BMJ Open

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

Saranda Ombashi [ID],[1] Melissa Srijanti Kurniawan,[1] Alexander Allori,[2] Banafsheh Sharif-Askary,[2] Carolyn Rogers-Vizena,[3] Maarten Koudstaal,[4] Marie-Christine Franken,[5] Aebele B Mink van der Molen,[6] Irene Mathijssen,[7] Anne Klassen [ID],[8] Sarah Lisa Versnel[9]

Correspondence to
Saranda Ombashi;
s.ombashi@gmail.com

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 International Consortium of Health Outcomes Measurement (ICHOM) has developed a set of standardised 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 Standard Set. Additionally, speech assessments in other cleft protocols and initiatives are discussed.

Design, setting and participants An international, multicentre study was set up including centres from the USA 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, a questionnaire developed and validated for patients with a cleft, 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 were 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.

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 *Cleft Audit Protocol for Speech Augmented*, is recommended.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ International, multicentre 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 categorised in multiple classification systems, of which a commonly used classification system includes four cleft types: a cleft lip (CL); a cleft lip and alveolus (CLA); a cleft palate (CP); and 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 has to face, treatment of patients with CL/P is ideally done by a specialised and multidisciplinary cleft team in which speech therapists, maxillofacial and plastic surgeons, otolaryngologists,

1

BMJ

paediatricians, psychologists, orthodontists, geneticists and specialised 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 CP, 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 CP (with or without a CL, 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 centres in a systematic manner. For speech assessment, an outcome set was included with both clinical measures and Patient Reported Outcome Measures (PROMs), being the patient's and parent's perspectives.

So far, the selected standardised speech outcome measures and their timing have not been evaluated. As an increasing number of centres are implementing this set, it is important to critically evaluate and optimise this ICHOM Standard Set. Three centres, 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. The CLEFT-Q is a PROM that is specifically developed and validated for patients with a cleft . Many scales are included in the ICHOM Standard Set.

The objective of this study was to evaluate the current standardised 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 centres.

METHODS
Patient population
Three centres (Boston Children's Hospital, Duke University Hospital and Erasmus Medical Center) each implemented the ICHOM Standard Set in 2015. All patients treated at these centres for a CP with a cleft lip/cleft alveolus (CL(A)P), or an isolated 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 (online supplemental appendix 1). Patients from the participating centres were excluded in case they could not sufficiently speak or write the language native to the centre'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 centre as part of regular clinical care. Data were pseudonymised 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
Patient Reported Outcome Measures
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 three 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 assesses 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 exist 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 categorised: 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 categorised as 'marginally incompetent'. When clinically significant problems were detected, suggesting surgical management and/or speech therapy, VPC was categorised 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 centres obtained ethical approval for the current study from their local ethics committees.

Data were 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 clinical outcome measures. During the data collection period, the included centres 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 (online supplemental 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 to 22 years old were included (online supplemental appendix 1).

Income status of the country according to the World Bank Classification was made within the field test. Data from the ICHOM centres were all categorised 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 were 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 0–100 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 centres 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 were categorised 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 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, besides the time points of the ICHOM protocol, additional time points were included in analyses. As CLEFT-Q outcome scores of all ages between 8 and 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 and 20–22 years (online supplemental appendix 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 worst (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 centres (table 1).

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 income and upper middle/lower middle income statuses of the field test were found and results are shown in online supplemental appendix 2.

Table 1 Demographics and phenotypes

	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 classification			
High income	1364 (79.2%)	777 (100%)	2141 (85.6%)
Upper middle income	199 (11.5%)	0 (0%)	199 (8.0%)
Lower middle income	160 (9.3%)	0 (0%)	160 (6.4%)

CL(A)P, cleft lip, alveolus and palate; CP, cleft palate.

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 centres (n=2141). The subgroup characteristics are included in online supplemental 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) (figure 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 weakly 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 −0.67 in patients with CP and CL(A)P, respectively).

Correlation PROMs and clinical outcomes

Moderate correlations were found between 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 (figure 1).

Comparing outcome measures between age groups
Speaking-Related Distress and Function

SDistress and SFunction showed the highest mean outcome (ie, the most favourable ratings) in the age group of 14–16 years old. From thereon, a slightly downward trend is seen (figure 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).

Intelligibility in Context Scale

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 (table 2).

