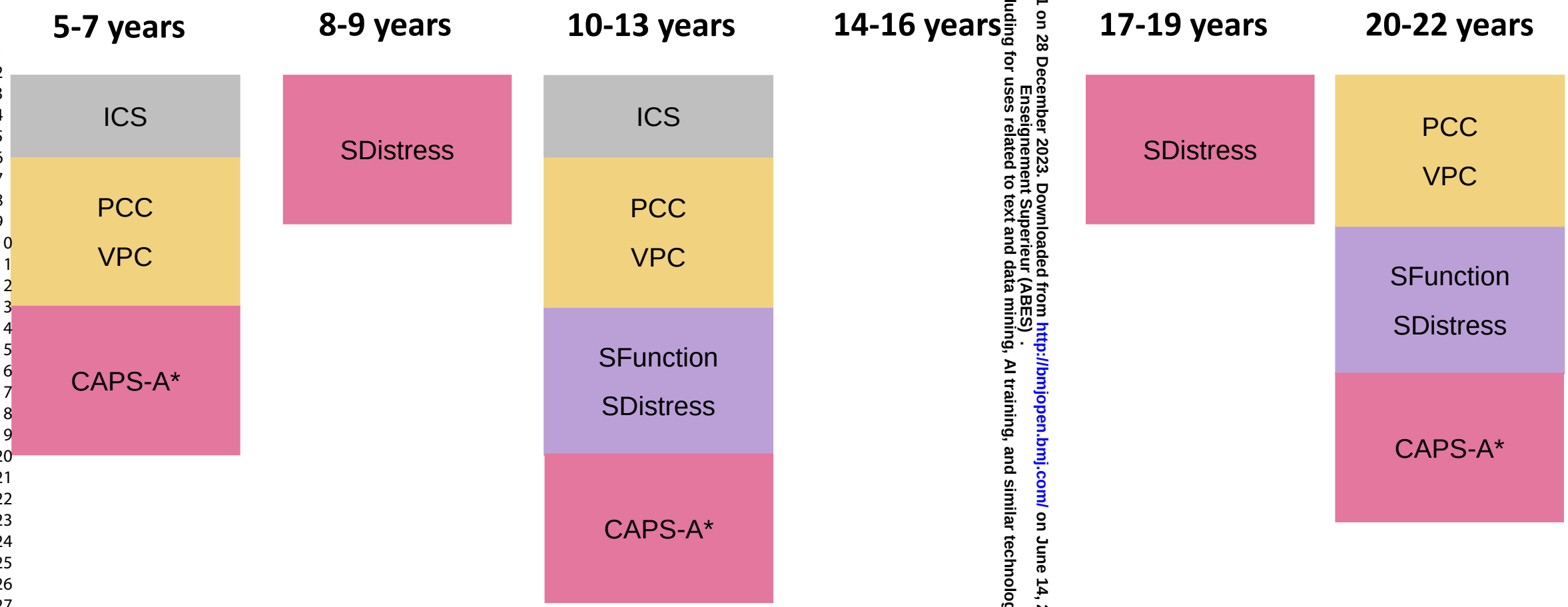


Figure 3. cross-sectional trend analyses of the age-groups. Analyses are presented per outcome measure, per cleft type.

199x299mm (300 x 300 DPI)



Appendix 1: differences of outcome scores based on income status of the country

	Income	Beta	SE	P-value	Difference between groups	P-value
Speaking -Related Distress	**High (ref)** **Middle**	72.10 -16.39	0.51 1.58	<0.001*	High VS Middle	<0.001*
	High (ref) **Low**	72.10 -13.08	0.51 1.68	<0.001*	High VS Low	<0.001*
					Middle VS Low	0.285
Speech Function	**High (ref)** **Middle**	70.87 -12.32	0.52 1.64	<0.001*	High VS Middle	<0.001*
	High (ref) **Low**	70.87 -13.18	0.52 1.71	<0.001*	High VS Low	<0.001*
					Middle VS Low	0.923

Univariate analyses. Outcome scores of patients deriving from countries with a high income status were taken as reference. *Significant difference

Appendix 2: descriptives for subgroup analyses

	CP (n=762)	CL(A)P (n=1379)	Total (n=2141)
Gender			
Male	308 (40.4%)	904 (65.6%)	1212 (56.6%)
Female	454 (59.6%)	475 (34.4%)	929 (43.4%)
SDistress			
Number	517 (67.8%	1068 (77.4%)	1585 (74.0%)
Missing	245 (32.2%)	311 (22.6%)	556 (26.0%)
SFunction			
Number	519 (68.1%)	1076 (78.0%)	1595 (74.5%)
Missing	243 (31.9%)	303 (22.0%)	546 (25.5%)
VPC			
Number	173 (22.7%)	213 (15.4%)	386 (18.0%)
Missing	589 (77.3%)	1166 (84.6%)	1755 (82.0%)
ICS			
Number	169 (22.2%)	228 (16.5%)	397 (18.5%)
Missing	593 (77.8%)	1151 (83.5%)	1744 (81.5%)
PCC score			
Number	152 (19.9%)	192 (13.92%)	344 (16.1%)
Missing	610 (80.1%)	1187 (86.1%)	1797 (83.9%)

Descriptives of patients deriving from high income countries, including all patients from the ICHOM centers.

Appendix 3: maximum scores and VPC overview

		Number of measurments with max score	Total measurements
CP	SDistress	18.8% (97)	517
	SFunction	19.8% (103)	519
	ICS	29.0% (49)	169
CL(A)P	SDistress	15.2% (162)	1068
	SFunction	15.9% (172)	1076
	ICS	18.0% (41)	228

Maximum scores

	CP				CL(A)P		
	Competent (0)	Marginally incompetent (1)	Incompetent (2)		Competent (0)	Marginally incompetent (1)	Incompetent (2)
5 years	43 (50%)	21 (24.4%)	22 (25.6%)		11 (15.5%)	17 (23.9%)	43 (60.6%)
12 years	39 (65%)	19 (31.7%)	2 (3.3%)		46 (52.3%)	28 (31.8%)	14 (15.9%)
22 years	11 (40.7%)	13 (48.1%)	3 (11.1%)		26 (48.1%)	19 (35.2%)	9 (16.7%)

VPC overview

BMJ Open

What is the optimal assessment of speech? A multicenter, international evaluation of speech assessment in 2500 patients with a cleft.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-071571.R1
Article Type:	Original research
Date Submitted by the Author:	04-Oct-2023
Complete List of Authors:	Ombashi, Saranda; Erasmus Medical Center, plastic and reconstructive surgery Kurniawan, Melissa; Erasmus Medical Center, plastic and reconstructive surgery Allori, Alexander; Duke University Hospital & children's health center, Plastic, Maxillofacial & Oral Surgery Sharif-Askary, Banafsheh; Duke University Hospital & children's health center, Plastic, Maxillofacial & Oral Surgery Rogers-Vizena, Carolyn; Boston Children's Hospital, Plastic and Oral Surgery Koudstaal, Maarten; Erasmus MC Sophia Children Hospital, Department of Oral and Maxillofacial Surgery Franken, Marie-Christine; Erasmus Medical Center, Ear, Nose and Throat, Speech Therapy Mink van der Molen, Aebele; University Medical Center Utrecht, Department of Plastic Surgery Mathijssen, Irene; Erasmus MC Sophia Children Hospital, Department of Plastic and Reconstructive Surgery Klassen, Anne; McMaster University, Pediatrics Versnel, Sarah; Erasmus MC Sophia Children Hospital, plastic and reconstructive surgery
Primary Subject Heading:	Paediatrics
Secondary Subject Heading:	Dentistry and oral medicine, Evidence based practice, Paediatrics, Patient-centred medicine, Surgery
Keywords:	ORAL & MAXILLOFACIAL SURGERY, Paediatric head & neck surgery < PAEDIATRIC SURGERY, Paediatric oral & maxillofacial surgery < PAEDIATRIC SURGERY, Paediatric plastic & reconstructive surgery < PAEDIATRIC SURGERY, PLASTIC & RECONSTRUCTIVE SURGERY, Quality of Life

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

What is the optimal assessment of speech? A multicenter, international evaluation of speech assessment in 2500 patients with a cleft.

S. Ombashi[1,2], M. Kurniawan[1,], A.C. Allori[3], B. Sharif-Askary[3], C.R. Rogers-Vizena[4], M.J. Koudstaal[2, 5, 6], M.C. Franken[7], A.B. Mink van der Molen[2,8] , I.M.J. Mathijssen [1,2,6], A.F. Klassen[9] , S.L. Versnel [1,2,6]

[1] Department of Plastic and Reconstructive Surgery, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands

[2] European Reference Network for rare and/or complex craniofacial anomalies and ear, nose and throat disorders

[3] Department of Plastic, Maxillofacial & Oral Surgery, Duke University Hospital & Children's Health Center, Durham, North Carolina, United States of America

[4] Department of Plastic and Oral Surgery, Boston Children's Hospital, Boston, Massachusetts, United States of America

[5] Department of Oral and Maxillofacial Surgery, Special Dental Care and Orthodontics, Erasmus University Medical Center, Rotterdam, The Netherlands

[6] Dutch Craniofacial Center, Erasmus University Medical Center, Rotterdam, The Netherlands

[7] Department of Ear, Nose and Throat, Speech Therapy, Erasmus University Medical Center, Rotterdam, The Netherlands

[8] Department of Plastic and Reconstructive Surgery, Univeristy Medical Center Utrecht, Utrecht, The Netherlands

[9] Department of Pediatrics, McMaster University, Hamilton, Canada

Numer of words: 4436

Number of tables: 2

Number of figures: 3

Number of appendices: 4

Corresponding Author: S. Ombashi, MD. Dr. Molewaterplein 40, 3015 GD Rotterdam, Erasmus University Medical Center; Tel: +316-31989928; E-mail: s.ombashi@erasmusmc.nl

ABSTRACT

Objectives

Speech problems in patients with a cleft palate are often complex and multifactorial. Finding the optimal way of monitoring these problems is challenging. ICHOM(International Consortium of Health Outcomes

Measurement) developed a set of standardized outcome measures at specific ages for patients with a cleft lip and/or palate, including measures of speech assessment. This study evaluates the type and timing of speech outcome measures currently included in this ICHOM. Additionally, speech assessments in other cleft protocols and initiatives are discussed.

