Orthodontics & Craniofacial Research

ORIGINAL ARTICLE

A. M. Mercado
C. Phillips
K. W. L. Vig
C.-A. Trotman

The effects of lip revision surgery on nasolabial esthetics in patients with cleft lip

Authors' affiliations:
A. M. Mercado, K. W. L. Vig, College of Dentistry, The Ohio State University, Columbus, OH, USA
C. Phillips, C.-A. Trotman, School of Dentistry, University of North Carolina, Chapel Hill, NC, USA
C.-A. Trotman, Tufts University School of Dental Medicine, Boston, MA, USA

Correspondence to:
Ana M Mercado
Division of Orthodontics
The Ohio State University College of Dentistry
305 West 12th Avenue
Columbus, OH 43210, USA
E-mail: mercado.13@osu.edu

Date:
Accepted 4 April 2014

DOI: 10.1111/ocr.12046

Mercado A. M., Phillips C., Vig K. W. L., Trotman C.-A. The effects of lip revision surgery on nasolabial esthetics in patients with cleft lip
Orthod Craniofac Res 2014; **17**: 216–225. © 2014 John Wiley & Sons A/S. Published by John Wiley & Sons Ltd

Structured Abstract

Objectives – 1) To determine the concordance among surgeons on subjective assessments of nasolabial esthetics in children with repaired cleft lip; and 2) to evaluate longitudinal changes in nasolabial esthetics in relation to cleft lip revision surgery.

Setting and Sample Population – School of Dentistry at University of North Carolina, Chapel Hill. Children with repaired unilateral cleft lip: 32 had lip revision surgery and 27 did not have surgery.

Materials and Methods – Retrospective observational study from a non-randomized clinical trial. Ratings of nasolabial esthetics performed by six surgeons using the Asher-McDade scale at baseline and 12-month follow-up.

Results – Concordance among surgeons ranged from poor to acceptable. Nasolabial ratings at follow-up were better in the Revision group than in the Non-Revision group, although differences were small. The most prevalent change in the Revision Group was improvement in one or more units on the scale, while 'no change' was most prevalent in the Non-Revision group. Participants in the Revision group were more likely to receive a 'no' in relation to the need for lip or nose revision at the follow-up visit.

Conclusion – There were mild esthetic improvements observed in relation to lip revision surgery, which should be interpreted with caution given the subjectivity of the rating method used.

Key words: cleft lip; lip revision surgery; nasolabial esthetics; treatment outcome

Introduction

Children born with a cleft lip with or without cleft palate have an initial lip repair soon after birth, during which the surgeon reconstructs the soft tissue anatomy and attempts to normalize the function and esthetics of the upper lip and nose. In most

instances, the short-term surgical results are acceptable. As the child continues to grow, however, the esthetics may become less than ideal as residual distortions of the soft tissues become more evident. The child and his/her family may notice features such as unsightly scarring of the upper lip, discontinuity of the lip contour, and distorted nasal features. The plastic surgeon examines and discusses these facial characteristics with the family during regular clinic visits. This examination by the surgeon is subjective and may lead to a recommendation for lip and nasal revision surgery.

When electing to proceed with revision surgery, the family and surgeon have expectations that the surgery will have a noticeable change in the nasolabial region resulting in a more 'normal' function and esthetics and that the inherent risks of surgery—such as infection, bleeding, dehiscence, and scarring—will be minimal. However, questions remain as to how best to assess the outcomes of nasolabial appearance in patients with cleft lip with or without cleft palate, and how effective is revision surgery at achieving the desired expectations/outcomes.

Previous studies have used subjective scales based on two-dimensional (2-D) images of patients with increasing degrees of severity of cleft features as a comparison during ratings (1–5). Perhaps the most popular scale of this type, the Asher-McDade, has been used extensively in cross-sectional, multicenter, outcome studies (6–11). Clinicians use this scale to rate frontal and profile facial images of children with complete unilateral cleft lip and palate. In the study by Brattström et al. (8), the Asher-McDade scale was used on subjects ages 9, 12, and 17 years, and intercenter comparisons were carried out of the pooled longitudinal ratings. However, there are no reports using the Asher-McDade scale to assess longitudinal changes in individual patients over time.