Percent Consonants Correct

Observing the trends in the clinical outcome measures, an upward trend regarding PCC score was seen (figure 2). In both cleft types, PCC scores differed significantly between the age-groups (table 2).
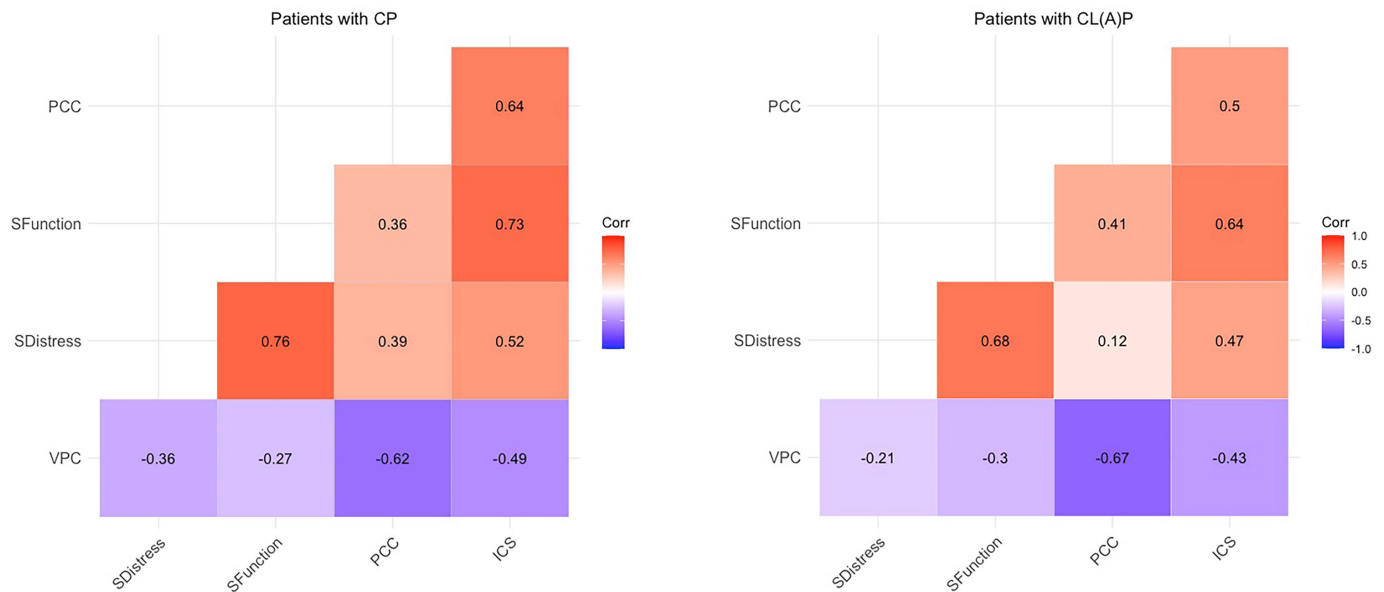
Figure 1 Correlations in patients with CP and CL(A)P. All correlations in both cleft types appeared significant (p<0.05), except for the correlation between the PCC and CLEFT-Q SDistress in patients with CL(A)P (p=0.285). Note that VPC is inversely scored (higher numbers correspond to worse outcomes), thus accounting for the negative correlations with the other scales. CP, cleft palate; CL(A)P, cleft lip, alveolus and palate; CLEFT-Q, —; ICS, Intelligibility in Context Scale; PCC, Percent Consonants Correct; SDistress, Speaking-Related Distress; SFunction, Speech Function; VPC, Velopharyngeal Competence Rating.

