The Cleft Palate–Craniofacial Journal 54(5) pp. 555–561 September 2017
© Copyright 2017 American Cleft Palate–Craniofacial Association

ORIGINAL ARTICLE

The Development of the Cleft Aesthetic Rating Scale: A New Rating Scale for the Assessment of Nasolabial Appearance in Complete Unilateral Cleft Lip and Palate Patients

David G.M. Mosmuller, M.D., Lisette M. Mennes, M.D., Charlotte Prahl, D.D.S., Ph.D., Gem J.C. Kramer, D.D.S., Ph.D., Melissa A. Disse, D.D.S., Gijs M. van Couwelaar, M.D., Frank B. Niessen, M.D., Ph.D., J.P.W. Don Griot, M.D., Ph.D.

Objective: **The development of the Cleft Aesthetic Rating Scale, a simple and reliable photographic reference scale for the assessment of nasolabial appearance in complete unilateral cleft lip and palate patients.**

Design: **A blind retrospective analysis of photographs of cleft lip and palate patients was performed with this new rating scale.**

Setting: **VU Medical Center Amsterdam and the Academic Center for Dentistry of Amsterdam.**

Patients: **Complete unilateral cleft lip and palate patients at the age of 6 years.**

Main Outcome Measures: **Photographs that showed the highest interobserver agreement in earlier assessments were selected for the photographic reference scale. Rules were attached to the rating scale to provide a guideline for the assessment and improve interobserver reliability. Cropped photographs revealing only the nasolabial area were assessed by six observers using this new Cleft Aesthetic Rating Scale in two different sessions.**

Results: **Photographs of 62 children (6 years of age, 44 boys and 18 girls) were assessed. The interobserver reliability for the nose and lip together was 0.62, obtained with the intraclass correlation coefficient. To measure the internal consistency, a Cronbach alpha of .91 was calculated. The estimated reliability for three observers was .84, obtained with the Spearman Brown formula.**

Conclusion: **A new, easy to use, and reliable scoring system with a photographic reference scale is presented in this study.**

KEY WORDS: *cleft lip and palate, nasolabial appearance, scoring system*

Dr. Mosmuller is Ph.D. student, Department of Plastic, Reconstructive and Hand Surgery, VU University Medical Center; Dr. Mennes is Ph.D. student, Department of Plastic, Reconstructive and Hand Surgery, VU University Medical Center; Dr. Prahl is Orthodontist, Department of Orthodontics, Academic Center for Dentistry; Dr. Kramer is Orthodontist, Department of Orthodontics, Academic Center for Dentistry; Dr. Disse is Orthodontist, Department of Orthodontics, Academic Center for Dentistry; Dr. van Couwelaar is Plastic Surgeon, Department of Plastic, Reconstructive and Hand Surgery, VU University Medical Center; Dr. Niessen is Plastic Surgeon, Department of Plastic, Reconstructive and Hand Surgery, VU University Medical Center; and Dr. Griot is Plastic Surgeon, Department of Plastic, Reconstructive and Hand Surgery, VU University Medical Center, Amsterdam, the Netherlands.

Mosmuller and Mennes contributed equally to this work and share first authorship.

The manuscript was presented orally by Dr. Mennes at the Scientific Meeting of the Dutch Association Cleft Palate and Craniofacial Anomalies, November 16, 2013, Ghent, Belgium. The manuscript was presented orally by Dr. Mosmuller at the Scientific Meeting of the Dutch Association for Cleft Palate and Craniofacial Anomalies, November 15, 2014, Rotterdam, the Netherlands. Manuscript was presented orally by Dr. Mosmuller at the 72nd Annual Meeting of the American Cleft Palate–Craniofacial Association, April 24, 2015, Palm Springs, California.

Submitted September 2015; Revised January 2016; Accepted January 2016.

Address correspondence to: Dr. David G.M. Mosmuller, Department of Plastic, Reconstructive and Hand Surgery, VU University Medical Center, De Boelelaan 1117, 1081 HV Amsterdam, the Netherlands. E-mail davidmosmuller@hotmail.com.

Cleft lip and palate is a congenital deformity caused by abnormal facial development during gestation. Orofacial clefts are one of the most common congenital malformations, occurring in about 1.7 in 1000 live births (Mossey et al., 2009). A complex combination of many unknown environmental and genetic factors contributes to the etiology of cleft lip and palate (Dixon et al., 2011; de Ladeira et al., 2012; Farronato et al., 2014).