The primary aim of this study was to have surgeons evaluate longitudinal changes in nasolabial appearance due to lip revision surgery in patients with repaired unilateral cleft lip with or without cleft palate using the Asher-McDade scale. A secondary aim was to assess the level of agreement among surgeons with the use of this scale. The hypotheses of the study were that lip revision surgery would result in an improved static or 'at rest' appearance of the nasolabial region and that agreement among surgeons would be excellent using the scale.

Methods

The data for this observational retrospective study were based on a subset of participants from a larger non-randomized clinical trial conducted at the University of North Carolina School of Dentistry (UNCSOD) that studied outcomes of lip revision surgery. The overall trial design included three groups of participants: 1) participants with non-syndromic repaired complete unilateral cleft lip with or without a cleft palate who were recommended by the surgeon to have, and who elected to undergo, lip revision surgery (Revision group); 2) participants with non-syndromic repaired complete unilateral cleft lip with or without a cleft palate who, either did not have or, elected not to have a revision lip revision surgery (Non-Revision); and 3) a group of non-cleft 'control' participants (Non-Cleft group). The clinical trial procedures, participant selection criteria, and surgical details were reported previously by Trotman et al. (12, 13) based on STROBE guidelines. All lip revision surgeries were performed by the same surgeon, who was experienced in cleft care. Surgeries were either full-thickness (full muscle takedown) or partial-thickness (partial division of the muscle) lip revisions, with concomitant rhinoplasties when indicated by the surgeon. This study included only the participants in the Revision and Non-Revision groups. From those groups, only participants with full sets of quality digital facial images taken at the two time points of interest were included in the study. The study was approved by the UNCSOD, the Ohio State University, and University of Maryland, Baltimore Institutional Review Boards.

Data collection and processing

The data for the study consisted of longitudinal, two-dimensional, digital, color, facial

images of Revision and Non-Revision participants. For the revision participants, images obtained at baseline or just before revision surgery and then at 12 months after surgery were included in the study. For the non-revision participants, images obtained at time points similar to the revision participants were included. At each time point, each participant had four facial image views captured with the face in a relaxed 'at rest' state. These views were frontal, submental vertex (or alar), right profile, and left profile.

The images then were processed using Adobe Photoshop 6.0 (San Jose, CA, USA). Frontal images were leveled medio-laterally based on the interpupillary line. To remove identifying features of the participants and eliminate possible influences of surrounding facial traits, images were cropped to show only the inner canthus, nasal bridge, nostrils, philtrum, and upper lip of each participant at each time point. This process resulted in a set of four cropped images for a given participant and visit which were arranged as a composite on a single slide using Microsoft Office PowerPoint 2007 (Mountain View, CA, USA; Fig. 1). Subsequently, all the participants' composite slides captured at baseline and at the 12-month follow-up were randomized, coded, and burned on a compact disk (CD) for rating by the surgeons.

Fig. 1. Composite slide showing a participant's frontal, alar, and profile views at rest.

Image rating process

Six plastic surgeons from central Ohio, all experienced in cleft care, agreed to rate the facial images. To calibrate the surgeons, individual practice sessions were conducted using the slides of 10 non-study participants. Then, surgeons were asked to rate each participant's composite slides using the Asher-McDade scale. Five features were rated—nasal form, nasal symmetry, right nasolabial profile, left nasolabial profile, and the shape of the vermilion border—using a 5-point ordinal (Likert) scale where 1 = very good, 2 = good, 3 = fair, 4 = poor, and 5 = very poor. Reference pictures for each feature were provided to help distinguish among the categories on the severity scale (5). An overall Asher-McDade score was calculated as the sum of the responses given for nasal form, nasal symmetry, vermilion border, and the affected side nasolabial profile. In addition to the Asher-McDade ratings, the 'overall' facial disfigurement was rated by the surgeons using a 5-point ordinal scale where 1 = near normal, 2 = mild, 3 = moderate, 4 = severe, and 5 = very severe. Also, the surgeons indicated whether each participant had facial stigmata, needed a nose revision, and/or needed a lip revision using a binary scale of 'yes' vs. 'no'. During the ratings, the surgeons were blinded to the participants' identity, group designation, and visit. The scores were entered onto scoring sheets created with Teleform (San Jose, CA, USA), and the sheets were scanned digitally. To determine the intrasurgeon concordance, baseline and follow-up composite images of eight revision and eight non-revision participants were randomly selected from the total sample, and burned on a second CD to be rated by the surgeons at least 1 day after their first rating. To determine the intersurgeon concordance, ratings given by each surgeon were compared between all possible pairs of surgeons.