Velopharyngeal Competence Rating

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 age group of 22 years, 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 online supplemental 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 outcomes measures, 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 that 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
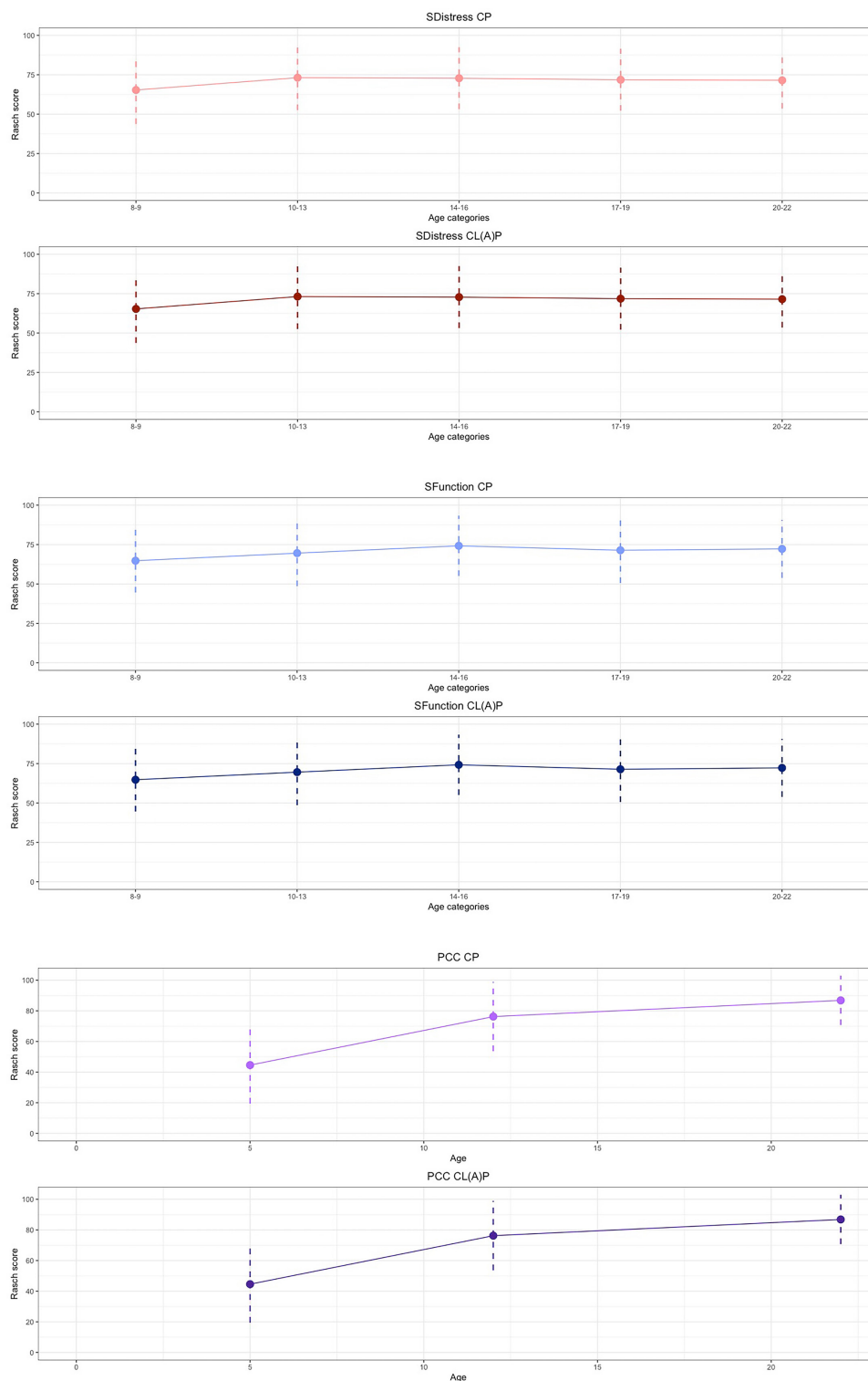
Figure 2 Cross-sectional trend analyses of the age groups. Analyses are presented per outcome measure, per cleft type. CP, cleft palate; CL(A)P, cleft lip, alveolus and palate; PCC, Percent Consonants Correct; SDistress, Speaking-Related Distress; SFunction, Speech Function.

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

Table 2 Mean outcomes per age group, per cleft type

	Age categories	SDistress	SFunction	ICS	PCC
CP	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)*	92.02 (11.48)*
	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)*
	5–7			4.03 (0.50)	46.06 (25.59)
CL(A)P	8–9	65.32 (21.61)	64.75 (20.16)		
	10–13	73.16 (20.62)*	69.57 (20.96)	4.32 (0.55)*	76.25 (22.63)*
	14–16	72.81 (19.66)*	74.22 (19.14)*		
	17–19	71.82 (19.67)*	71.41 (20.79)*		
	20–22	71.49 (18.00)	72.26 (18.40)*		86.83 (16.13)*†

The variables in the table present the mean outcomes of the SDistress, SFunction, ICS and PCC. Outcomes were categorised into cleft type and age group.
*p<0.05 compared to ages 5–7 or 8–9 years.
†p<0.05 compared to ages 10–13 years.
CP, cleft palate; ICS, Intelligibility in Context Scale; PCC, Percent Consonants Correct; SDistress, Speaking-Related Distress; SFunction, Speech Function.