Design, setting and participants

An international, multicenter study was set up including centers from the United States and The Netherlands. Outcomes of clinical measures and Patient Reported Outcome Measures(PROMs) were collected retrospectively according to the ICHOM set. PROM data from a field test of the CLEFT-Q were collected, including participants from countries with all sorts of income statuses, to examine the value of additional moments of measurement that are used in other cleft initiatives.

Data from 2500 patients was included. Measured outcomes contained univariate regression analyses, trend analyses, T-tests, correlations and floor and ceiling effects.

Results

PROMs correlated low to moderate with clinical outcome measures. Clinical outcome measures correlated low to moderate with each other too. In contrast, two CLEFT-Q scales correlated strongly with each other. All PROMs and the Percent Consonants Correct(PCC) showed an effect of age. In patients with an isolated cleft palate, a ceiling effect was found in the intelligibility in context scale(ICS).

Conclusion

Recommendations for an optimal speech outcome assessment in cleft patients are made. Measurement moments of different cleft protocols and initiatives are considered in this proposition. Concerning the type of measures, adjustment of the current PCC score outcome

seems appropriate. For centres with adequate resources and specific interest in research, translation and validation of an upcoming tool, the CAPS-A, is recommended.

Strengths and limitations of this study

- **International, multicenter setting**
- **Data analyses per cleft type and age group**
- **Cross-sectional data analyses**
- **2500 participants with a cleft**
- **Evaluating both PROMs and clinical outcome measures**

INTRODUCTION

A cleft lip and/or palate (CL/P) is the most common congenital craniofacial anomaly, with varying incidence rates among Asians (1:500), Caucasians (1:1000) and patients of African descent (1:2500) (1-4). Causes of a cleft are multifactorial, as both environmental and genetic factors have been reported(4). Clefts can be categorized in multiple classification systems, of which a commonly used classification system includes 4 cleft types: a cleft lip (CL); a cleft lip and alveolus (CLA); a cleft palate (CP); a cleft lip, alveolus and palate (CL(A)P)(5). In addition, clefts can occur unilaterally or bilaterally(3).

Due to the facial defects, functional and appearance-related problems can occur, of which the extent may depend on the cleft type; the severity of the cleft; and the coping of the individual and his/her environment(6). Functional problems such as speech problems, hearing impairment and orodental problems are often reported. As a result of the latter, difficulties with eating, drinking and breathing can occur as well (5, 7).

Given the broad range of problems a patient with a cleft may have to face, treatment of patients with CL/P is ideally done in a specialized and multidisciplinary cleft team in which speech therapists; maxillofacial and plastic surgeons; otolaryngologists; pediatricians; psychologists; orthodontists; geneticists; and specialized nurses are involved(7). Treatment and monitoring patients with CL/P consists of multiple surgical interventions to close the defect and to improve appearance if the patient desires so. Follow up of hearing function is indicated in case of a cleft palate, and placement of moppets is regularly done if necessary. Furthermore, psychological guidance is often indicated while the child grows up. Moreover, speech monitoring and long-term, intensive speech therapy is often necessary to improve the eligibility of the child(5, 7).

The development of speech is often complex in patients with a cleft palate (with or without a cleft lip, CP±L). Persistent velopharyngeal incompetence, residual fistula, adenoid atrophy, surgical intervention

and hearing problems influence speech disorder severity in this population(8-12). Speech problems in patients with CP±L can have a large impact on an individual's life, as proper speech skills play an essential role in activities, social functioning and participation in society(13). Many treatment pathways are focused on speech improvement to ameliorate Quality of Life (QoL)(14). Logically, speech assessment is an important parameter in cleft care.

However, no consensus has been reached regarding best diagnostic speech outcome measures and their timing in this population(5). Developing scientifically solid instruments to assess speech in an objective manner is complicated, because listener's perception of speech deficits, even by experts, may differ substantially(15). An additional challenge is systematic assessment of the patient's perspective, which is essential to include due to the impact of speech problems on the individual(16). Although widely-accepted agreement seems essential for improvement of cleft care, finding consensus is complex, especially since speech outcomes should be comparable between different languages to facilitate international collaboration.

Recently, the International Consortium for Health Outcomes Measurement (ICHOM) developed the ICHOM Standard Set for Cleft Lip and Palate (ICHOM Standard Set), with different pathways for varying cleft types(5). Based on patient and expert consensus, a minimal, accessible set of outcome measures was established to enable benchmarking between cleft centers in a systematic manner. For speech assessment, an outcome set was included with both clinical measures and Patient Reported Outcome Measure (PROMs), being the patient's and parent's perspectives.

So far, the selected standardized speech outcome measures and their timing have not been evaluated. As an increasing number of centers are implementing this set, it is important to critically evaluate and

optimize this ICHOM Standard Set. Three centers, the Boston Children's Hospital (Boston, USA), Duke University Hospital (Durham, USA), and the Erasmus Medical Center (Rotterdam, The Netherlands), started clinical implementation and an international collaboration in 2015. The overarching aim of this collaboration is to share data and knowledge obtained by using the set in standard care. Additionally, they collaborate with McMaster University (Hamilton, Canada), who developed and field tested the CLEFT-Q questionnaire, of which many scales are included in the ICHOM Standard set.

The objective of this study was to evaluate the current standardized speech outcome measures of the ICHOM Standard Set for patients with CP±L. More specifically, the value of every speech outcome measure was examined, as well as the best age intervals for assessment of these outcome measures. In addition, other speech assessment tools are discussed. Finally, recommendations are made for an optimal and complete assessment of speech in patients with CP±L, that is efficient and accessible for all cleft centers.

METHODS

Patient population

Three centers (Boston Children's Hospital, Duke University Hospital, and Erasmus Medical Center,) each implemented the ICHOM Standard Set in 2015. All patients treated at these centers for a cleft palate with a cleft lip/cleft alveolus (CL(A)P), or an isolated cleft palate (CP) who were assessed according to the ICHOM Standard Set (age range 5-22 years), were included. In addition, another patient group derived from an international field test of the CLEFT-Q, by McMaster University(17). According to the age cut-off of the ICHOM Standard Set, only outcomes from CLEFT-Q scales of field test patients with a CP±L up to 22 years old were included in the current study(appendix 1). Patients from the participating centers were excluded in case they could not sufficiently speak or write the language native to the center's country .

Patient and public involvement

In the development of the ICHOM Standard Set, patients were actively involved. The ICHOM Standard Set was implemented in each center as part of regular clinical care. Data was pseudonymized and collected retrospectively, and ethical approval was obtained to do so without explicit consent from each patient and parent. Results of this study may be of use to further improve the currently used ICHOM Standard Set, and therewith regular clinical care.

Outcome measures

PROMs

CLEFT-Q Scales

The CLEFT-Q is developed specifically to assess QoL from the patient's perspective in patients with a CP±L. A literature review, patient interviews and psychometric testing, established the final content of the scales, which covers several overarching domains(18-20). Speech is assessed through two scales, each covering a different domain. Both scales have 3 response options for each item (always; sometimes; never); a lower score equals a worse outcome. Completing the scales can be done online; it will take the patient several minutes.

Speaking-Related Distress (SDistress) is part of the psychosocial domain. The scale contains 10 items that relate to the psychosocial part of speech difficulties, like nervousness or frustration(20).

Speech Function (SFunction) focuses on the functional speech difficulties that patients themselves identify, for example the ability to say certain letters or words. The scale consists of 12 items that belong to the facial function domain(20).

Intelligibility in Context Scale(ICS) is a measure that assess the intelligibility of the child. It is a 7-item, parent-reported questionnaire designed to be scored by speech pathologists. The score indicates a child's level of functional intelligibility, by assessing the degree to which the speech of the patient is understood by different communication partners. The total score is calculated by the averages of the items completed. ICS appeared to be a valid and reliable tool for children with speech disorders, (21, 22), but not specifically designed nor validated for patients with CP±L(23). It is available in several languages, and normative data exists for English-speakers(24, 25).

Clinical outcome measures

Percent Consonants Correct(PCC) is developed to detect speech sound errors. PCC scores are calculated by using a standard, crossecitonally translated set of words that include all speech problems children with CP±L often tend to have.

In case of any problems, their severity can be categorized: PCC scores of 85-100% indicate mild to no problems; scores of 65-84.9% indicate mild-moderate problems; scores of 50-64.9% indicate moderate-severe problems; and scores <50% indicate severe problems(20). PCC is suitable for usage in patients with CP±L when assessed by well-trained clinicians(8).

Velopharyngeal Competence rating(VPC) discriminates between three categories: 'competent', 'marginally incompetent' and 'incompetent'. The outcome is determined by the speech therapist based on the PCC test and spontaneous speech. In case of any clinical evidence of minor problems regarding the competence, VPC was categorized as 'marginally incompetent'. When clinically significant problems were detected, suggesting surgical management and/or speech therapy, VPC was categorized as 'incompetent'. Prior studies found VPC to be suitable as a first clinical choice for the assessment of velopharyngeal dysfunction and is recommended for both clinical follow-up and research(26).

Data collection

All participating centers obtained ethical approval for the current study from their local ethics committees.

