

Functional Outcomes of Cleft Lip Surgery. Part I: Study Design and Surgeon Ratings of Lip Disability and Need for Lip Revision

Carroll-Ann Trotman, B.D.S., M.A., M.S., Ceib Phillips, M.P.H., Ph.D., Greg K. Essick, D.D.S., Ph.D., Julian J. Faraway, Ph.D., Steven M. Barlow, Ph.D., H. Wolfgang Losken, M.B.Ch.B., John van Aalst, M.D., M.A., Lyna Rogers, R.D.H., B.S.

Objective: Children with a cleft of the upper lip exhibit obvious facial disfigurement. Many require multiple lip surgeries for an optimal esthetic result. However, because the decision for lip revision is based on subjective clinical criteria, clinicians may disagree on whether these surgeries should be performed. To establish more reliable, functionally relevant outcome criteria for evaluation and treatment planning, a clinical trial currently is in progress. In this article, the design of the clinical trial is described and results of a study on subjective evaluations of facial form by surgeons for or against the need for lip revision surgery are presented.

Design: Parallel, three-group, nonrandomized clinical trial and subjective evaluations/ratings of facial views by surgeons.

Subjects: For the clinical trial, children with repaired cleft lip and palate scheduled for a secondary lip revision, children with repaired cleft lip and palate judged not to need lip revision, and noncleft children. For the subjective evaluations, surgeons' facial ratings of 21 children with repaired cleft lip.

Analysis: Descriptive and Kappa statistics of surgeons' ratings of repeated facial views and a recommendation of revision on viewing the prerevision and postrevision views.

Results: The surgeons' consistency in rating repeated views was moderate to excellent; however, agreement among the surgeons when rating individual participants was low to moderate.

Conclusions: The findings suggest that the agreement among surgeons was poor and support the need for more objective measures to assess the need for revision surgery.

KEY WORDS: *clinical trial, functional outcomes, lip revision surgery*

Ms. Trotman is Professor, Department of Orthodontics, University of North Carolina at Chapel Hill, North Carolina. Dr. Phillips is Professor, Department of Orthodontics, University of North Carolina at Chapel Hill, North Carolina. Dr. Essick is Professor, Department of Prosthodontics and Curriculum in Neurobiology, University of North Carolina at Chapel Hill, North Carolina. Dr. Faraway is Professor, Department of Mathematical Sciences, the University of Bath, Bath, England. Dr. Barlow is Professor of Speech-Language-Hearing Sciences and Disorders, Professor in the Program in Neuroscience and Human Biology, and Director, Communication Neurosciences Laboratory, University of Lawrence, Kansas. Mr. Losken is Professor, Department of Plastic and Reconstructive Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina. Dr. van Aalst is Assistant Professor and Director of Pediatric and Craniofacial Services, Department of Plastic and Reconstructive Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina. Ms. Rogers is Research Analyst, Department of Orthodontics, University of North Carolina at Chapel Hill, North Carolina.

Submitted July 2006; Accepted February 2007.

This study was supported by grant R01 DE13814-01A1 from the National Institute of Dental Research.

Address correspondence to: Ms. Carroll-Ann Trotman, CB #7450, 275 Brauer Hall, Chapel Hill, NC 27599-7450. E-mail carroll-ann.trotman@dentistry.unc.edu.

DOI: 10.1597/06-124.1

For a child with a cleft of the lip with or without a cleft palate, the decision to surgically revise the lip is based on a subjective evaluation of lip form and function that is made by the surgeon either independently or in conjunction with the patient and parents (Marsh, 1990). Subjective evaluation defines the current standard of care for patients who are candidates for lip revision; however, recent research has demonstrated many limitations with the use of subjective assessments. Of particular concern are the lack of agreement among clinicians (Asher-McDade et al., 1991; Tobiason et al., 1991; Ritter et al., 2002; Marrant and Shaw, 1996) and a tendency for the assessment of lip form to confound the assessment of lip function (Ritter et al., 2002). For example, when the severity of the deformity of static faces (i.e., faces at rest) was rated subjectively by clinicians, interexaminer agreement ranged from low (Asher-McDade et al., 1991) to good (Tobiason et al., 1991; Ritter et al., 2002), while for the subjective evaluations of cleft faces during movement, agreement among clinicians was consistently reported to be poor (Marrant and Shaw, 1996; Ritter et al., 2002). Moreover, the extent of lip scarring influenced ob-

surgeons' perceptions of impaired movements of the lips: the more severe the perception of scarring, the more severe was the perception of impaired movement (Ritter et al., 2002). These findings call into question the reliance on subjective evaluations as the only determinant for subsequent soft tissue surgeries (Trotman et al., 2003) and highlight the need for objective measures of lip disability to supplement subjective evaluations.