this PROM can be assessed) and 17–19 years 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 in the age groups of 8–9 and 17–19 years (figure 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.
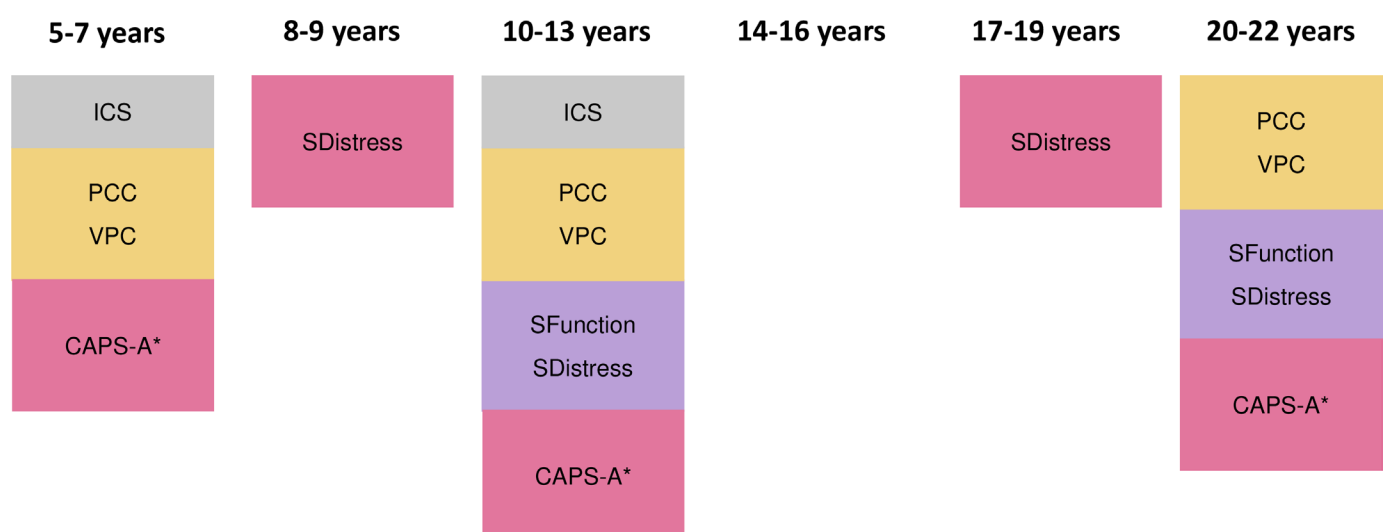

Figure 3 Overview of the new proposed ICHOM Standard Set concerning speech assessment. Newly made recommendations are coloured in pink. *Suggestion for centres that have adequate resources to implement and are interested in research with speech outcomes. CAPS-A, *Cleft Audit Protocol for Speech Augmented*; ICHOM, International Consortium of Health Outcomes Measurement; ICS, Intelligibility in Context Scale; PCC, Percent Consonants Correct; SDistress, Speaking-Related Distress; SFunction, Speech Function; VPC, Velopharyngeal Competence Rating.

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. Twenty-two-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 standardised outcome measures is still debatable, as intercentre and intracentre reliabilities have not been investigated thoroughly in all participating centres so far.[15] Future research should include an examination of scoring and interpreting PCC scores in different centres and/or different countries.

Future considerations regarding alternative speech outcome measures

In order to establish an optimal cleft set for speech assessment, other standardised 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 centres that seek an efficient minimal cleft set that comprises all important speech parameters.[17–19] The standardised approach for translation and validation of the CLEFT-Q questionnaire enables accessibility of the PROM even for centres 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 dichotomised 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 centres 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 centres.

The currently implemented *PCC* lacks any categorisation of consonant errors. The *Eurocleft Speech Group* created a research protocol with a phonetic framework, which was used in six centres and five different languages.[39] It also included consonant production, but assessed on sentence level instead of single words. It is categorised into three groups (correct, almost correct and 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 and other).[39] Moreover, general speech quality was assessed concerning hypernasality and hyponasality, and voice quality.[40] Expert rating of these outcomes requires periodic training of sufficient inter-rater reliability. However, it might be too detailed for implementation in an efficient, clinically oriented cleft set. Therefore, we suggest to further categorise 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 intercentre 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 standardised outcome measures. The Americleft Speech Project found that an acceptable inter-rater and intrarater reliability can be achieved.[43 45] Furthermore, it is suitable for assessment in 5-year-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 centres interested in a minimal and efficient cleft

set. However, centres with experience and resources are highly recommended to implement this tool in order to promote further international standardisation of elaborate speech assessment in patients with CP±L (figure 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 centres that consider the implementation of the CAPS-A is to assess it at ages 5–7, 10–13 and 20–22 years in order to enable long-term follow-up.