Data was collected restrospectively over a 6 year period (2015-2020) and extracted from the electronic patient files in 2018 and 2020 (as a data update). Both video and audio records were used for the evaluation of the clincal outcome measures. During the data collection period, the included center cooperated together, and regular meetings were held (both online and live).

According to the ICHOM protocol, both CLEFT-Q Speech scales were assessed at ages 12 and 22 years(appendix 1). Both PCC and VPC were scored at ages 5, 12 and 22 years, and ICS at ages 5 and 12 years.

The field test cross-sectionally collected data from patients with a cleft across 12 different countries with different income-statuses(17). As 8 years is the minimum age to complete the CLEFT-Q, both CLEFT-Q Speech scales from field test patients with CP±L from 8-22 years old were included (appendix 1).

Income status of the country according to the World Bank Classification was made within the field test. Data from the ICHOM centers were all categorized as deriving from high-income countries.

Baseline characteristics that were collected included gender, type of cleft, and age at the time of assessment.

Data analysis

Data was analysed in R-studio, a free software environment for statistical computing and graphics(27). Psychometric validation of the SDistress and SFunction confirmed suitability to use a 0 to 100 scale deriving from the sum scores for analysis(17).

For analysis of the ICS questionnaire, the average score of the seven items was used. VPC was used as an ordinal variable, whereas PCC scores were expressed as proportions.

All participating ICHOM centers are high-income countries, whereas part of the field test data was collected in upper middle and lower middle income countries. To prevent possible influence of income-status on the outcomes, univariate regression analyses were used to examine differences in outcome scores of the SDistress and SFunction before further analyses. Data was categorized according to the income status of the country where the data had been collected.

In order to examine the added value of each PROM and clinical outcome measure in regard with the other measures, correlations were examined between the PROMs; between the clinical outcome measures; and between the PROMs and the clinical outcome measures. Pearson correlations were used, and outcomes were analysed per cleft type. Correlations were considered strong in case r>0.7; moderate between r=0.5 and R= 0.7; and weak in case r<0.5.

Analyses within and between different age-groups were done to explore whether the current outcome measures are assessed at the optimal age-points, and whether additional measurement moments are indicated either for PROMs or clinical outcome measures. Therefore, not only time-points of the ICHOM protocol were included in analyses. As CLEFT-Q outcome scores of all ages between 8-22 years were included from the field test data, time-points used by other large initiatives as Eurocleft, Scandcleft and Americleft were considered as well (28-30). Doing so, the following age-groups were set up: 5-7 years; 8-9 years; 10-13 years; 14-16 years; 17-19 years, 20-22 years(app. 2).

Per age-group, possible differences between scale scores were examined with independent T-tests. Bonferroni correction was applied for multiple testing. Trend analyses were performed to identify potential problems in specific age-groups.

Floor- and ceiling effects were examined to identify the suitability of the outcome measures in our population. A floor or ceiling effect is seen when a considerable amount of the outcome scores are either scored the best (in this case a maximum score, thus a ceiling effect), or the worse (in this case a minimum score, thus a floor effect). Both effects result in a truncated distribution of the outcomes on either side of the scale (31, 32). Minimum and maximum score outcomes of all PROMs and the PCC were evaluated. A percentage of 20% or more of the patients scoring the minimum or maximum outcome score was considered as a ceiling effect. In VPC, the outcome distributions were examined.

RESULTS

Characteristics of the included participants

A total of 2500 patients were included in the study; 1723 derived from the field test, and 777 from the ICHOM centers (table 1).

Table 1: Demographics and phenotypes

Descriptives total

	Field test (n = 1723)	ICHOM (n = 777)	Overall (n = 2500)
Sex			
Male	981 (56.9%)	444 (57.1%)	1425 (57.0%)
Female	742 (43.1)	333 (42.9%)	1075 (43.0%)
Cleft Type			
CP	517 (30.0%)	301 (38.7%)	818 (32.7%)
CL(A)P	1206 (70.0%)	476 (61.3%)	1682 (67.3%)
Income classifications			
High Income	1364 (79.2%)	777 (100%)	2141 (85.6%)
Up Middle income	199 (11.5%)	0 (0%)	199 (8.0%)
Low Middle Income	160 (9.3%)	0 (0%)	160 (6.4%)

There were slightly more males than females, and relatively more patients with a CL(A)P than with a CP. Significant differences between countries with a High and Upper-/Lower-Middle income status of the field test were found and results are shown in appendix 2.

Therefore, further analyses were done only with the patient population of the field test deriving from countries with a high income status, like all participating ICHOM Cleft centers (n = 2141). The subgroup characteristics are included in appendix 3.

Associations between the outcome measures

Correlations between all outcome measures (clinical and PROMs), in both cleft types (CP and CL(A)P), appeared significant (P<0.05), except for the correlation between PCC and SDistress (p = 0.285)(fig. 1).

Correlation PROMs: The SDistress and SFunction showed a strong correlation in patients with CP (r = 0.76) and a moderate correlation in patients with CL(A)P (r = 0.68). The ICS and SFunction correlated strongly (r = 0.73) in patients with CP, and moderately (r = 0.64) in patients with CL(A)P; whereas the ICS and the SDistress correlated moderately in patients with CP (r = 0.52) and weak in patients with CL(A)P (r = 0.47).

Correlation clinical outcomes: VPC and PCC were (negatively) moderately correlated in both cleft types (r = -0.62 and r = -0.67 in patients with CP and CL(A)P respectively).

Correlation PROMs and clinical outcomes: Moderate correlations were found between the PCC and ICS in patients with CP (r = 0.64), and in patients with CL(A)P (r= 0.5). VPC and ICS had a (negative) weak correlation in patients with CP (r = -0.49), and CL(A)P (r = -0.43). The SDistress and SFunction were weakly correlated with VPC and PCC (negatively) in both cleft types (fig. 1).

Comparing outcome measures between age-groups

SDistress and SFunction: showed the highest mean outcome (i.e. the most favourable ratings) in the age-group of 14-16 years old. From thereon, a slightly downward trend is seen (fig. 2). In both CLEFT-Q Speech scales the lowest mean outcome scores were found in the age-group of 8-9 years old, which was significantly different in comparison with the other age-groups in patients with CL(A)P (p<0.05, table 2).

Table 2 Mean outcomes per age-group, per cleft type

		Age categories	SDistress	SFunction	ICS	PCC
Cleft Palate		5 – 7			4.08 (0.73)	72.49 (31.14)
		8 – 9	70.44 (20.64)	66.12 (22.12)		
		10 – 13	74.84 (19.91)	72.56 (21.52)	4.45 (0.62)[1]	92.02 (11.48)[1]

	14 – 16	75.10 (22.77)	74.66 (24.32)		
	17 – 19	70.59 (16.92)	73.00 (19.02)		
	20 – 22	70.27 (21.93)	71.69 (24.98)		90.92 (15.33)[1]
	5 – 7			4.03 (0.50)	46.06 (25.59)
Cleft Lip Palate	8 – 9	65.32 (21.61)	64.75 (20.16)		
	10 – 13	73.16 (20.62)[1]	69.57 (20.96)	4.32 (0.55)[1]	76.25 (22.63)[1]
	14 – 16	72.81 (19.66)[1]	74.22 (19.14)[1]		
	17 – 19	71.82 (19.67)[1]	71.41 (20.79)[1]		
	20 – 22	71.49 (18.00)	72.26 (18.40)[1]		86.83 (16.13)[1,2]

The variables in the table present the mean outcomes of the SDistress, SFunction, ICS and PCC. Outcomes were categorized into cleft type and age group.

1 $p < 0.05$ compared to 5-7 or 8-9 age

2 $P < 0.05$ compared to 10 – 13 age

ICS : Both patient groups showed a significant difference between the two age-groups 5 and 12 years, ICS was significantly lower at 5 years than at 12 years in both cleft types (tab. 2).

PCC: Observing the trends in the clinical outcome measures, an upward trend regarding PCC score was seen (fig. 2). In both cleft types, PCC scores differed significantly between the age-groups(tab. 2).

VPC: In the age-group of 5 years, 25.6% of the patients with CP and 60.6% of the patients with CL(A)P were scored as incompetent. In 22 year-olds, this percentage was 11.1% in patients with CP and 16.7% in patients with CL(A)P.

No floor effects were found in any of the PROMs. In patients with CP, the ICS showed a ceiling effect (29.0%, n = 169). No ceiling effects were observed in patients with CL(A)P. An overview of all maximum scores and the VPC score distribution is shown in appendix 4.

DISCUSSION

Evaluation of the value of the current ICHOM speech outcome measures

All correlations between PROMs were moderate, except for the strong correlation of the SFunction with both the SDistress and the ICS in patients with a CP. The fact that the correlation between the SFunction and SDistress is stronger in patients with CP than in patients with CL(A)P, suggests that the visibly different appearance in patients with CL(A)P plays a significant role in SDistress as well; in a social context, looking differently may cause additional or more distress besides having speech problems. This is supported by our finding that the ICS correlated moderately with SFunction, but weakly with SDistress in the CL(A)P group. Parent reported speech intelligibility correlated higher to children's self report of their speech function, than it did to the speech distress the children themselves experience. In the latter, distress about appearance could be included. This finding suggests that the ICS can give an indication of 'patient-reported' SFunction in young children who cannot complete a PROM themselves yet (7 years and younger).

The PROMs showed weak correlations with the clinical reported outcomes; except for the moderate correlation that was seen between the ICS and the PCC in both patient groups. Based on these findings, PROMs appear to be of added value, as they provide different information than captured with the clinical outcome measures included in the Standard Set. They add a unique dimension to speech outcome measurement – a subjective dimension related to the patient's experiences with everyday speaking situations. While clinical measures objectively appraise the quality of speech, they will probably be insufficient to adequately capture the more nuanced social, emotional, and psychological aspects of SDistress and SFunction. With this additional self-report and parental information, clinicians

can more comprehensively explore the patients' problems concerning speech in order to find out whether additional treatment or guidance is indicated.