Statistical analysis

Intra- and intersurgeon concordances were assessed using weighted Kappa. The comparison of the Revision and Non-Revision groups

for the responses given by the surgeons at the follow-up visit was performed using a proportional odds model based on cumulative logits for each of the ordinal measures, logistic regression for each of the binary measures, and linear regression for the overall Asher-McDade score. For all models, the response at the first visit was included as a covariate, group and surgeon were considered fixed factors, and the interaction between group and surgeon was included. SAS 9.1 Proc Logistic (Cary, NC, USA) was employed for the ordinal and binary outcomes, and Proc GLM was used for the continuous outcome. To assess whether the interaction term should be included in the model, backward selection was used for the ordinal and binary outcomes, and partial *F*-test was used for the continuous outcome. The level of significance was set at 0.05 for all analyses.

Results

The final study sample consisted of 32 participants in the Revision group and 27 in the Non-Revision group (Table 1). For the within-surgeon concordance (Table 2), 87% of the weighted Kappa values were in the acceptable range or higher ($K > 0.45$) when the five facial features were rated. Surgeon C had only fair within-surgeon concordance ($0.24 < K < 0.45$) for overall disfigurement and for three facial features, resulting in the lowest average weighed Kappa (0.43) among all surgeons. The average weighted Kappa calculated for each feature across all surgeons was lowest for nasal symmetry (0.53), although it was in the acceptable range.

Table 1. **Participant demographics**

	Revision (n = 32)	Non-revision (N = 27)
Gender (%)		
Male	18 (56.2)	17 (63.0)
Female	14 (43.8)	10 (37.0)
Age (median and IQR)	11.9 (8.2–15.8)	11.6 (9.7–14.5)

Results for the intersurgeon concordance in Table 3 are given as the minimum and maximum weighed Kappa statistics for each outcome compared between all possible surgeon pairs. The intersurgeon concordances for all features were considerably lower when compared with the intrasurgeon concordances especially for nasal symmetry (maximum Kappa was 0.36, in the fair range). The best concordance among surgeons occurred in the rating of nasolabial profile and overall disfigurement (maximum Kappa values were 0.54 for each profile and 0.55 for overall disfigurement—in the acceptable range). Surgeon F accounted for five of the six minimum weighed Kappa values—within the poor to fair agreement range.

The scores averaged across surgeons were analyzed to assess the impact of lip revision on the outcomes of the Revision group and maturation on the Non-Revision group. Figure 2 illustrates the proportion of cases that had improvement, no change, and worsening over time per outcome and study group. For all ordinal outcomes, there was a higher frequency of cases that improved by one or more units in the Revision group as compared to the Non-Revision group. Also for all outcomes, there was a higher frequency of cases that had no changes in the Non-Revision group as compared to the Revision group. Cases that worsened by one or more units were more prevalent in the Non-Revision group. As the rating scale used in this study is a 5-point scale, changes in 1-unit magnitude may be within the errors related to the sensitivity of the scale and the low intersurgeon agreement, because all surgeons' distributions were averaged. Therefore, differences between the baseline and follow-up ratings equal to or >2 units may be more clinically noticeable. At least 10% of the Revision group improved by two or more units, while 5.6% or less of the Non-Revision group improved by that amount.

Results of the statistical analyses are shown in Table 4. The interaction term of group and surgeon was not statistically significant for any of the outcomes, except nasolabial profile, and was removed from those models. For the nasolabial profile, the pattern of scores given the two

Table 2. **Within-surgeon concordance for each outcome as indicated by the weighted Kappa statistic**

	Within-surgeon concordance							
Surgeon	Nasal form	Nasal symmetry	Vermilion border	Nasolabial profile right	Nasolabial profile left	Overall disfigurement	Average Kappa	# <0.45
A	0.63	0.58	0.46	0.55	0.64	0.53	0.57	0
B	0.47	0.35	0.59	0.66	0.61	0.42	0.52	2
C	0.44	0.59	0.48	0.38	0.39	0.27	0.43	4
D	0.78	0.66	0.59	0.84	0.67	0.66	0.70	0
E	0.58	0.53	0.51	0.76	0.69	0.69	0.63	0
F	0.73	0.47	0.59	0.62	0.76	0.65	0.64	0
Average Kappa	0.61	0.53	0.54	0.64	0.63	0.54	–	–
# <0.45	1	1	0	1	1	2	–	–