Limitations of the study

Data were analysed cross-sectionally. 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 were 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 ages 8–9 and 17–19 years as well. Further adjustments of the set could comprise an additional categorisation of the PCC score, based on the framework of Eurocleft and adjusted for clinical usage.

Author affiliations
[1]Plastic and Reconstructive Surgery, Erasmus Medical Center, Rotterdam, The Netherlands
[2]Plastic, Maxillofacial and Oral Surgery, Duke University Hospital and Children's Health Center, Durham, North Carolina, USA
[3]Plastic and Oral Surgery, Boston Children's Hospital, Boston, Massachusetts, USA
[4]Department of Oral and Maxillofacial Surgery, Erasmus MC Sophia Children Hospital, Rotterdam, The Netherlands
[5]Ear, Nose and Throat, Speech Therapy, Erasmus Medical Center, Rotterdam, The Netherlands
[6]Department of Plastic Surgery, University Medical Center Utrecht, Utrecht, The Netherlands
[7]Department of Plastic and Reconstructive Surgery, Erasmus MC Sophia Children Hospital, Rotterdam, The Netherlands
[8]Pediatrics, McMaster University, Hamilton, Ontario, Canada
[9]Plastic and Reconstructive Surgery, Erasmus MC Sophia Children Hospital, Rotterdam, The Netherlands

Twitter Alexander Allori @allorimd and Anne Klassen @anneklassen

ORCID iDs
Saranda Ombashi http://orcid.org/0000-0001-7988-052X
Anne Klassen http://orcid.org/0000-0003-4720-0096

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:e0215931.
2 Tanaka SA, Mahabir RC, Jupiter DC, *et al*. Updating the epidemiology of cleft lip with or without cleft palate. *Plast Reconstr Surg* 2012;129:511e–8e.
3 Tolarová MM, Cervenka J. Classification and birth prevalence of orofacial clefts. *Am J Med Genet* 1998;75:126–37.
4 Wang W, Guan P, Xu W, *et al*. Risk factors for oral clefts: a population-based case-control study in Shenyang, China. *Paediatr Perinat Epidemiol* 2009;23:310–20.

5 Allori AC, Kelley T, Meara JG, *et al*. A standard set of outcome measures for the comprehensive appraisal of cleft care. *Cleft Palate Craniofac J* 2017;54:540–54.

6 Ombashi S, Kurniawan M, Koudstaal MJ, *et al*. Most efficient and meaningful patient reported appearance assessment in different cleft types and age-groups with cleft-Q. *Plast Reconstr Surg* April 11, 2023.

7 Mossey PA, Little J, Munger RG, *et al*. Cleft lip and palate. *Lancet* 2009;374:1773–85.

8 Kummer AW. Types and causes of Velopharyngeal dysfunction. *Semin Speech Lang* 2011;32:150–8.

9 Knight J, Cassell CH, Meyer RE, *et al*. Academic outcomes of children with isolated orofacial clefts compared with children without a major birth defect. *Cleft Palate Craniofac J* 2015;52:259–68.

10 Van Lierde KM, Dhaeseleer E, Luyten A, *et al*. 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;41:192–9.

11 Peterson-Falzone SJ. Speech outcomes in adolescents with cleft lip and palate. *Cleft Palate Craniofac J* 1995;32:125–8.

12 Baek R-M, Kim B-K, Jeong JH, *et al*. 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;70:646–52.

13 Markham C, van Laar D, Gibbard D, *et al*. Children with speech, language and communication needs: their perceptions of their quality of life. *Int J Lang Commun Disord* 2009;44:748–68.

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;13:127:127.:.

15 Sell D, Sweeney T. Percent consonant correct as an outcome measure for cleft speech in an intervention study. *Folia Phoniatr Logop* 2020;72:143–51.

16 Klassen AF, Tsangaris E, Forrest CR, *et al*. Quality of life of children treated for cleft lip and/or palate: a systematic review. *J Plast Reconstr Aesthet Surg* 2012;65:547–57.