Evaluation of the impact of age of assessment on measurement outcomes

In both CLEFT-Q Speech scales, the age-group of 8-9 years enholds the worst scores. Speech improvement due to speech therapy or late closure of the hard palate (in certain protocols around the age of 9 years when alveolar bone grafting is performed), might explain the higher, better scores in the age-groups of 10-13 and 14-16 years. In age-groups 17 and up however, CLEFT-Q scores appeared to decline whereas PCC scores improved. This finding suggests that (almost) adult patients with CP±L develop feelings of insecurity concerning their speech, although their speech sound production remains good, or even improves. This is in line with speech therapists' experiences in the outpatient clinic, where patients were seen in person at the age of 22, but not at age of 17-19. Quite often, when discussing outcomes of the CLEFT-Q scales as well as the PCC with the patient, (s)he reacted surprised when told that no (cleft-related) problems were present in their speech. Taking the lower CLEFT-Q scores in 8-9-, 17-19- and 20–22-year-olds that were found in the field test into consideration, additional assessment of a PROM at the age-groups of 8-9 (youngest age at which this PROM can be assessed) and 17-19 years old should be considered for implementation in the ICHOM Standard Set. Therewith, monitoring patients more closely will be enabled, and any concerns of patients with CP±L regarding their speech can be discussed timely.

The two CLEFT-Q speech scales showed to capture overlapping information as they strongly correlate in patients with CP. Questions deriving from the SDistress are not measurable in any other manner, whereas SFunction from the patient's perspective might be less of added value for a PROM questionnaire. Therefore, implementation of the CLEFT-Q SDistress scale in patients with both cleft types is recommended at the age-groups of 8-9 years and 17-19 years (fig. 3).

A ceiling effect in **ICS** outcomes of patients with CP, without clear differences between average scores in patients with CP and CL(A)P, suggests that the group with CP contains a diverse population and severity of the speech problems vary widely. Furthermore, since ICS is not specifically developed for a population with CP±L, it is debatable whether this tool captures the information necessary to point out all relevant speech problems in the patient group.

However, exclusion of ICS could mean that a large part of the speech problems in the population with CP would remain undetected. Assessment at 5 and 12 years in patients with both cleft types, which is the current timing in the ICHOM Standard Set, appears therefore appropriate despite the ceiling effect.

Although **VPC** scores were relatively favourable in patients with CP, no changes regarding the implementation of the VPC scores are recommended as the outcomes showed to vary. VPC can serve as a suitable screening tool and outcomes are easily gathered by the observation of a clinician. Hence, patient-burden is low and the tool efficiently detects any velopharyngeal problems.

PCC scores that were found indicated speech sound problems especially in the younger age-groups of the patients with CL(A)P. 22-year-olds with both cleft types showed mild speech sound problems in general. Therefore time points as currently implemented in the ICHOM Standard Set appear adequate. In contrast, the suitability of PCC-assessment in a cleft set focusing on standardized outcome measures is still debatable, as inter- and intra-center reliability have not been investigated thoroughly in all participating centers so far(15). Future research should include an examination of scoring and interpreting PCC scores in different centers and/or different countries.

Future considerations regarding alternative speech outcome measures

In order to establish an optimal cleft set for speech assessment, other standardized outcome measures should be considered. Based on clinical experience with ICHOM Standard Set, possible suggestions for additional outcome measures are discussed here.

Regarding **PROMs** for speech assessment in patients with CP±L, the CLEFT-Q scales seem to be the most suitable PROMs available. Their comprehensive psychometric examination and cross-cultural character make them accessible for all cleft centers that seek an efficient minimal cleft set that comprises all important speech parameters(17-19). The standardized approach for translation and validation of the CLEFT-Q questionnaire enables accessibility of the PROM even for centers that still need to translate the CLEFT-Q into their native language(33, 34). Another cleft-specific PROM is the Cleft Hearing and Speech Questionnaire (CHASQ). Whereas the psychometric properties of the CLEFT-Q were examined throughout Rasch measurement theory, classical test theory was used for the CHASQ(35). A recent cross-sectional questionnaire study that compared the CLEFT-Q with the CHASQ, found that the majority of the patients with CP±L preferred the CLEFT-Q(35). Therefore, implementation of the CHASQ speech does not seem to be of added value in the current cleft set.

Besides the used **VPC** measure, a more elaborate variant exists, namely the VPC-Summary (VPC-Sum). This includes assessment of hypernasality, passive VPI symptoms and the transcriptions of active non-oral consonant errors(36). VPC-sum can either be reported as a score between 0 and 6, or as a dichotomized outcome (velopharyngeal competence or incompetence)(36).

Calculation of the VPC-Sum is based on single-words, whereas VPC-rate is based on observation of spontaneous speech (37). VPC-Sum would be an interesting measure due to its efficiency, although it may not be achievable to implement VPC-Sum in all centers in the near future as only five different

languages are currently available(31). Other alternatives such as nasopharyngoscopy or MRI are invasive, expensive and enlarge the patient burden(38), and therefore not easy accessibility for all centers.

The currently implemented **PCC** lacks any categorization of consonant errors. The **Eurocleft speech group** created a research protocol with a phonetic framework, which was used in six centers and five different languages(39). It also included consonant production, but assessed on sentence level instead of single words. It categorized into 3 groups (correct; almost correct; incorrect). Further division into 21 error categories that were sampled in five groups was done in case of incorrect consonants (Nasal airflow; glottal realisations; alveolar deviations; sibilant deviations; other)(39). Moreover, general speech quality was assessed concerning hyper- and hyponasality, and voice quality(40). Expert rating of these outcomes requires periodic training of sufficient interrater reliability. However, it might be too detailed for implementation in an efficient, clinically oriented cleft set. Therefore, we suggest to further categorize the PCC score, although not as detailed as in the Eurocleft studies. Based on clinical experience with the ICHOM Standard Set, it is recommended that speech pathologists report whether any cleft related, phonological, or phonetic problems are detected.

Another clinical outcome measure, the **Great Ormond Street Speech Assessment 1998** (GOS.SP.ASS'98), provides a comprehensive view of all speech associated features for patients with CP±L(41, 42). Its suitability for inter-center comparison would make it interesting for the ICHOM Standard Set(5); however, it is too detailed for clinical audit(43). In succession the **Cleft Audit Protocol for Speech Augmented** (CAPS-A) was developed for cleft-related problems, and could be an alternative for PCC(44). Seen its rigorous psychometric assessment, it fits well into a set that seeks standardized outcome measures. The Americleft Speech Project found that an acceptable inter- and intrarater reliability can be

achieved(43, 45). Furthermore, it is suitable for assessment in 5-years-olds, which enables detection of speech problems at an earlier age(46). However, the CAPS-A is limited in types of statistical analyses due to the scaling type used (equal appearing interval)(47). A more practical challenge concerning implementing the CAPS-A would be the required training of all involved speech therapists, and the amount of time the assessment takes (15 min.)(44). Moreover, the CAPS-A is developed and applicable for English-speaking countries, necessitating translation and validation in other languages(43). The CAPS-A is not ideal for centers interested in a minimal and efficient cleft set. However, centers with experience and resources are highly recommended to implement this tool in order to promote further international standardization of elaborate speech assessment in patients with CP±L(fig. 3). Implementation of the CAPS-A would also enable the use of the recently developed and validated CAPS-A-VPC-Sum score to reliably measure velopharyngeal function(48). Our suggestion for centers that consider the implementation of the CAPS-A, is to assess it at the ages of 5-7, 10-13 and 20-22 years in order to enable long-term follow-up.

Limitations of the study

Data was analyzed cross-sectional. Longitudinal analyses to explore development of speech and for benchmarking will be possible in the future, since data collection continues. Moreover, because this study included data from the CLEFT-Q field test, a higher number of outcome data from the CLEFT-Q scales was available for analyses than from the other outcome measures included in the ICHOM Standard Set.

CONCLUSION

From the current study, it can be concluded that the current ICHOM Standard Set is informative and efficient. PROMs were shown to be of added value, and the CLEFT-Q appeared to be the most suitable PROM. Therefore, continuation of collecting the current outcome measures and time points is recommended. Furthermore, a minor extension is suggested: In addition to the current time-points of assessment, it is recommended to implement the CLEFT-Q SDistress scale at the age of 8-9 and 17-19 as well. Further adjustments of the set could comprise an additional categorization of the PCC score, based on the framework of Eurocleft and adjusted for clinical usage.

ACKNOWLEDGMENTS

We would like to thank all children and caregives that participated in the study, that contributed by filling out the questionnaires and allowing the research team to use their data for the current study. Furthermore, we would like to thank the speech pathologists from all our participating centers, that contributed tremendously by collecting the data consequently and by advising the research team on important clinical considerations

Contributor statement

Saranda **Ombashi**: Data collection in Dutch Center; correspondence to other centers; data analyses; writing and adjusted the manuscript based on advice of the other authors.*

Melissa S.I.C. **Kurniawan**: Data analyses, co-writer on the methods and results section of the manuscript.*

Alexander C. **Allori:** Data collection at Duke University Hospital, advisor of project plan and manuscript**.***

B. **Sharif-Askary:** Data collection at Duke University Hospital, advisor of project plan and manuscript.*

Carolyn R. **Rogers:** Data collection at Harvard University Hospital, advisor of project plan and manuscript.*

Maarten J. **Koudstaal**: Writer of the Ethical Board Approval, Supervision of data analyses. Advisor and supervisor on project plan and manuscript.*

M.C. **Franken:** Data collection at Erasmus Medical Cewnter, advisor of project plan and manuscript, with special focus on the clinical content.*

A.B. **Mink van der Molen**: advisor of project plan and manuscript.*

Irene M.J. **Mathijssen:** Overall supervision of project plan and manuscript.*

Anne F. **Klassen:** Data collection Field Test Mc Master University, advisor on project plan and manuscript.*

Sarah L. **Versnel:** Project leader, correspondence with all other centers. Supervision on project plan, data analyses and manuscript.*

*All authors approved the final version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved

Competing interests:

The CLEFT-Q is owned by McMaster University and the Hospital for Sick Children. Prof. Anne Klassen is a co-developer of the CLEFT-Q and would receive a share of any licence revenues as royalties based on her institutions' inventor sharing policy if the CLEFT-Q was used in a for profit study.