Several studies report no major psychosocial problems in cleft lip and palate patients above that of the noncleft population (Hunt et al., 2005; Collett and Speltz, 2006; Berger and Dalton, 2009; Feragen and Borge, 2010; Pisula et al., 2014). However, two systematic reviews report that patients with cleft lip and palate are at higher risk for behavioral problems, dissatisfaction with facial appearance, and impairment of certain aspects (marriage and friendships) of social functioning (Hunt et al., 2005; Queiroz Herkrath et al., 2015). Another recent review of psychosocial difficulties experienced by cleft lip and palate patients pointed out that among other things, appearance

DOI: 10.1597/15-274

concerns, expectations of surgery and orthodontics, and difficulties such as teasing and bullying are specific issues to which attention should be paid in the course of treatment (Rumsey and Stock, 2013). Hence, nasolabial appearance is an important factor in treatment outcome.

The surgical repair of the cleft lip is in most cases the first operation performed (de Ladeira et al., 2012; Farronato et al., 2014). Achieving symmetry and improvement of nasolabial appearance is an important goal of this operation, and a reliable scoring system is needed to report on outcomes. However, for the assessment of the aesthetic outcome of cleft-related facial deformities, a widely accepted, reliable scoring system is not available (Al-Omari et al., 2005; Sharma et al., 2012; Mosmuller et al., 2013).

One of the most frequently used scoring systems in the literature is the system proposed by Asher-McDade et al. (1991) that uses frontal- and lateral-view masked prints of the nasolabial area. A 5-point ordinal scale is used to assess four facial features: nasal form, nasal symmetry, nasolabial profile, and the alignment of the vermilion border. Later, reference photographs were added for each of the four nasolabial features to facilitate the rating task (Kuijpers-Jagtman et al., 2009). Reported reliability varies in the literature. Satisfactory pooled coefficients of reliability of .60 and .66, respectively, were reported using the 5-point scale when the sum of the four subscores was totaled (Asher-McDade et al., 1992; Mosmuller et al., 2015). Interobserver reliability scores from moderate to good were reported when using the scoring system proposed by Asher-McDade et al. (1991) combined with the photographic reference scale by Kuijpers-Jagtman et al. (2009) (Mercado et al., 2011). More recently, multiple new reference photographs were added to a modified Asher-McDade rating scale and named the Nasolabial Yardstick (Mercado et al., 2016). However, the interrater reliability only slightly improved to .67.

The scoring systems by Asher-McDade et al. (1991) and by Mercado et al. (2016) are time-consuming because three or four different features for each patient have to be assessed. This results in a significant burden when assessing a large number of photographs.

The use of three-dimensional (**3D**) imaging is up and coming, and this seems to be the most reliable in assessing cleft related facial deformities (Al-Omari et al., 2005; Sharma et al., 2012; Mosmuller et al., 2013).

However, scoring two-dimensional (**2D**) photographs is easier to perform and more applicable in daily practice because all cleft patients are photographed during their treatment period at predetermined intervals (Mosmuller et al., 2013). Moreover, 3D equipment is expensive and not available everywhere. Also, when one wants to analyze the results of the last decades only 2D photographs are available. The existing scoring systems for the assessment of the aesthetic results of cleft patients with 2D photographs are rather difficult to use and time-consuming. Therefore, a new scoring system to reliably and quickly assess large numbers of patients in order to compare the aesthetic results of different treatment protocols and different treatment teams would be very useful. The aim of the current study is the development of a new scoring system using reference photographs that is reliable, easy to use, and less time-consuming.

FIGURE 1 Cropped photograph with a circle around the nasolabial area.

Methods

In this study a new reference-photograph scoring system was developed and named the Cleft Aesthetic Rating Scale (CARS). In the following paragraphs the development and the validation of this new scoring system are comprehensively described.

Selection Procedure of the Reference Photographs

Photographs from patients with nonsyndromic, complete unilateral cleft lip and palate were selected from the database of the Academic Center for Dentistry of Amsterdam (**ACTA**). All photographs were taken at the age of 6 years, and all clefts were presented as being left-sided. To reduce the influence of surrounding features, the photographs were cropped, revealing only the nasolabial area (Fig. 1). The use of anonymous patient data collected during routine patient care is in accordance with Dutch law on medical research, and the principles outlined in the Declaration of Helsinki were followed.