17 Klassen AF, Riff KWW, Longmire NM, *et al*. Psychometric findings and normative values for the CLEFT-Q based on 2434 children and young adult patients with cleft lip and/or palate from 12 countries. *CMAJ* 2018;190:E455–62.

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

19 Tsangaris E, Wong Riff KWY, Goodacre T, *et al*. Establishing content validity of the CLEFT-Q: A new patient-reported outcome instrument for cleft lip/palate. *Plast Reconstr Surg Glob Open* 2017;5:e1305.

20 Wong Riff KWY, Tsangaris E, Goodacre TEE, *et al*. What matters to patients with cleft lip and/or palate: an international qualitative study informing the development of the CLEFT-Q. *Cleft Palate Craniofac J* 2018;55:442–50.

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;46:266–76.

22 McLeod S. Intelligibility in context scale: cross-linguistic use, validity, and reliability. *Speech, Language and Hearing* 2020;23:9–16.

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;55:648–56.

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, *et al*. Validity of auditory perceptual assessment of Velopharyngeal function and dysfunction – the VPC-sum and the VPC-rate. *Clinical Linguistics & Phonetics* 2017;31:589–97.

27 The R Foundation. The R project for statistical computing. 2021. Available: https://www.r-project.org

28 Shaw W, Semb G. The Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 11. What next? *J Plast Surg Hand Surg* 2017;51:88–93.

29 Daskalogiannakis J, Mercado A, Russell K, *et al*. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 1. *Cleft Palate Craniofac J* 2011;48:252–8.

30 Shaw WC, Semb G, Nelson P, *et al*. The Eurocleft project 1996-2000: overview. *J Craniomaxillofac Surg* 2001;29:131–40;

31 Lim CR, Harris K, Dawson J, *et al*. Floor and ceiling effects in the OHS: an analysis of the NHS Proms data set. *BMJ Open* 2015;5:e007765.

32 Stucki G, Liang MH, Stucki S, *et al*. Application of statistical Graphics to facilitate selection of health status measures for clinical practice and evaluative research. *Clin Rheumatol* 1999;18:101–5.

33 Tsangaris E, Riff K, Vargas F, *et al*. Translation and cultural adaptation of the CLEFT-Q for use in Colombia, Chile, and Spain. *Health Qual Life Outcomes* 2017;15:228.

34 Tsangaris E, Wong Riff KWY, Dreise M, *et al*. Translation and cultural adaptation of the CLEFT-Q into Arabic, Dutch, Hindi, Swedish, and Turkish. *Eur J Plast Surg* 2018;41:643–52.

35 Stiernman M, Klintö K, Persson M, *et al*. 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. *Cleft Palate Craniofac J* 2021;58:736–45.

36 Hammarström IL, Nyberg J, Alaluusua S, *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;57:458–69.

37 Lohmander A, Willadsen E, Persson C, *et al*. 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;46:347–62.

38 Shprintzen RJ, Marrinan E. Velopharyngeal insufficiency: diagnosis and management. *Curr Opin Otolaryngol Head Neck Surg* 2009;17:302–7.

39 Brondsted K, Grunwell P, Henningsson G, *et al*. A phonetic framework for the cross-linguistic analysis of cleft palate speech. *Clinical Linguistics & Phonetics* 1994;8:109–25.

40 Grunwell P, Brondsted K, Henningsson G, *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;34:219–29.

41 Sell D, Harding A, Grunwell P. A screening assessment of cleft palate speech (great Ormond street speech assessment). *Eur J Disord Commun* 1994;29:1–15.

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;34:17–33.

43 John A, Sell D, Sweeney T, *et al*. The cleft audit protocol for speech-augmented: a validated and reliable measure for auditing cleft speech. *Cleft Palate Craniofac J* 2006;43:272–88.

44 Ahl R, Harding-Bell A. Comparing Methodologies in a series of speech outcome studies: challenges and lessons learned. *Cleft Palate Craniofac J* 2018;55:35–44.

45 Chapman KL, Baylis A, Trost-Cardamone J, *et al*. The Americleft speech project: A training and reliability study. *Cleft Palate Craniofac J* 2016;53:93–108.

46 Britton L, Albery L, Bowden M, *et al*. 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;51:431–51.

47 Baylis A, Chapman K, Whitehill TL, *et al*. 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;52:660–70.

48 Pereira VJ, Tuomainen J, Lee KYS, *et al*. 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;56:754–67.