The other authors have no disclosures.

Funding

This project received no funding. However, S. Ombashi works as a PhD-student for an European Union-funded Network, the 'Eureopean Reference Network for Craniofacial anomalies and Ear, Nose and Throat disorders'. The results and findings from this project are of help in further European alignment concerning standardized outcome measures in cleft care.

Data sharing statement

A data sharing agreement was set up and signed by all participating centers that shared data for this project.

Ethics approval statement

All participating centers obtained ethical approval for the current study from their local ethics committees (ethics committees of the Erasmus University Medical Center; Utrecht University Medical Center; Duke University Hospital & Children's Health Center; Boston Children's Hospital; McMaster University) IRB approval nr MEC-2016-156.

References

1. Hlongwa P, Levin J, Rispel LC. Epidemiology and clinical profile of individuals with cleft lip and palate utilising specialised academic treatment centres in South Africa. PLOS ONE. 2019;14(5):e0215931.
2. 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-8e.
3. Tolarová MM, Cervenka J. Classification and birth prevalence of orofacial clefts. Am J Med Genet. 1998;75(2):126-37.
4. Wang W, Guan P, Xu W, Zhou B. Risk factors for oral clefts: a population-based case-control study in Shenyang, China. Paediatr Perinat Epidemiol. 2009;23(4):310-20.
5. Allori AC, Kelley T, Meara JG, Albert A, Bonanthaya K, Chapman K, et al. A Standard Set of Outcome Measures for the Comprehensive Appraisal of Cleft Care. Cleft Palate Craniofac J. 2017;54(5):540-54.
6. Ombashi S, Kurniawan M, Koudstaal MJ, Allori AC, Jansson K, Rogers-Vizena CR, et al. "Most efficient and meaningful patient reported appearance assessment in different cleft types and age-groups with Cleft-Q". Plast Reconstr Surg. 2023.
7. Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC. Cleft lip and palate. Lancet. 2009;374(9703):1773-85.
8. Kummer AW. Types and causes of velopharyngeal dysfunction. Semin Speech Lang. 2011;32(2):150-8.
9. Knight J, Cassell CH, Meyer RE, Strauss RP. Academic outcomes of children with isolated orofacial clefts compared with children without a major birth defect. Cleft Palate Craniofac J. 2015;52(3):259-68.
10. Van Lierde KM, Dhaeseleer E, Luyten A, Van De Woestijne K, Vermeersch H, Roche N. Parent and child ratings of satisfaction with speech and facial appearance in Flemish pre-pubescent boys and girls with unilateral cleft lip and palate. Int J Oral Maxillofac Surg. 2012;41(2):192-9.
11. Peterson-Falzone SJ. Speech outcomes in adolescents with cleft lip and palate. Cleft Palate Craniofac J. 1995;32(2):125-8.
12. Baek RM, Kim BK, Jeong JH, Ahn T, Park M, Han J. The effect of age at surgery and compensatory articulation on speech outcome in submucous cleft palate patients treated with double-opposing Z-plasty: A 10-year experience. J Plast Reconstr Aesthet Surg. 2017;70(5):646-52.
13. Markham C, van Laar D, Gibbard D, Dean T. Children with speech, language and communication needs: their perceptions of their quality of life. Int J Lang Commun Disord. 2009;44(5):748-68.
14. Nicola K, Watter P. Health-related quality of life from the perspective of children with severe specific language impairment. Health Qual Life Outcomes. 2015;13:127-.
15. Sell D, Sweeney T. Percent Consonant Correct as an Outcome Measure for Cleft Speech in an Intervention Study. Folia Phoniatr Logop. 2019:1-9.
16. Klassen AF, Tsangaris E, Forrest CR, Wong KW, Pusic AL, Cano SJ, et al. Quality of life of children treated for cleft lip and/or palate: a systematic review. J Plast Reconstr Aesthet Surg. 2012;65(5):547-57.

17. Klassen AF, Riff KWW, Longmire NM, Albert A, Allen GC, Aydin MA, 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-E62.
18. Wong Riff KW, Tsangaris E, Goodacre T, Forrest CR, Pusic AL, Cano SJ, 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.
19. Tsangaris E, Wong Riff KWY, Goodacre T, Forrest CR, Dreise M, Sykes J, 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.
20. Wong Riff KWY, Tsangaris E, Goodacre TEE, Forrest CR, Lawson J, Pusic AL, 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;55(3):442-50.
21. McLeod S, Crowe K, Shahaeian A. Intelligibility in Context Scale: Normative and Validation Data for English-Speaking Preschoolers. Lang Speech Hear Serv Sch. 2015;46(3):266-76.
22. McLeod S. Intelligibility in Context Scale: cross-linguistic use, validity, and reliability. Speech, Language and Hearing. 2020;23(1):9-16.
23. 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):648-56.
24. McLeod S, Harrison LJ, McCormack J. Skattning av förståelighet i kontext: Svenska [Intelligibility in Context Scale: Swedish] (T. Lagerberg, Trans.) Bathurst, NSW, Australia: Charles Sturt University. 2012.
25. McLeod S, Harrison LJ, McCormack J. Schaal voor verstaanbaarheid in de context: Nederlands [Intelligibility in Context Scale: Dutch] (A. van Doornik, Trans). Bathurst, NSW, Australia: Charles Sturt University. 2013.
26. Lohmander A, Hagberg E, Persson C, Willadsen E, Lundeborg I, Davies J, et al. Validity of auditory perceptual assessment of velopharyngeal function and dysfunction – the VPC-Sum and the VPC-Rate. Clinical Linguistics & Phonetics. 2017;31(7-9):589-97.
27. The R Project for Statistical Computing: The R Foundation; 2021 [Available from: https://www.r-project.org/.
28. Semb G, Enemark H, Friede H, Paulin G, Lilja J, Rautio J, et al. A Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 1. Planning and management. J Plast Surg Hand Surg. 2017;51(1):2-13.
29. Long RE, Jr., Hathaway R, Daskalogiannakis J, Mercado A, Russell K, Cohen M, 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(3):239-43.
30. Shaw WC, Semb G, Nelson P, Brattstrom V, Molsted K, Prahl-Andersen B, et al. The Eurocleft project 1996-2000: overview. J Craniomaxillofac Surg. 2001;29(3):131-40; discussion 41-2.
31. Lim CR, Harris K, Dawson J, Beard DJ, Fitzpatrick R, Price AJ. Floor and ceiling effects in the OHS: an analysis of the NHS PROMs data set. BMJ open. 2015;5(7):e007765-e.

32. Stucki G, Liang MH, Stucki S, Katz JN, Lew RA. Application of statistical graphics to facilitate selection of health status measures for clinical practice and evaluative research. Clin Rheumatol. 1999;18(2):101-5.
33. Tsangaris E, Riff K, Vargas F, Aguilera MP, Alarcon MM, Cazalla AA, et al. Translation and cultural adaptation of the CLEFT-Q for use in Colombia, Chile, and Spain. Health Qual Life Outcomes. 2017;15(1):228.
34. Tsangaris E, Wong Riff KWY, Dreise M, Stiernman M, Kaur MN, Piplani B, et al. Translation and cultural adaptation of the CLEFT-Q into Arabic, Dutch, Hindi, Swedish, and Turkish. European Journal of Plastic Surgery. 2018;41(6):643-52.
35. Stiernman M, Klintö K, Persson M, Becker M. Comparison of Corresponding Scores From the Cleft Hearing Appearance and Speech Questionnaire (CHASQ) and CLEFT-Q in Swedish Patients With Cleft Lip and/or Palate. The Cleft Palate-Craniofacial Journal. 2020:105566562096412.
36. Hammarström IL, Nyberg J, Alaluusua S, Rautio J, Neovius E, Berggren A, et al. Scandcleft Project Trial 2-Comparison of Speech Outcome in 1- and 2-Stage Palatal Closure in 5-Year-Olds With UCLP. Cleft Palate Craniofac J. 2020;57(4):458-69.
37. Lohmander A, Willadsen E, Persson C, Henningsson G, Bowden M, Hutters B. Methodology for speech assessment in the Scandcleft project--an international randomized clinical trial on palatal surgery: experiences from a pilot study. Cleft Palate Craniofac J. 2009;46(4):347-62.
38. Shprintzen RJ, Marrinan E. Velopharyngeal insufficiency: diagnosis and management. Curr Opin Otolaryngol Head Neck Surg. 2009;17(4):302-7.
39. Brondsted K, Grunwell P, Henningsson G, Jansonius K, Karling J, Meijer M, et al. A phonetic framework for the cross-linguistic analysis of cleft palate speech. Clinical Linguistics & Phonetics. 1994;8(2):109-25.
40. Grunwell P, Brondsted K, Henningsson G, Jansonius K, Karling J, Meijer M, et al. A six-centre international study of the outcome of treatment in patients with clefts of the lip and palate: the results of a cross-linguistic investigation of cleft palate speech. Scand J Plast Reconstr Surg Hand Surg. 2000;34(3):219-29.
41. Sell D, Harding A, Grunwell P. A screening assessment of cleft palate speech (Great Ormond Street Speech Assessment). Eur J Disord Commun. 1994;29(1):1-15.
42. Sell D, Harding A, Grunwell P. GOS.SP.ASS.'98: an assessment for speech disorders associated with cleft palate and/or velopharyngeal dysfunction (revised). Int J Lang Commun Disord. 1999;34(1):17-33.
43. John A, Sell D, Sweeney T, Harding-Bell A, Williams A. The cleft audit protocol for speech-augmented: A validated and reliable measure for auditing cleft speech. The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association. 2006;43(3):272-88.
44. Ahl R, Harding-Bell A. Comparing Methodologies in a Series of Speech Outcome Studies: Challenges and Lessons Learned. The Cleft Palate-Craniofacial Journal. 2017;55(1):35-44.
45. Chapman KL, Baylis A, Trost-Cardamone J, Cordero KN, Dixon A, Dobbelsteyn C, et al. The Americleft Speech Project: A Training and Reliability Study. Cleft Palate Craniofac J. 2016;53(1):93-108.