In an earlier part of this study, all photographs were independently assessed by six observers. In this assessment the four features of nasolabial appearance (nasal

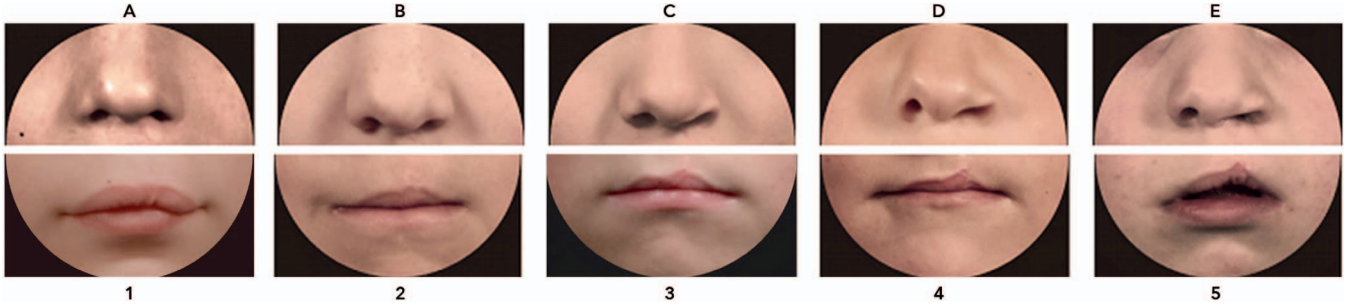

FIGURE 2 The Cleft Aesthetic Rating Scale.

form, nasal symmetry, nasolabial profile, and vermilion border) proposed by Asher-McDade et al. (1991) were scored using a 5-point ordinal scale: A score of 1 represents *very good appearance*; whereas, a score of 5 represents *very poor appearance* (Asher-McDade et al., 1991; Mosmuller et al., 2015).

After the ratings, five photographs with the highest interobserver and intraobserver scores were selected for each component (nasal form, nasal symmetry, vermilion border, and nasolabial profile) and each number of the 5-point scale, resulting in 20 photographs.

For the final photographic scale, five photographs of the nasal form were chosen to score nasal symmetry as well as nasal form. The scorings of the nasal form had a higher intraobserver and interobserver reliability in the earlier ratings, and also the scorings of the nasal form significantly correlated with the overall scoring of the nose in our previous study (Mosmuller et al., 2015). Next, five photos representing the vermilion border were chosen for the photographic scale.

Development of Photographic Reference Scale and Additional Rules

The 10 selected photographs for the nose and lip were used to create a sliding photographic scale, making it possible to create different faces (Fig. 2). For the nose, scores ranged from A to E and for the lip the scores ranged from 1 to 5, representing a *very good* to *very poor appearance*, respectively.

In the pilot sessions in which this reference scale was used, there was disagreement among the observers about which features to assess, leading to an unacceptably low degree of interobserver reliability (Table 1). For this reason, complementary rules were attached to the rating scale (Table 2).

For the assessment of the nose, rules regarding the width and flaring of the affected nostril and symmetry of the nasal tip were added to the scale (Table 2). Widening and flaring of the affected nostril is a common problem in patients with cleft lip and palate. In some cases however, the affected nostril is too small. In these cases a fixed score of C was given. For the assessment of the lip, rules regarding symmetry and the continuity of the vermilion border and length of the philtrum were added. Just as with a small nostril, an extended philtrum is an exception in patients with cleft lip and palate; therefore, in this case a fixed score of 3 was given.

In addition, two other features were to be disregarded in the scorings: volume of the upper lip and the color and thickness of the scar. Volume of the upper lip is considered to be a feature that naturally varies among people and can, for example, also be a result of maxillary protrusion or retrusion. For this reason, the volume of the upper lip should not be a parameter in assessing the aesthetic outcome of cleft lip repair.