Abnormalities in facial (lip) form and function may be attributed to three separate mechanisms: (1) mechanical limitations in movement of the perioral tissues due to the effects of scarring, (2) impaired muscle force dynamics, and (3) impaired sensorimotor integration. Regarding the first mechanism, the effects of scarring, objective measurements confirmed that patients with a cleft of the lip have impairments in the maximum extent to which they can move their facial tissues (Trotman et al., 1998; Trotman and Faraway, 1998; Trotman et al., 2000). Relevant to the second mechanism, muscle force dynamics, a number of studies have demonstrated that measures of lip force can be used to characterize both the strength and fine motor control of facial muscles in normal individuals and in patients with somatomotor disorders of the central nervous system (Barlow and Abbs, 1983, 1984, 1986; Barlow and Rath, 1985; Barlow and Netsell, 1986). The results of pilot studies have demonstrated impairments in maximum force capacity and motor control of the lips of children with cleft lip (D'Antonio et al., 1994, 1995). Equally important is the concern related to the third mechanism, sensory integrative (sensorimotor) function, which is defined as the "use of all sensory input as meaningful information which the individual reacts with a well-organized adaptive response" (Chapparo et al., 1981). Dysfunction is characterized by a failure of different bilateral structures in the face and structures in the vertical thirds of the face (i.e., middle to lower facial thirds) to function in a coordinated manner. Normal sensory function and sensation are requisite for normal perioral motor function (Stranc et al., 1987; Essick, 1998). Importantly, the above studies found impairments in motor control not only in the upper lip but also in the lower lip, and our initial studies indicated that many patients with cleft of the upper lip do not have normal sensation (Essick et al., 2005).

To date, however, the gold standard for the assessment of lip function in the patients with a repaired cleft of the lip continues to be subjective clinical assessments. Objective measures that quantify facial soft tissue function in terms of movement, muscle force dynamics, or sensory integrative function have not been employed in formal studies or to monitor surgical treatment outcomes. The objective of this article is twofold: (1) to describe the overall design of a clinical trial to assess functional outcomes of cleft lip surgery and discuss issues related to the conduct of the trial and (2) in participants with a repaired cleft lip, to compare the results of subjective evaluations by surgeons for or against the need for lip revision surgery.

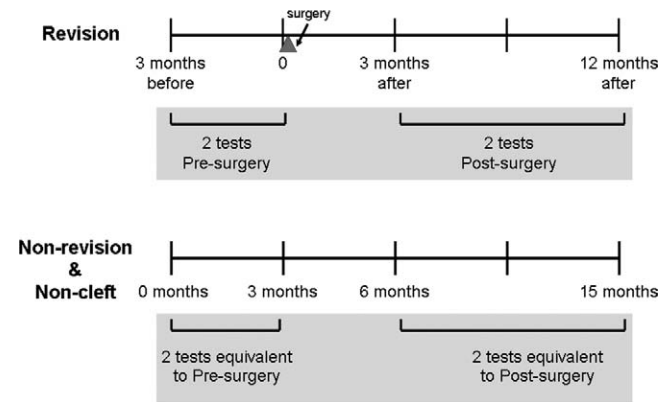


FIGURE 1 Schematic of the testing times for the three participant groups. The revision participants are tested twice before revision (–3 and 0 months) and twice after revision (3 and 12 months). The nonrevision and noncleft participants are tested at similar time intervals (0, 3, 6, and 15 months).