46. Britton L, Albery L, Bowden M, Harding-Bell A, Phippen G, Sell D. A cross-sectional cohort study of speech in five-year-olds with cleft palate ± lip to support development of national audit standards: benchmarking speech standards in the United Kingdom. Cleft Palate Craniofac J. 2014;51(4):431-51.
47. Baylis A, Chapman K, Whitehill TL, The Americleft Speech G. Validity and Reliability of Visual Analog Scaling for Assessment of Hypernasality and Audible Nasal Emission in Children With Repaired Cleft Palate. Cleft Palate Craniofac J. 2015;52(6):660-70.
48. Pereira VJ, Tuomainen J, Lee KYS, Tong MCF, Sell DA. A perceptual outcome measure of velopharyngeal function based on the Cleft Audit Protocol for Speech-Augmented (CAPS-A VPC-Sum): Validation through a speech osteotomy study. Int J Lang Commun Disord. 2021;56(4):754-67.

FIGURE LEGENDS

Figure 1. Correlations in patients with CP and CL(A)P

All correlations in both cleft types appeared significant (P<0.05), except for the correlation between the PCC and CLEFT-Q SDistress in patients with CL(A)P (p = 0.285). Note that VPC is inversely scored (higher numbers correspond to worse outcomes), thus accounting for the negative correlations with the other scales.

Figure 2: cross-sectional trend analyses of the age-groups

Analyses are presented per outcome measure, per cleft type.

Figure 3 Overview of the new proposed ICHOM Cleft Lip and Palate set concerning speech assessment

Newly made recommendations are coloured in pink

*Suggestion for centres that have adequate resources to implement and are interested in research with speech outcomes.

Appendices

Appendix 1. Speech outcomes of ICHOM Cleft Lip and Palate and field test data

PROMs filled out by the patients are colored in purple, and grey if filled out by the parent(s). Clinical outcome measures are colored in yellow.

*Data derived from the field test only in these cohorts, as there is no measurement moment included in ICHOM Cleft Lip and palate in these age-groups.

Appendix 2. differences of outcome scores based on income status of the country

Univariate analyses. Outcome scores of patients deriving from countries with a high income status were taken as reference. *Significant difference

Appendix 3. descriptives for subgroup analyses

Descriptives of patients deriving from high income countries, including all patients from the ICHOM centers.

Appendix 4: maximum scores and VPC overview

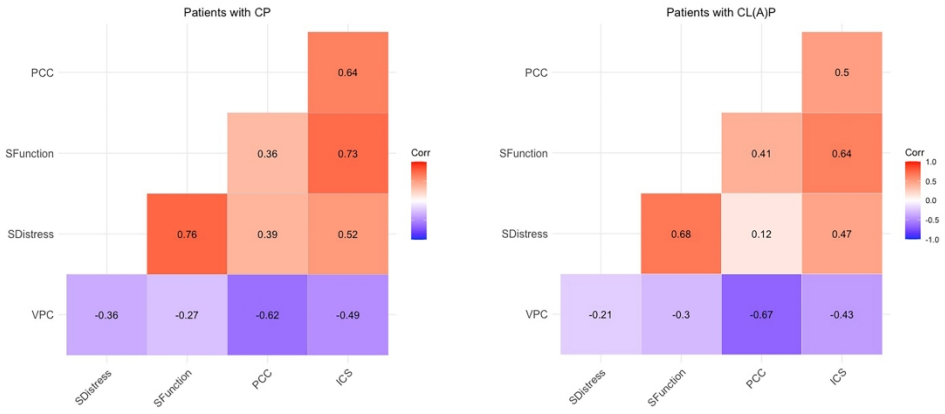


Figure 1. Correlations in patients with CP and CL(A)P
All correlations in both cleft types appeared significant (P<0.05), except for the correlation between the PCC and CLEFT-Q SDistress in patients with CL(A)P (p = 0.285). Note that VPC is inversely scored (higher numbers correspond to worse outcomes), thus accounting for the negative correlations with the other scales.

270x119mm (300 x 300 DPI)

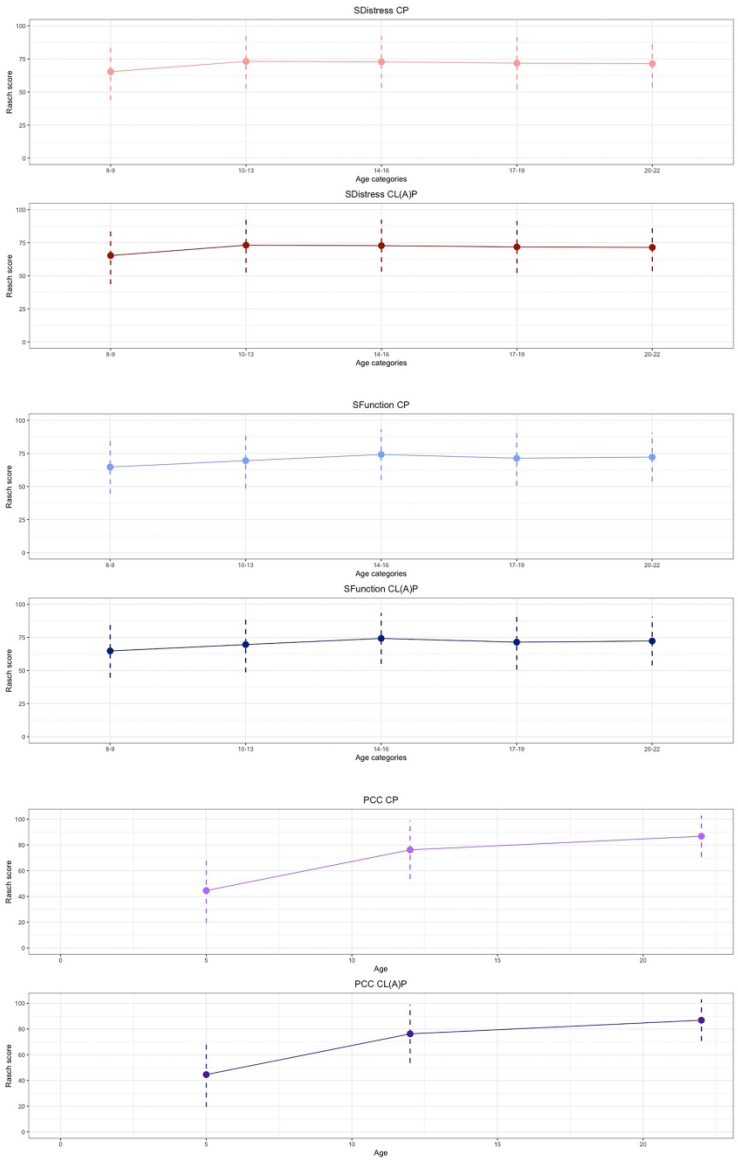
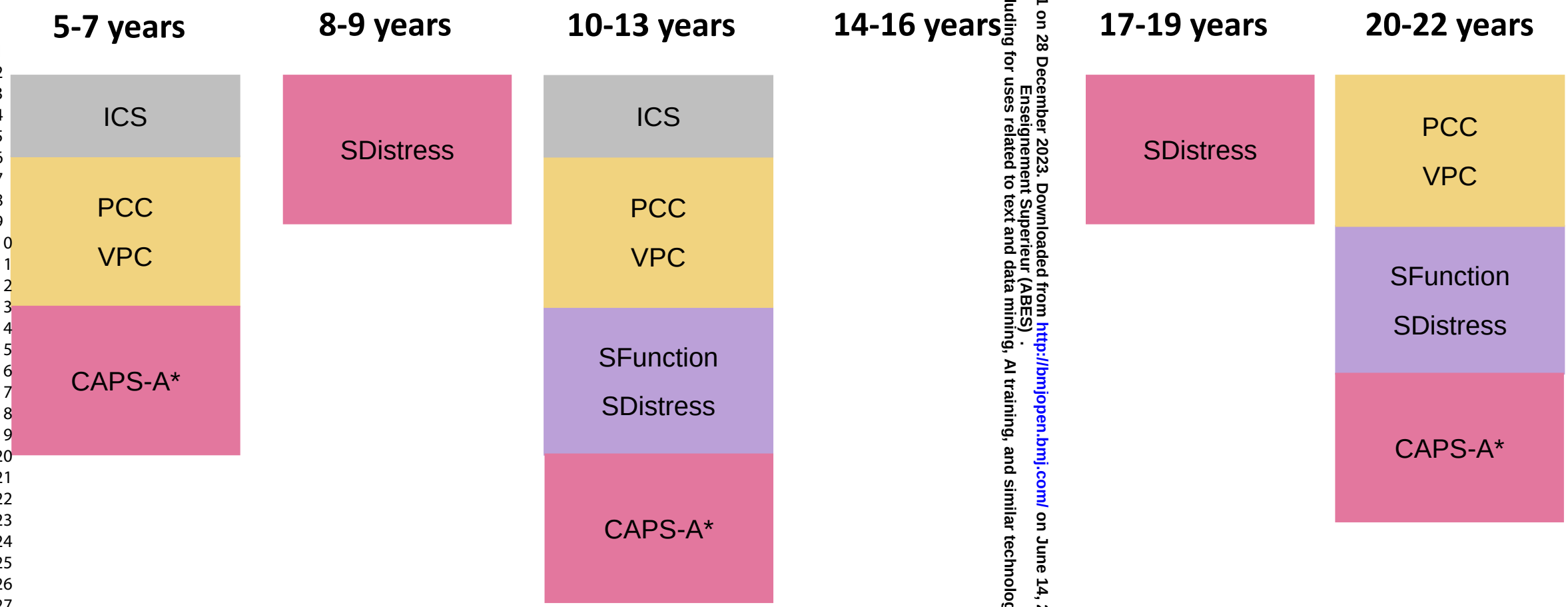
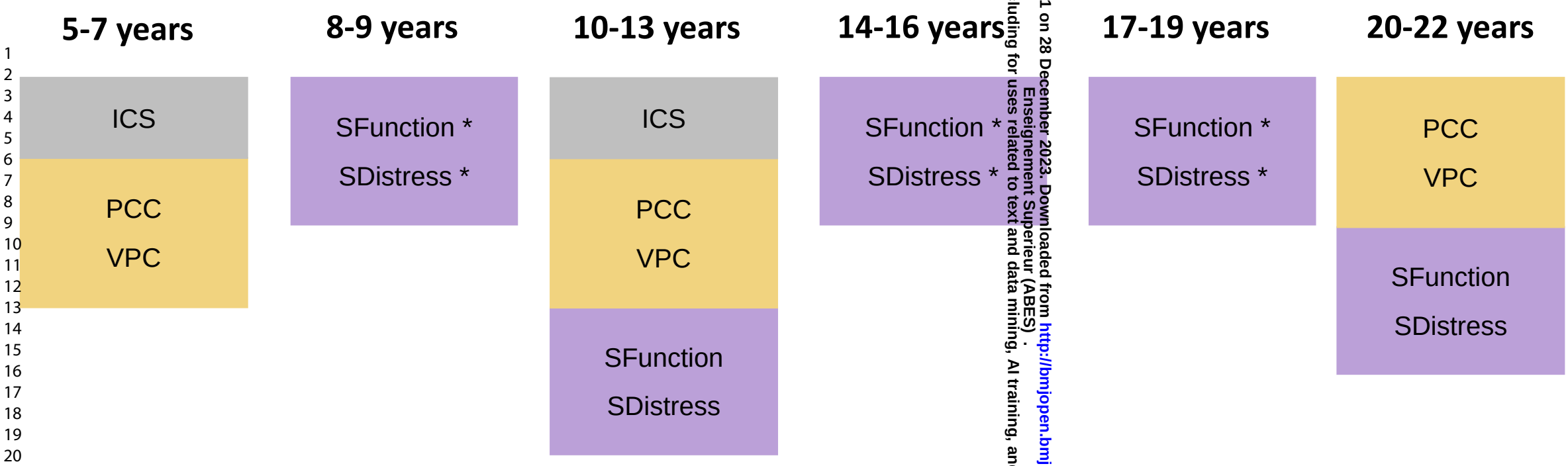