Table 3. **Concordance between all possible pairs of surgeons as indicated by the weighted Kappa statistic**

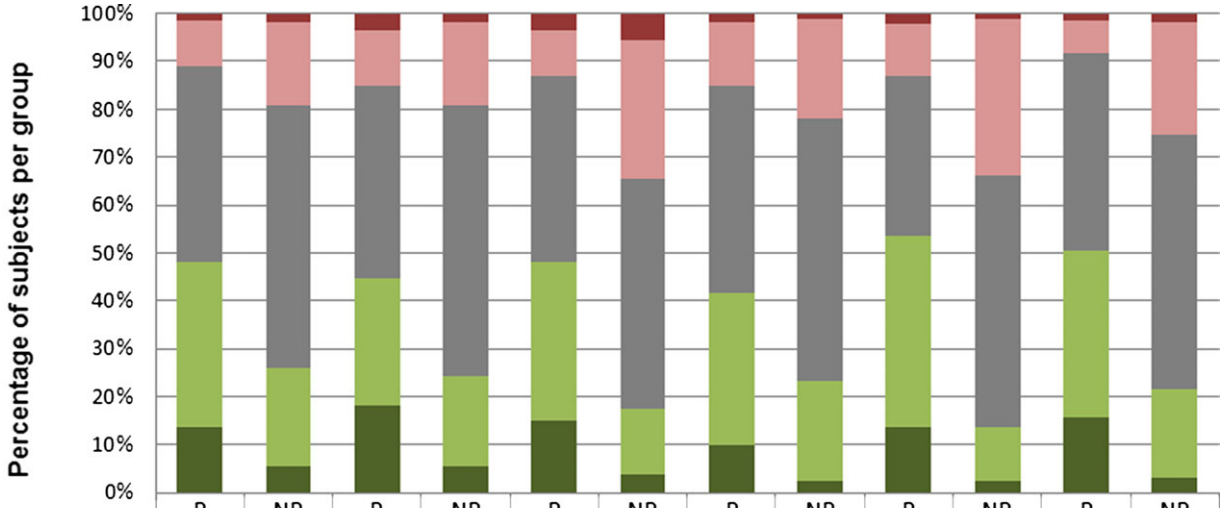
	Between surgeon concordance			
Outcome	Minimum	Surgeon pair	Maximum	Surgeon pair
Nasal form	0.13	B_C	0.50	D_E
Nasal symmetry	0.04	B_F	0.36	A_D
Vermilion border	0.19	C_F	0.49	B_D
Nasolabial profile right	0.23	D_F	0.54	A_E
Nasolabial profile left	0.21	D_F	0.54	B_D
Overall disfigurement	0.25	E_F	0.55	D_E

groups was not the same for all surgeons, and no further analyses were performed. For all other outcomes, the pattern of responses given the two groups was similar for all surgeons. For the ordinal outcomes of nasal form, nasal symmetry, vermilion border, and overall disfigurement, the Revision group had 2.84, 3.40, 2.85, and 4.13 times higher odds, respectively, of receiving lower scores, indicating better appearance, compared with the Non-Revision group at the follow-up time point after controlling for surgeon and the respective baseline score. Likewise, the Revision group had 2.76, 4.15, and 3.93 times higher odds of receiving a 'no' with respect to the presence of a stigmata, the need for a nasal revision, and the need for a lip revision compared with the Non-Revision group at the follow-up time point after controlling for surgeon and the perception at baseline. The average Asher-McDade score for the Revision group was 0.53 units lower at follow-up, indicating a better overall appearance than the Non-Revision group after controlling for surgeon and the baseline score.