The assessment of scars was excluded because reliable assessment of scars on 2D images does not seem to be possible (Asher-McDade et al., 1991). For the assessment of scars, the Patient and Observer Scar Assessment

TABLE 1 Comparison of the Photographic Rating Scale With Attached Rules, Photographic Rating Scale Without Attached Rules, and Asher-McDade Rating Scale for the Total Score

Measure	*Rules Attached Nose and Lip*	*Nose and Lip Without Rules*	*Asher-McDade on Same Set of Photographs*	*Asher-McDade*
Sample size	N = 62	N = 55	N = 55	N = 31
Cronbach alpha	.91	.87	.92	NR*
Interobserver score†	.62	.52	.66	.60
Intraobserver score†	.59–.75	.60–.74	.56–.76	NR
Estimated reliability for 3 observers‡	.84	.80	.89	.82

* NR = not reported
† Obtained using the intraclass correlation coefficient
‡ Obtained using the Spearman Brown Formula

TABLE 2 Rules Attached to the Photographic Rating Scale

Assessment of the nose	
Nose tip	Look at symmetry
Nostrils	Look at width and flaring nostril NB: nostril too small = score C
Assessment of the lip	
Vermilion border	Look at symmetry and continuity
Length philtrum cleft side	Look at shortening of philtrum NB: philtrum too long = score 3
Features that should not be assessed	
Volume of the upper lip	
The color and thickness of the scar	

Scale can be used (Draaijers et al., 2004; Frans et al., 2012).

How to Use the Scale

The scale can be used to assess 2D photographs. It is advisable to use cropped photographs in which surrounding features are excluded. In each frontal image the nose and lip needs to be assessed. The reference photographic scale can be used as a sliding scale, and in this way different faces can be created (Fig. 3). Besides this, the rules as presented in Table 2 should be taken into account when using the scale. The full color and high resolution photographic reference scale can be obtained, without any cost, by contacting the first author.

Testing the Reliability

Prior to photograph assessments, the rules were explained to the observers: three plastic surgeons from the VU University Medical Center Amsterdam and three orthodontists from the ACTA. Before the experimental assessments, four photographs of patients outside the trial were shown to familiarize the judges with the material and the scoring system. The photographs were presented in a random order on color slides using PowerPoint. One frontal photograph of the nasolabial area was shown for each patient for 30 seconds. Between each photograph, a blank slide was shown to prevent the assessment of one photograph being influenced by comparison with the previous photograph.

The observers scored the nose and lip separately using the 5-point scale. Thus, for each patient, two scores were filled in using a score sheet.

After 2 weeks, the same photographs were assessed again in an altered random order by the same observers. The time frame of 2 weeks was used to prevent memory bias.

Statistical Analysis

To analyze the data retrieved from testing the scoring system, the statistical program SPSS 20.0 (IBM Corp., Armonk, NY) was used. Relevant outcomes are the