Figure 2: cross-sectional trend analyses of the age-groups
Analyses are presented per outcome measure, per cleft type

199x299mm (300 x 300 DPI)





Appendix 1: differences of outcome scores based on income status of the country

	Income	Beta	SE	P-value	Difference between groups	P-value
Speaking -Related Distress	High (ref) Middle	72.10 -16.39	0.51 1.58	<0.001*	High VS Middle	<0.001*
	High (ref) Low	72.10 -13.08	0.51 1.68	<0.001*	High VS Low	<0.001*
					Middle VS Low	0.285
Speech Function	High (ref) Middle	70.87 -12.32	0.52 1.64	<0.001*	High VS Middle	<0.001*
	High (ref) Low	70.87 -13.18	0.52 1.71	<0.001*	High VS Low	<0.001*
					Middle VS Low	0.923

Univariate analyses. Outcome scores of patients deriving from countries with a high income status were taken as reference. *Significant difference

Appendix 2: descriptives for subgroup analyses

	CP (n=762)	CL(A)P (n=1379)	Total (n=2141)
Gender			
Male	308 (40.4%)	904 (65.6%)	1212 (56.6%)
Female	454 (59.6%)	475 (34.4%)	929 (43.4%)
SDistress			
Number	517 (67.8%	1068 (77.4%)	1585 (74.0%)
Missing	245 (32.2%)	311 (22.6%)	556 (26.0%)
SFunction			
Number	519 (68.1%)	1076 (78.0%)	1595 (74.5%)
Missing	243 (31.9%)	303 (22.0%)	546 (25.5%)
VPC			
Number	173 (22.7%)	213 (15.4%)	386 (18.0%)
Missing	589 (77.3%)	1166 (84.6%)	1755 (82.0%)
ICS			
Number	169 (22.2%)	228 (16.5%)	397 (18.5%)
Missing	593 (77.8%)	1151 (83.5%)	1744 (81.5%)
PCC score			
Number	152 (19.9%)	192 (13.92%)	344 (16.1%)
Missing	610 (80.1%)	1187 (86.1%)	1797 (83.9%)

Descriptives of patients deriving from high income countries, including all patients from the ICHOM centers.

Appendix 3: maximum scores and VPC overview

		Number of measurments with max score	Total measurements
CP	SDistress	18.8% (97)	517
	SFunction	19.8% (103)	519
	ICS	29.0% (49)	169
CL(A)P	SDistress	15.2% (162)	1068
	SFunction	15.9% (172)	1076
	ICS	18.0% (41)	228

Maximum scores

	CP				CL(A)P		
	Competent (0)	Marginally incompetent (1)	Incompetent (2)		Competent (0)	Marginally incompetent (1)	Incompetent (2)
5 years	43 (50%)	21 (24.4%)	22 (25.6%)		11 (15.5%)	17 (23.9%)	43 (60.6%)
12 years	39 (65%)	19 (31.7%)	2 (3.3%)		46 (52.3%)	28 (31.8%)	14 (15.9%)
22 years	11 (40.7%)	13 (48.1%)	3 (11.1%)		26 (48.1%)	19 (35.2%)	9 (16.7%)

VPC overview

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.	Relevant text from manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2	An international, multicenter study was set up including centers from the United States and The Netherlands. Outcomes of clinical measures and Patient Reported Outcome Measures(PROMs) were collected retrospectively according to the ICHOM set
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found		Data from 2500 patients was included. Measured outcomes contained univariate regression analyses, trend analyses, T-tests, correlations and floor and ceiling effects. Results PROMs correlated low to moderate with clinical outcome measures. Clinical outcome measures correlated low to moderate with each other too. In contrast, two CLEFT-Q scales correlated strongly with each other. All PROMs and the Percent Consonants Correct(PCC) showed an effect of age. In patients with an

			isolated cleft palate, a ceiling effect was found in the intelligibility in context scale(ICS).
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Recently, the International Consortium for Health Outcomes Measurement (ICHOM) developed the ICHOM Standard Set for Cleft Lip and Palate (ICHOM Standard Set), with different pathways for varying cleft types(5). Based on patient and expert consensus, a minimal, accessible set of outcome measures was established to enable benchmarking between cleft centers in a systematic manner. For speech assessment, an outcome set was included with both clinical measures and Patient Reported Outcome Measure (PROMs), being the patient's and parent's perspectives. So far, the selected standardized speech outcome measures and their timing have not been evaluated. As an increasing

				number of centers are implementing this set, it is important to critically evaluate and optimize this ICHOM Standard Set. Three centers, the Boston Children's Hospital (Boston, USA), Duke University Hospital (Durham, USA), and the Erasmus Medical Center (Rotterdam, The Netherlands), started clinical implementation and an international collaboration in 2015. The overarching aim of this collaboration is to share data and knowledge obtained by using the set in standard care. Additionally, they collaborate with McMaster University (Hamilton, Canada), who developed and field tested the CLEFT-Q questionnaire, of which many scales are included in the ICHOM Standard set.
Objectives	3	State specific objectives, including any prespecified hypotheses	6	The objective of this study was to evaluate the current standardized speech outcome measures of the ICHOM Standard Set for patients with CP±L. More specifically, the value of every speech outcome measure was examined, as well

				as the best age intervals for assessment of these outcome measures. In addition, other speech assessment tools are discussed. Finally, recommendations are made for an optimal and complete assessment of speech in patients with CP±L, that is efficient and accessible for all cleft centers.
Methods				
Study design	4	Present key elements of study design early in the paper		Three centers (Boston Children's Hospital, Duke University Hospital, and Erasmus Medical Center,) each implemented the ICHOM Standard Set in 2015. All patients treated at these centers for a cleft palate with a cleft lip/cleft alveolus (CL(A)P), or an isolated cleft palate (CP) who were assessed according to the ICHOM Standard Set (age range 5-22 years), were included. In addition, another patient group derived from an international field test of the CLEFT-Q, by McMaster University(17).
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	9, 10	Data was collected restrospectively over a 6 year