Discussion

In general, the intrasurgeon concordance for all surgeons fell within the acceptable range, with the ratings for nasal symmetry having the lowest concordance. This latter finding is similar to those of Brattström et al. (6) who reported that nasal symmetry was more difficult to assess than the other nasolabial features. Nasal symmetry, or deviation, may be the most variable of all the Asher-McDade outcomes, as it considers several features from the center of the nasal bridge to the nasal tip. 'Nasolabial profile' (right and left) obtained the highest intrasurgeon concordance, as shown in previous studies (4), which may be attributable to the fact that this outcome is an evaluation of the contour or outline of the nose and lip, not influenced by symmetry.

As seen in previous studies (4, 6, 11), the weighed Kappa values for all features decreased considerably for the intersurgeon concordance when compared with the intrasurgeon concor-



	Nasal form		Nasal symmetry		Vermilion border		Nasolabial profile right		Nasolabial profile left		Overall disfigurement	
	R	NR	R	NR	R	NR	R	NR	R	NR	R	NR
Worse ≥ 2	1.6	1.9	3.6	1.9	3.7	5.6	1.7	1.3	2.3	1.2	1.6	1.9
Worse = 1	9.4	17.3	11.5	17.3	9.4	29.0	13.3	20.5	10.7	32.5	6.8	23.5
No change	41.1	54.9	40.1	56.8	38.7	48.1	43.3	55.1	33.6	52.6	41.1	53.1
Improved = 1	34.4	20.4	26.6	18.5	33.0	13.6	31.7	20.5	39.8	11.1	34.9	18.5
Improved ≥ 2	13.5	5.6	18.2	5.6	15.1	3.7	10.0	2.6	13.7	2.6	15.6	3.1

Fig. 2. Distribution of cases by amount of change, as indicated by average across all surgeons, from baseline to follow-up.

Table 4. ***P*-values from the proportional odds, logistic and linear regression analyses and estimates of better appearance in the Revision group compared with the Non-Revision group after controlling for surgeon and the baseline score**

		Ordinal					Binary			Continuous
Effect	df	Nasal form	Nasal symmetry	Vermilion border	Nasolabial profile	Overall disfigurement	Stigmata	Need for nasal revision	Need for lip revision	Average rating
Group	1	<0.0001	<0.0001	<0.001	0.05	<0.0001	0.0002	<0.0001	<0.001	<0.0001
Surgeon	5	<0.0001	<0.0001	<0.0001	0.50	<0.001	<0.0001	<0.0001	<0.0001	<0.0001
Baseline	1	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Group x Surgeon	5	Removed	Removed	Removed	0.004	Removed	Removed	Removed	Removed	Removed
Odds*		2.84	3.40	2.85	NA†	4.13	2.76	4.15	3.93	Beta = −0.53

*Ordinal outcomes: Estimates for the likelihood of the Revision group to be perceived as more attractive than the Non-Revision group. Binary outcomes: Estimates for the likelihood of the Revision group to be perceived as having less stigmata or less in need of treatment than the Non-Revision group. Continuous outcome: Estimate of the unit change, on average, indicating a more favorable perception overall in the Revision group compared with the Non-Revision group.
†Odds ratio assessing the two groups was not calculated because the group by surgeon interaction was statistically significant.

dance. In particular, the maximum intersurgeon concordance for nasal symmetry was only fair, demonstrating that surgeons differed the most in their subjective analysis of nasal symmetry compared with other features. Intersurgeon concordance was higher for 'Nasolabial Profile', as

shown in other studies (4). One surgeon accounted for the majority of the lowest weighed Kappa values suggesting that this surgeon had different perceptions of severity in most of the outcomes rated, even after the practice session and the reference photographs. Because these surgeons were similarly trained in the Asher-McDade method, the low concordance between them demonstrates a considerable limitation when using this scale, introducing observer bias. Not all surgeons could rate within acceptable concordance to their colleagues because the subjective scale is susceptible to modifying factors such as personal and professional biases.

Our findings showed that, on average, lip revision surgery does improve the appearance of certain features of the nasolabial region. Nasal form, nasal symmetry, the shape of the vermilion border, and the overall disfigurement were more likely to show improvements at 12-month follow-up in the Revision group than in the Non-Revision group, suggesting that lip revision surgery produces beneficial effects on the 'static' lip esthetics when compared to lip maturation alone. This finding is similar to that for objective measures of nasolabial/circumoral function that also indicated an average improvement in movement of the region (14). These average findings on improvement concur with the expectation of all patients and their families about lip/nose revision surgery; however, they do not necessarily pertain to the individual patient in the Revision group. A Revision patient may have been rated as 'improved' in one of the categories and yet as 'worsened' in another category. In addition, the surgeries in this study varied from partial-thickness to full-thickness lip revisions, and rhinoplasties were included for either type of revision. This potential source of variability in the surgical procedures within the Revision group was not controlled for and thus constitutes a confounding variable. A participant who received a rhinoplasty in addition to lip revision may be more likely to be rated as improved in the nasal features (form or symmetry) than a participant who received lip revision only.