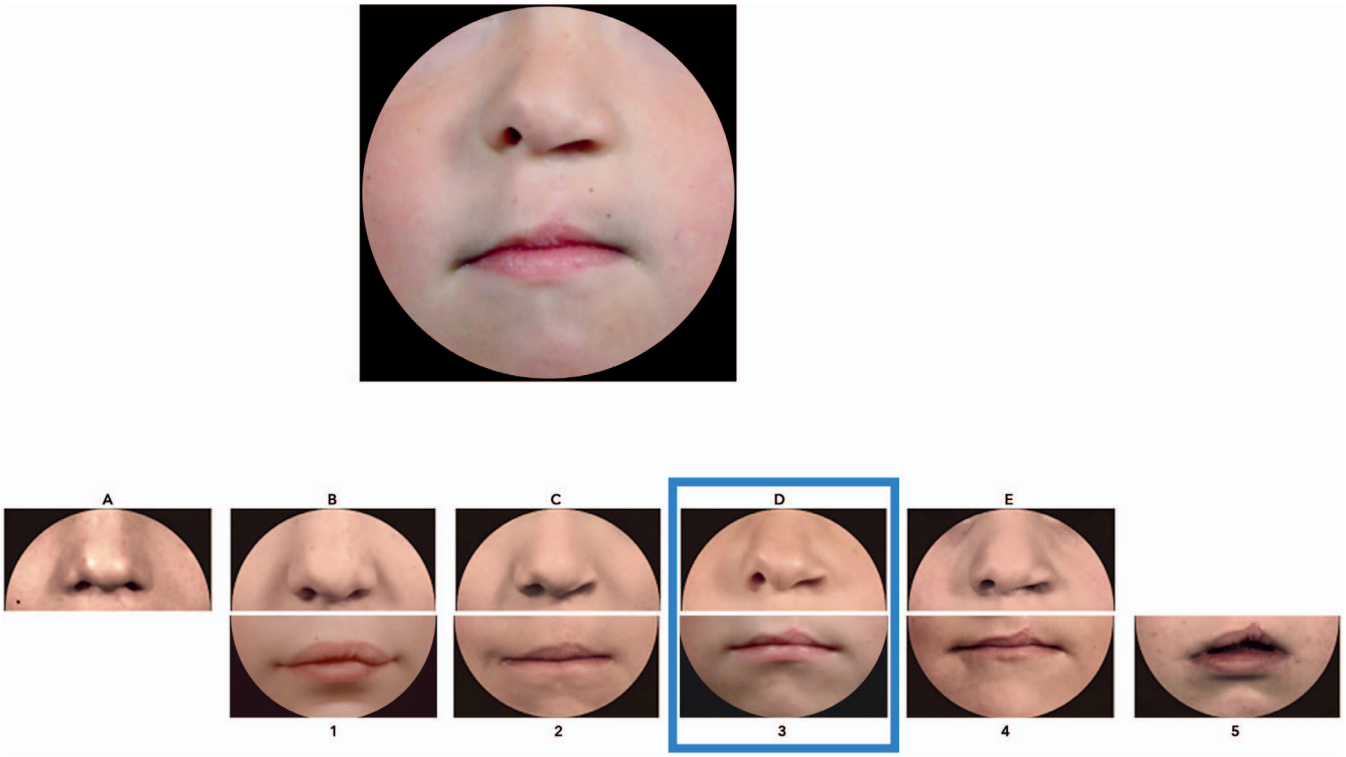

FIGURE 3 Example of scoring using the Cleft Aesthetic Rating Scale.

TABLE 3 Results: Interobserver and Intraobserver Reliability for the Photographic Rating Scale With Attached Rules

Measure	*Nose and Lip*	*Nose*	*Lip*
Interobserver reliability			
Sample size	N = 62	N = 62	N = 62
Cronbach alpha	.91	.90	.90
Interobserver reliability*	.62	.59	.61
95% confidence interval	.52 – .72	.48–.69	.51–.71
Estimated reliability for 3 observers†	.84	.81	.80
Intraobserver reliability			
Plastic surgeon 1	.75	.67	.77
Plastic surgeon 2	.73	.60	.80
Plastic surgeon 3	.67	.70	.63
Orthodontist 1	.71	.73	.64
Orthodontist 2	.69	.71	.72
Orthodontist 3	.59	.60	.54

* Obtained using the intraclass correlation coefficient
† Obtained using the Spearman Brown Formula

intraobserver and interobserver reliabilities, which were assessed using the intraclass correlation coefficient (ICC). *Interobserver reliability* is the degree of agreement among observers when measuring the same subject with the same instrument. *Intraobserver reliability* is the degree of agreement among multiple repetitions of a test performed by a single observer (Streiner et al., 2008). Values approaching the upper limit of ICC = 1.0 indicate a high degree of agreement. The minimal acceptable level of ICC for a scoring system is considered to be .6 (Tinsley and Weiss, 1975). To estimate the reliability for three observers, the Spearman Brown formula was used. The overall reliability of the test was determined using the Cronbach alpha, which is a measure for the internal consistency of the test. An acceptable level for the Cronbach alpha is .7 (Tinsley and Weiss, 1975).

Results

Photographs of 75 six-year-old patients (54 boys and 21 girls) with complete unilateral cleft lip and palate were available. Eleven patients had a right-sided cleft and 64 patients had a left-sided cleft. Thirteen photographs were excluded from the assessment: four photographs were used for practicing the assessment and nine photographs were used with the 5-point rating scale. Eventually, 62 photographs were used for the assessment (44 boys and 18 girls; 10 right-sided and 52 left-sided clefts).

The interobserver reliability was .62 for the nose and lip together (Table 3). The interobserver reliability for the nose and lip scored separately were .59 and .61, respectively. A Cronbach alpha of .91 was calculated for the nose and lip together. The intraobserver reliabilities for the nose and lip together varied from .59 to .75 (Table 3). High to very high scores were obtained for the estimated interobserver reliability for three observers (>.80).

The orthodontists scored more negatively than did the plastic surgeons, with an average score of 2.94 (on a scale from 1 to 5) compared with 2.67. This difference was statistically significant ($P < .05$).

The rating scale with attached rules showed a higher interobserver reliability compared with the rating scale without rules, with scores, respectively, of .62 and .54 for the nose and lip together (Table 1). The interobserver score for the developed rating scale with rules was slightly higher (.62 versus .60) than the interobserver score reported by Asher-McDade et al. (1991) and just slightly lower in comparison with the use of the Asher-McDade aesthetic index by the same observers of the same patients (Table 1) (Mosmuller et al., 2015).