				period (2015-2020) and extracted from the electronic patient files in 2018 and 2020 (as a data update). (…) According to the ICHOM protocol, both CLEFT-Q Speech scales were assessed at ages 12 and 22 years(fig. 1). Both PCC and VPC were scored at ages 5, 12 and 22 years, and ICS at ages 5 and 12 years. The field test cross-sectionally collected data from patients with a cleft across 12 different countries with different income-statuses(17).
Participants	6	~~(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up~~ ~~*Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls~~ *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	7	All patients treated at these centers for a cleft palate with a cleft lip/cleft alveolus (CL(A)P), or an isolated cleft palate (CP) who were assessed according to the ICHOM Standard Set (age range 5-22 years), were included. In addition, another patient group derived from an international field test of the CLEFT-Q, by McMaster University(17). According to the age cut-off of the ICHOM Standard Set, only outcomes from CLEFT-Q scales

				of field test patients with a CP±L up to 22 years old were included in the current study(fig 1). Patients from the participating centers were excluded in case they could not sufficiently speak or write the language native to the center's country .
		~~(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed~~ ~~*Case-control study*—For matched studies, give matching criteria and the number of controls per case~~		
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	11	To prevent possible influence of income-status on the outcomes, univariate regression analyses were used to examine differences in outcome scores of the SDistress and SFunction before further analyses. Data was categorized according to the income status of the country where the data had been collected. In order to examine the added value of each PROM and clinical outcome measure in regard with the other measures, correlations were examined between the PROMs; between the clinical outcome measures; and between the PROMs and

the clinical outcome measures. Pearson correlations were used, and outcomes were analysed per cleft type. Correlations were considered strong in case r>0.7; moderate between r=0.5 and R= 0.7; and weak in case r<0.5. Analyses within and between different age-groups were done to explore whether the current outcome measures are assessed at the optimal age-points, and whether additional measurement moments are indicated either for PROMs or clinical outcome measures. Therefore, not only time-points of the ICHOM protocol were included in analyses. As CLEFT-Q outcome scores of all ages between 8-22 years were included from the field test data, time-points used by other large initiatives as Eurocleft, Scandcleft and Americleft were considered as well (28-30). Doing so, the following age-groups were set up: 5-7 years; 8-9 years; 10-13 years; 14-16 years; 17-19 years, 20-22 years(fig. 1).

Per age-group, possible differences between scale scores

were examined with independent T-tests. Bonferroni correction was applied for multiple testing. Trend analyses were performed to identify potential problems in specific age-groups.

Floor- and ceiling effects were examined to identify the suitability of the outcome measures in our population. A floor or ceiling effect is seen when a considerable amount of the outcome scores are either scored the best (in this case a maximum score, thus a ceiling effect), or the worse (in this case a minimum score, thus a floor effect). Both effects result in a truncated distribution of the outcomes on either side of the scale (31, 32). Minimum and maximum score outcomes of all PROMs and the PCC were evaluated. A percentage of 20% or more of the patients scoring the minimum or maximum outcome score was considered as a ceiling effect. In VPC, the outcome distributions were examined.

Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7, 9	Outcome measures PROMs CLEFT-Q Scales The CLEFT-Q is developed specifically to assess QoL from the patient's perspective in patients with a CP±L. A literature review, patient interviews and psychometric testing, established the final content of the scales, which covers several overarching domains(18-20). Speech is assessed through two scales, each covering a different domain. Both scales have 3 response options for each item (always; sometimes; never); a lower score equals a worse outcome. Completing the scales can be done online; it will take the patient several minutes. Speaking-Related Distress (SDistress) is part of the psychosocial domain. The scale contains 10 items that relate to the psychosocial part of speech difficulties, like nervousness or frustration(20). Speech Function (SFunction) focuses on the functional speech difficulties that patients

themselves identify, for example the ability to say certain letters or words. The scale consists of 12 items that belong to the facial function domain(20).

Intelligibility in Context Scale(ICS) is a measure that assess the intelligibility of the child. It is a 7-item, parent-reported questionnaire designed to be scored by speech pathologists. The score indicates a child's level of functional intelligibility, by assessing the degree to which the speech of the patient is understood by different communication partners. The total score is calculated by the averages of the items completed. ICS appeared to be a valid and reliable tool for children with speech disorders, (21, 22), but not specifically designed nor validated for patients with CP±L(23). It is available in several languages, and normative data exists for English-speakers(24, 25).

Clinical outcome measures

Percent Consonants Correct(PCC) is developed to detect speech sound errors. PCC scores are calculated by using a standard, crossecitonally translated set of words that include all speech problems children with CP±L often tend to have.
In case of any problems, their severity can be categorized: PCC scores of 85-100% indicate mild to no problems; scores of 65-84.9% indicate mild-moderate problems; scores of 50-64.9% indicate moderate-severe problems; and scores <50% indicate severe problems(20). PCC is suitable for usage in patients with CP±L when assessed by well-trained clinicians(8).

Velopharyngeal Competence rating(VPC) discriminates between three categories: ‘competent’, ‘marginally incompetent’ and ‘incompetent’. The outcome is determined by the speech therapist based on the PCC test and spontaneous speech. In case of any clinical

				evidence of minor problems regarding the competence, VPC was categorized as 'marginally incompetent'. When clinically significant problems were detected, suggesting surgical management and/or speech therapy, VPC was categorized as 'incompetent'. Prior studies found VPC to be suitable as a first clinical choice for the assessment of velopharyngeal dysfunction and is recommended for both clinical follow-up and research(26).
Bias	9	Describe any efforts to address potential sources of bias	N/A	
Study size	10	Explain how the study size was arrived at	N/A (see data collection, which is cited at point 7 'participants')	

Continued on next page

Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	See citations from point 7 'variables'
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	See citations from point 7 'variables'
		(*b*) Describe any methods used to examine subgroups and interactions	See citations from point 7 'variables'
		(*c*) Explain how missing data were addressed	N/A
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	See citations from point 7 'variables'
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	N/A

		(b) Give reasons for non-participation at each stage	N/A, as our data was collected as part of regular care protocols	
		(c) Consider use of a flow diagram		
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	12, table 1	
		(b) Indicate number of participants with missing data for each variable of interest	N/A	
		(c) ~~*Cohort study*—Summarise follow-up time (eg, average and total amount)~~		
Outcome data	15*	~~*Cohort study*—Report numbers of outcome events or summary measures over time~~		

	Fieldtest (n = 1723)	ICHOM (n = 777)	Overall (n = 2500)
Sex			
Male	981 (56.9%)	444 (57.1%)	1425 (57.0%)
Female	742 (43.1%)	333 (42.9%)	1075 (43.0%)
Cleft Type			
CP	517 (30.0%)	301 (38.7%)	818 (32.7%)
CL(A)P	1206 (70.0%)	476 (61.3%)	1682 (67.3%)
Income classifications			
High Income	1364 (79.2%)	777 (100%)	2141 (85.6%)
Up Middle income	199 (11.5%)	0 (0%)	199 (8.0%)
Low Middle Income	160 (9.3%)	0 (0%)	160 (6.4%)

		~~*Case-control study*—Report numbers in each exposure category, or summary measures of exposure~~	
		Cross-sectional study—Report numbers of outcome events or summary measures	See point 14 'descriptive data', where table 1 is cited.
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	N/A
		(*b*) Report category boundaries when continuous variables were categorized	Page 10: Psychometric validation of the SStress and SFunction confirmed suitability to use a 0 to 100 scale deriving from the sum scores for analysis(17).
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not relevant for the current study

Continued on next page

Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	10	All participating ICHOM centers are high-income countries, whereas part of the field test data was collected in upper middle and lower middle income countries. To prevent possible influence of income-status on the outcomes, univariate regression analyses were used to examine differences in outcome scores of the SDistress and SFunction before further analyses. Data was categorized according to the income status of the country where the data had been collected.
Discussion				
Key results	18	Summarise key results with reference to study objectives	21	From the current study, it can be concluded that the current ICHOM Standard Set is informative and efficient. PROMs were shown to be of added value, and the CLEFT-Q appeared to be the most suitable PROM. Therefore, continuation of collecting the current outcome measures and time points is recommended. Furthermore, a minor extension is suggested: In addition to the current time-points of assessment, it is recommended to implement the CLEFT-Q SDistress scale at the age of 8-9

				and 17-19 as well. Further adjustments of the set could comprise an additional categorization of the PCC score, based on the framework of Eurocleft and adjusted for clinical usage.
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		Data was analyzed cross-sectional. Longitudinal analyses to explore development of speech and for benchmarking will be possible in the future, since data collection continues. Moreover, because this study included data from the CLEFT-Q field test, a higher number of outcome data from the CLEFT-Q scales was available for analyses than from the other outcome measures included in the ICHOM Standard Set.
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	15-20 (entire discussion, where first results of the current study are discussed for different ages, for each outcome measure. Then future	

			considerations and limitations are discussed.)	
Generalisability	21	Discuss the generalisability (external validity) of the study results	21	From the current study, it can be concluded that the current ICHOM Standard Set is informative and efficient. PROMs were shown to be of added value, and the CLEFT-Q appeared to be the most suitable PROM. Therefore, continuation of collecting the current outcome measures and time points is recommended. Furthermore, a minor extension is suggested: In addition to the current time-points of assessment, it is recommended to implement the CLEFT-Q SDistress scale at the age of 8-9 and 17-19 as well. Further adjustments of the set could comprise an additional categorization of the PCC score, based on the framework of Eurocleft and adjusted for clinical usage.
Other information				
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	23	This project received no funding. However, S. Ombashi works as a PhD-student for an European Union-funded Network, the

		'Eureopean Reference Network for Craniofacial anomalies and Ear, Nose and Throat disorders'. The results and findings from this project are of help in further European alignment concerning standardized outcome measures in cleft care.

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.