The allocation of the participants into the two study groups was not random and may have been subject to selection bias. Group allocation was based on 1) surgeon's recommendation for revision surgery and 2) patient's decision on surgery. Thus, depending on how those criteria came into effect, the study groups were not necessarily similar in the severity of the nasolabial deformity at baseline. In the Revision group, all patients had lip/nose distortions that were severe enough to need revisions, while in the Non-Revision group, there was a mix of patients who needed revisions (but declined to have surgery) and patients who did not need revisions. It should be noted that patients in each group were compared individually with themselves over time, with no attempt to compare the Non-Revision patients directly with the Revision patients. Patients in the Non-Revision group provided a control over time for maturation. The inequality between the groups is considered as a confounding variable because patients in the Revision group may have been more likely to show improvements because they may have been initially more severely distorted than patients in the Non-Revision group.

The Asher-McDade scale was initially developed and has been used in several cross-sectional studies to compare patients from different cleft palate centers that perform different treatment regimens (4, 6–11). The shortcomings of the method have been described (7, 8), including low inter-rater agreement, the requirement of cropped standardized photographs of the nasolabial region at rest, inconsistency in some profile views taken from the non-affected side and others taken from the affected side, limited suitability of using two-dimensional images to assess three-dimensional facial features, and the inability to assess facial features during animation. In the present study, efforts were made to account for some of these limitations. All images were taken in a standard fashion by a single operator, with a single camera, under the same lighting conditions. Each composite slide included profile images taken from both the affected and the non-affected side. In addition, submental vertex (or alar) views were included to add a dimension of 'depth' in the assessment of nasal form. Despite the precautions taken in

this study, most of the improvements noted were small (1-unit changes in the 5-unit scale). One explanation for this finding may be that the Asher-McDade scale lacks the sensitivity to detect subtle differences due to surgery in participants from a single cleft palate center. Another possible explanation may have been related to the types of clefts in the sample. In this study, the sample included participants with unilateral cleft lip with or without cleft palate. The Asher-McDade scale was validated in patients with complete unilateral cleft lip and palate, a population that may be considered more homogeneous than the present study sample. Including clefts of the lip only with clefts of the lip and palate may be a limitation by influencing the applicability of the Asher-McDade scale for the present study. It is proposed that the intra-oral extent of the cleft (whether or not it included the alveolus or the secondary palate) would have little or no influence on the effect of the lip revision surgery on nasolabial esthetics over the relatively short period of time (1 year) between ratings, especially because none of the participants had other cleft-related surgeries during the study period.

Another limitation of this study is that age was not included in the statistical model to verify its association, if any, with the observed outcomes. It was not anticipated that age would be a factor as there is no difference in the median ages between Revision (11.9 years) and Non-Revision (11.6 years) groups. Brattström et al. (8) reported deterioration of average nasolabial esthetics over a 3-year span, a 5-year span, and cumulatively over an 8-year span. However, their report refers to the trend seen in the average Asher-McDade ratings from one time point to another, with no statistical comparisons to support the observations. The present study evaluated longitudinal ratings from the same individuals over a 1-year period. Thus, it could be expected that any changes in nasolabial appearance would be subtle. Indeed, the most predominant observation in the Non-Revision group over the 1-year period was 'no change' (Fig. 2), with fewer cases that had improved or worsened in the same time interval.