Each photograph was shown to the observers for 30 seconds. This was enough to score both the lip and the nose. When using the scoring system of Asher-McDade et al. (1991), approximately 90 seconds were needed for each patient to score the four different facial features.

Discussion

In previous studies, several methods of rating the facial appearance in patients with unilateral complete cleft lip and palate have been described. Most studies used 2D photographs for the assessment of facial appearance. However, there is still no internationally accepted scoring system for the aesthetic evaluation of cleft lip and palate patients after surgical repair (Al-Omari et al., 2005; Sharma et al., 2012; Mosmuller et al., 2013).

In the current study, the development and evaluation of the CARS is presented: the assessment of nasolabial appearance in patients with complete unilateral cleft lip and palate.

The CARS showed an interobserver reliability of .62 for the nose and lip together, resulting in a test strength of moderate to good (ICC = .52 to .72). This is comparable to the reliability of the rating scale by Asher-McDade et al. (see Table 1; Asher-McDade et al., 1991; Mosmuller et al., 2015).

Because rating facial appearance is based on a subjective notion, the development of a reliable scoring system is a complex process. By attaching complementary rules to the different categories of the scoring system, an objective element is introduced into a primarily subjective scoring system. The introduction of an objective element improved the interobserver reliability, which is shown by the higher reliability scores for the rating scale with attached rules (interobserver reliability of .62) compared with the rating scale without attached rules (interobserver reliability of .52). On the other hand, this makes the scale slightly more complicated. If the rules are well explained, this should pose no problem.

Three other studies used reference photographs for a rating scale (Johnson and Sandy, 2003; Kuijpers-Jagtman et al., 2009; Mercado et al., 2016). As mentioned before, the

rating scale developed by Kuijpers-Jagtman et al. (2009) uses reference photographs for four different nasolabial features. Mercado et al. (2016) uses three different features for the Nasolabial Yardstick, and for each feature four reference photographs are used. This results in a time-consuming scoring system with a significant burden when assessing a large quantity of photographs. Moreover, the reference photographs presented by Mercado et al. (2016) are rather confusing, given that for each feature both the nose and lip are displayed. Where, for instance, all lips are from the same category, the noses differ significantly and can influence the total score. Another limitation of this last study is the use of the Q-sort method, which makes it, in our view, even more complicated and time-consuming. Therefore, it is not applicable in multidisciplinary team meetings or audits. A major advantage of the scoring system developed in the present study is that it is user-friendly and less time-consuming, with 20 to 30 seconds for the assessment of each photograph.

Recommended views by Eurocleft to assess the results in cleft lip and palate patients are frontal, both laterals, three-quarter oblique, and inferior columellar (Shaw et al., 2000). In the Eurocleft study itself, the Asher-McDade rating system was used in which only the frontal and lateral photographs were assessed and relatively low interobserver scores were obtained. Another problem that occurs is that in most centers only frontal and profile photographs are available (Brattström et al., 2005). At the ACTA, standard views in photographing cleft lip and palate patients are frontal, lateral, and inferior columellar. The worm's-eye view is probably the best way to assess symmetry of the nostrils from a professional point of view. This view does not express a person's whole nasolabial attractiveness because this view is rarely shown in social circumstances. In all methods of assessing cleft-related deformities, at least the frontal view is used (Al-Omari et al., 2005; Sharma et al., 2012; Mosmuller et al., 2013).

To keep the new scale pragmatic and less time-consuming, it was decided to only use the frontal view. This decision makes the CARS a good assessment method for assessing the overall aesthetic results in large caseloads. For example, this tool can be used to assess the results of both different cleft lip and palate teams and different surgeons. When researching a particular surgical technique, the CARS is probably not specific enough, and in that case it would be better to look at the scar formation or symmetry.

A more recent method is 3D assessment. In the reviews by Al Omari et al. (2005), Sharma et al. (2012), Mosmuller et al. (2013), and Kuijpers et al. (2014), several studies are described that use 3D imaging for the assessment of nasolabial appearance. These studies concluded that 3D imaging is a promising tool for the assessment of nasolabial appearance. In the systematic review by Kuijpers et al. (2014) it is concluded that "available evidence implies that 3D imaging methods can be used for documentations of CLP [cleft lip and palate] patients. However, there is no data yet showing that 3D methods are more informative than conventional 2D methods and that further research is essential to clarify this and to enable the development of new guidelines for documentation and record taking in cleft lip and palate patients." Three-dimensional imaging is expensive and has a limited availability; whereas, 2D imaging is low in costs, widely available, and part of the diagnostic routine for cleft lip and palate patients. To make large-scale research possible, the availability of diagnostic equipment is important. For this reason 2D imaging still seems the best method for the assessment of cleft-related facial deformities in daily routine.

The strength of the CARS reflects itself in the easy usability and the possibility to assess large numbers of patients in a short period of time. It has a moderate to good reliability, which is in concordance with existing scoring systems (Asher-McDade et al., 1991; Mosmuller et al., 2015; Mercado et al., 2016). Moreover, with three or more observers the interobserver reliability rises to excellent ($>.80$), meaning that a very reliable and fast assessment of large numbers of patients can be performed by using only three observers. An ideal assessment method would be reliable even when it is used by only one observer. Unfortunately, as with other 2D assessment methods, the current CARS needs to be performed by more than one observer to achieve high reliability.

The following recommendations for further improvement of the presented scale can be identified:

First, the standardization of the photographs could be improved. During the photograph assessments, the observers were confronted with differences in head orientation. To enable the observers to assess the photographs consistently, all patients should be photographed from the same angle. However, as shown in the study by Vegter et al. (2000), true standardization of facial photographs is difficult to achieve in cleft lip and palate patients. Due to facial asymmetry in cleft lip and palate patients, the use of a grid or cephalostat will not lead to comparable photographs in these patients (Vegter et al., 2000).

Second, the reference photographs that were used for the 5-point scale could be improved. In most cases, the patients who received a poor or very poor score of the upper lip had a noticeable volume difference of the upper lip, causing asymmetry. The reference photographs for the poor and very poor outcomes did not reflect this volume difference of the upper lip.

Third, one more complementary rule could be attached to the rating scale. During the scoring, there was disagreement among the observers about how to score a whistle deformity. A possible rule could be to add 1 point (resulting in a poorer outcome) if the patient has a whistle deformity.

Fourth, further research should explore whether this scale can also be used for other age groups and patients with incomplete clefts and syndromic patients as well as whether

this scale is distinctive enough when, for example, different cleft teams are compared.

Conclusion

In the present study, a new reference-photograph scoring system was developed named the CARS. This rating scale has an acceptable level of interobserver reliability and an excellent level of internal consistency. The scale achieves good reliability when used by three or more observers. The main advantage of the CARS is that it is easy to use and is less time-consuming in comparison with existing scoring systems, making it a perfect tool for assessing large caseloads.

References

Al-Omari I, Millett DT, Ayoub AF. Methods of assessment of cleft-related facial deformity: a review. *Cleft Palate Craniofac J*. 2005;42:145–156.

Asher-McDade C, Brattström V, Dahl E, McWilliam J, Mølsted K, Plint DA, Prahl-Andersen B, Semb G, Shaw WC, The RP. A six-center international study of treatment outcome in patients with clefts of the lip and palate. Part 4: assessment of nasolabial appearance. *Cleft Palate Craniofac J*. 1992;29:409–412.

Asher-McDade C, Roberts C, Shaw WC, Gallager C. Development of a method for rating nasolabial appearance in patients with clefts of the lip and palate. *Cleft Palate Craniofac J*. 1991;28: 385–390.

Berger ZE, Dalton LJ. Coping with a cleft: psychosocial adjustment of adolescents with a cleft lip and palate and their parents. *Cleft Palate Craniofac J*. 2009;29:352–356.

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

Collett, BR, Speltz, ML. (2006). Social-emotional development of infants and young children with orofacial clefts. *Infants Young Child*. 2006;19:262–291.

de Ladeira PR, Alonso N. Protocols in cleft lip and palate treatment: systematic review. *Plast Surg Int*. 2012;2012:562892.

Dixon MJ, Marazita ML, Beaty TH, Murray JC. Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet*. 2011;12:167–178.

Draaijers LJ, Tempelman FR, Botman YAM, Tuinebreijer WE, Middelkoop E, Kreis RW, van Zuijlen PPM. The Patient and Observer Assessment Scale for clinical scar assessment: a reliable and feasible tool for scar evaluation. *Plast Reconstr Surg*. 2004;113:1960–1965.