Other studies have used subjective ratings from multiple, calibrated raters with experience in cleft palate care to compare outcomes before and after surgery (15–17). As part of their parallel, three-group, non-randomized clinical trial, Trotman et al. (12) recruited eight surgeons to perform subjective assessments of two groups of children with repaired cleft lip (one group had lip revision surgery and one group did not). Their results showed that the inter-rater agreement was predominantly poor to fair in the surgeons' recommendations for lip revision surgery. In addition, surgeons made the same recommendation before and after revision surgery for most of the participants in the Revision group. One possible explanation is that the surgeons could not perceive a substantial enough difference between viewings to change their recommendation. Likewise, the present study asked surgeons to rate the disfigurement of facial features on each participant before and after lip revision surgery and showed that the intersurgeon concordance for subjective measures was poor to acceptable. Although the analyses showed that lip revision surgery is more likely to result in improvements in appearance than maturation alone, analyses of individual patients' changes showed a high prevalence for 'no change' from baseline to follow-up. The results of Trotman's assessment using a binary 'yes/no' scale (12) as well as the results of the present study using the Asher-McDade 5-point ordinal scale suggest that, at least 1 year after surgery, subjective measures are limited at establishing if the lip revision procedure produced a real benefit in nasolabial esthetics. A longer time interval may be necessary (3 years or longer after revision surgery) to evaluate whether nasolabial esthetics of subjects in the Revision group deteriorate less over time than that of subjects in the Non-Revision group.

Perhaps objective measures of the nasolabial region rather than a subjective, two-dimensional, assessment would have been able to detect these subtle changes. There are several longitudinal studies that have evaluated subjects with cleft lip before and after surgery using objective measures derived from three-dimensional assess-

ment techniques (18–20). All of these studies have detected significant pre- and post-surgical changes in some of their measures of height, shape, contours, and symmetry demonstrating that objective three-dimensional measures are discriminating and sensitive for use in longitudinal studies. Trotman et al. (13) have shown that certain objective measures correlate well with subjective measures of lip form at rest and should be included in a comprehensive evaluation of nasolabial appearance of patients with clefts. Interestingly, Tanikawa et al. (21) also used objective three-dimensional measures to describe features of static lip form in participants with and without cleft lip. They were able to generate different mean lip categories that characterized their entire (cleft and non-cleft) sample ranging from normal to different severity categories of cleft lip. The authors proposed that the methodology could be used to assess outcomes of individual patients before and after surgery.

The clinical implications of this study are several. First, when a patient with repaired cleft lip is evaluated in a clinical setting, he or she is usually seen by a single plastic surgeon. The plastic surgeon will give the patient and his/her family a subjective assessment on the present level of facial disfigurement, an estimate of the changes/outcomes that may be expected from lip revision surgery, and a recommendation on whether or not to proceed with lip revision surgery. All of these assessments and recommendations will be, to a considerable extent, limited to that particular surgeon and may not be the same as the assessments and recommendations of another surgeon. Each surgeon's perceptions are influenced by factors such as training, experience, surgical skills, and level of competence, among others. According to the results of the present study, it may be predicted that improvement will be the most likely outcome of revision surgery; however, these 'improved' outcomes were related to the competence of a single surgeon performing all the revisions and should not be extrapolated to the outcomes that may be achieved by other surgeons. Moreover, a surgeon should use caution when advocating that revision surgery will result in drastic improvements in appearance. In many cases, the improvements may be slight or not noticeable. Less frequently, revision surgery may result in worsening of the appearance. Lastly, future studies should include objective, three-dimensional measures as a way to control for confounding factors that influence surgeons' subjective clinical evaluations.

Conclusions

Subjective evaluations by surgeons on the outcomes of lip revision surgery in children with cleft lip showed mild esthetic improvements. However, those evaluations revealed limited concordance among surgeons and hence should be interpreted with caution.

Clinical relevance

Children with repaired cleft lip undergo lip revision surgery based, in part, on a recommendation from surgeons about the likely benefits from surgery. It is unclear whether the effects of lip revision surgery on nasolabial esthetics can be accurately and consistently assessed. This study showed limited concordance among surgeons when using the Asher-McDade 5-point ordinal scale for ratings of nasolabial esthetics before and after revision surgery. Therefore, the mild esthetic improvements observed in relation to lip revision surgery should be interpreted with caution given the subjectivity of the rating method used.

Acknowledgement: This study was supported by a research supplement awarded to Dr. Ana Mercado as part of parent grant R01 DE013814-01A1 from the National Institute of Dental and Craniofacial Research to Dr. Carroll-Ann Trotman.