Farronato G, Kairyte L, Giannini L, Galbiati G, Maspero C. How various surgical protocols of the unilateral cleft lip and palate influence the facial growth and possible orthodontic problems. Which is the best timing of lip, palate and alveolus repair? Literature review. *Stomatologija*. 2014;16:53–60.

Frans FA, van Zuijlen PP, Griot JP, van der Horst CM. Assessment of scar quality after cleft lip closure. *Cleft Palate Craniofac J*. 2012;49:171–176.

Feragen KB, Borge AI. Peer harassment and satisfaction with appearance in children with and without a facial difference. *Body Image*. 2010;7:97–105.

Hunt O, Burden D, Hepper P, Johnston C. The psychosocial effects of cleft lip and palate: a systematic review. *Eur J Orthod*. 2005;27:274–285.

Johnson N, Sandy J. An aesthetic index for evaluation of cleft repair. *Eur J Orthod*. 2003;25:243–249.

Kuijpers MAR, Chiu Y-T, Nada RM, Carels CEL, Fudalej PS. Three-dimensional imaging methods for quantitative analysis of facial soft tissues and skeletal morphology in patients with orofacial clefts: a systematic review. *PLoS ONE*. 2014:9:e93442.

Kuijpers-Jagtman AM, Nollet PJ, Semb G, Bronkhorst EM, Shaw WC, Katsaros C. Reference photographs for nasolabial appearance rating in unilateral cleft lip and palate. *J Craniofac Surg*. 2009;20:1683–1686.

Mercado AM, Russell KA, Daskalogiannakis J, Hathaway RR, Semb G, Ozawa T, Smith A, Lin AY, Long RE Jr. The Americleft project: a proposed expanded nasolabial appearance yardstick for 5- to 7-year-old patients with complete unilateral cleft lip and palate (CUCLP). *Cleft Palate Craniofac J*. 2016;53:30–37.

Mercado A, Russell K, Hathaway R, Daskalogiannakis J, Sadek H, Long RE Jr, Cohen M, Semb G, Shaw W. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate. Part 4. Nasolabial aesthetics. *Cleft Palate Craniofac J*. 2011;48:259–264.

Mosmuller DG, Griot JP, Bijnen CL, Niessen FB. Scoring systems of cleft-related facial deformities: a review of literature. *Cleft Palate Craniofac J*. 2013;50:286–296.

Mosmuller DGM, Bijnen CL, Don Griot JPW, Kramer G, Disse M, Prahl C, Kuik DJ, Niessen FB. The Asher-McDade aesthetic index in comparison with two scoring systems in complete unilateral cleft lip and palate patients. *J Craniofac Surg*. 2015;26:1242–1245.

Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC. Cleft lip and palate. *Lancet*. 2009;374:1773–1785.

Pisula E, Lukowska E, Fudalej PS. Self-esteem, coping styles, and quality of life in Polish adolescents and young adults with unilateral cleft lip and palate. *Cleft Palate Craniofac J*. 201;51:290–299.

Prahl C, Prahl-Andersen B, van't Hof MA, Kuijpers-Jagtman AM. Infant orthopedics and facial appearance: a randomized clinical trial (Dutchcleft). *Cleft Palate Craniofac J*. 2006;43:659–664.

Queiroz Herkrath AP, Herkrath FJ, Rebelo MA, Vettore MV. Measurement of health-related and oral health-related quality of life among individuals with nonsyndromic orofacial clefts: a systematic review and meta-analysis. *Cleft Palate Craniofac J*. 2015;52:157–172.

Rumsey N, Stock NM. Living with a cleft: psychological challenges, support and intervention. In: Berkowitz S, ed. *Cleft Lip and Palate: Diagnosis and Management*. 3rd ed. Berlin: Springer-Verlag; 2013:907–915.

Sharma VP, Bella H, Cadier MM, Pigott RW, Goodacre TE, Richard BM. Outcomes in facial aesthetics in cleft lip and palate surgery: a systematic review. *J Plast Reconstr Aesthet Surg*. 2012;65:1233–1245.

Streiner DL, Norman GR. Reliability. *Health measurement scales*. Oxford: Oxford University Press; 2008:167–210.

Tinsley HE, Weiss DJ. Interrater reliability and agreement of subjective judgments. *J Couns Psychol*. 1975;22:358–376.

Vegter F, Hage JJ. Standardized facial photography of cleft patients: just fit the grid? *Cleft Palate Craniofac J*. 2000;37:435–440.