References

1. Glass L, Starr CD. A study of relationships between judgments of speech and appearance of patients with orofacial clefts. *Cleft Palate Craniofac J* 1979;16:436–40.
2. Schneiderman CR, Harding JB. Social ratings of children with cleft lip by school peers. *Cleft Palate Craniofac J* 1984;21:219–23.
3. Tobiasen JM, Hiebert JM, Boraz RA. Development of scales of severity of facial impairment. *Cleft Palate Craniofac J* 1991;28:419–24.
4. 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;28:385–91.
5. Kuijpers-Jagtman AM, Nollet PJPM, Semb G, Bronkhorst EW, Shaw WC, Katsaros C. Reference photographs for nasolabial appearance rating in unilateral cleft lip and palate. *J Craniofac Surg* 2009;20(Suppl. 2):1683–6.
6. Brattström V, McWilliam J, Larson O, Semb G. Craniofacial development in children with unilateral clefts of the lip, alveolus, and palate treated according to three different regimes. Assessment of nasolabial appearance. *Scand J Plast Reconstr Surg Hand Surg* 1992;26:313–9.
7. Asher-McDade C, Brattström V, Dahl E, McWilliam J, Mølsted K, Plint DA et al. 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;29:409–12.
8. 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.
9. Nollet PJPM, Kuijpers-Jagtman A, Chatzigianni A, Semb G, Shaw WC, Bronkhorst EM et al. Nasolabial appearance in unilateral cleft lip, alveolus and palate: a comparison with Eurocleft. *J Craniomaxillofac Surg* 2007;35:278–86.
10. Fudalej P, Katsaros C, Bongaarts C, Dudkiewicz Z, Kuijpers-Jagtman AM. Nasolabial esthetics in children with complete unilateral cleft lip and palate after 1- versus 3-stage treatment protocols. *J Oral Maxillofac Surg* 2009;67:1661–6.
11. Mercado AM, Russell KA, Hathaway RR, Daskalogiannakis J, Sadek H, Long RE Jr et al. 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;48:259–64.
12. Trotman C-A, Phillips C, Essick GK, Faraway JJ, Barlow SM, Losken HW et al. Functional outcomes of cleft lip surgery. Part 1: study design and surgeon ratings of lip disability and need for lip revision. *Cleft Palate Craniofac J* 2007;44:598–606.
13. Trotman C-A, Phillips C, Faraway JJ, Ritter K. Association between subjective and objective measures of lip form and function: an exploratory analysis. *Cleft Palate Craniofac J* 2003;40:241–8.
14. Trotman C-A, Faraway JJ, Phillips C, Aalst JV. Effects of lip revision surgery in cleft lip/palate patients. *J Dent Res* 2010;89:728–32.
15. Holtmann B, Wray RC. A randomized comparison of triangular and rotation-advancement unilateral cleft lip repairs. *Plast Reconstr Surg* 1983;71:172–9.
16. Reddy GS, Webb RM, Reddy RR, Reddy LV, Thomas P, Markus AF. Choice of incision for primary repair of unilateral complete cleft lip: a comparative study of outcomes in 796 patients. *Plast Reconstr Surg* 2008;121:932–40.
17. Reddy SG, Reddy RR, Bronkhorst EM, Prasad R, Kuijpers-Jagtman AM, Bergé S. Comparison of three incisions to repair complete unilateral cleft lip. *Plast Reconstr Surg* 2010;125:1208–16.
18. Yamada T, Mori Y, Minami K, Mishima K, Sugahara T. Three-dimensional facial morphology, following primary cleft lip repair using the triangular flap with or without rotation advancement. *J Craniomaxillofac Surg* 2002;30:337–42.
19. Okawachi T, Nozoe E, Nishihara K, Nakamura N. 3-dimensional analyses of outcomes following secondary treatment of unilateral cleft lip nose deformity. *J Oral Maxillofac Surg* 2010;69:322–32.
20. Hood CA, Bock M, Hosey MT, Bowman A, Ayoub AF. Facial asymmetry – 3D assessment of infants with cleft lip & palate. *Int J Pediatr Dent* 2003;13:404–10.
21. Tanikawa C, Takada K, Aalst JV, Trotman C-A. Objective three-dimensional assessment of lip form in patients with repaired cleft lip. *Cleft Palate Craniofac J* 2010;47:611–22.