DESIGN AND CONDUCT OF THE CLINICAL TRIAL

Objectives and Organization

This study is designed to determine the differences in soft tissue circumoral function between children with a cleft of the lip (with or without a cleft palate) and children who do not have a cleft and to determine whether lip revision is effective in normalizing soft tissue function. Also, given the subjective nature of clinical assessments of facial/perioral form and function by surgeons, an exploratory aim was included to obtain estimates of concordance in subjective assessments among experienced surgeons for planning of future studies aimed at improving these types of clinical assessments. These assessments of the face were made with the face at rest and during different movements and are described in greater detail later.

The investigative team is composed of individuals from the Facial Animation Laboratory, Sensory Laboratory, Craniofacial Center, and Clinical Research Data Center at the University of North Carolina School of Dentistry (UNCSOD), the Statistics Laboratory at the University of Bath, and the Communication Neuroscience Laboratories at the University of Kansas. In addition, an appointed Data Safety and Monitoring Board met annually to monitor patient safety and ethical considerations and to review patient recruitment, retention, and data analysis. The study was a nonrandomized, three-group, parallel clinical trial that was confined to a single center and a single surgeon. The quantitative measures and subjective clinical measures of facial form and function were collected over a 15-month period, and for the participants who had lip revision surgery, the measures were collected before and after the surgery (Fig. 1).

Enrollment

Children and adolescents with cleft lip and palate were recruited from the Craniofacial Center at the UNCSOD. There

were three general groups of participants: (1) those with cleft lip and palate judged by the surgeon to be candidates for secondary lip revision surgery based on the current standards of care and thus recommended for, and elected to undergo, lip revision surgery; (2) those with cleft lip and palate who had not received a lip revision surgery at the time of enrollment (this nonrevision group was composed of participants who may have been recommended for lip revision surgery by the surgeon but had chosen not to proceed with the surgery as well as those participants who were not recommended to receive surgery); and (3) a group of noncleft participants recruited as controls from clinics at the UNCSOD. To ensure a similar age and sex distribution among the three groups, the participants were matched on age and sex using group frequency matching.

Selection Criteria

The inclusion and exclusion criteria that apply to participants in all three groups were as follows.

Inclusion

- Interest/parent willingness to participate in the study
- An ability to comprehend verbal instructions
- An age range of 5 to 21 years

Exclusion

- Previous orthognathic surgery
- A diagnosis of a craniofacial anomaly other than cleft lip and palate
- A medical history of diabetes, collagen vascular disease, systemic neurologic impairment
- Mental or hearing impairment to the extent that comprehension or ability to perform the tests is hampered

Additional criteria that applied only to the participants with a cleft lip and palate were as follows.

Inclusion

- A previously repaired complete unilateral or bilateral cleft lip with or without cleft palate
- For the lip revision participants, the professional clinical recommendation by the Craniofacial Center's plastic surgeon for a lip revision

Exclusion

- Previous lip revision surgery or other facial soft tissue surgery within 2 years of the planned revision surgery

Participants who met the selection criteria were recruited and screened in the Craniofacial Center, the Graduate Orthodontics Clinic, the Pediatric Dentistry Clinic, and the Ortho-

odontics Faculty Practice at the University of North Carolina. No subject was excluded from participation on the basis of sex, race, or ethnic background. The purpose and protocol of the study was explained to the participant(s) and parent(s), and informed consent and assent were obtained. Consent and HIPAA documents were approved by the School of Dentistry Human Subjects Institutional Review Board. Group frequency matching was performed periodically (every 3 months) to track the adequacy of balance among the three groups.

Assessment Times

The participants in the three groups were followed longitudinally for a 15-month period and tested at four times (Fig. 1). Specifically, the participants who had revision surgery were tested in a run-in period of approximately 3 months (time -3) and just before surgery (baseline, time 0). These revision participants received surgery soon after time 0. Testing was repeated at 3 months (time 3) and 12 months after the surgery. The test at 3 months after surgery was selected to coincide with a time point used in evaluating patients after trigeminal nerve injuries (Van Boven and Johnson, 1994; Karas et al., 1990; Yoshida et al., 1989) and the early resolution of edema and swelling, and the test at 12 months after surgery was chosen because scar maturation is completed in approximately 12 months. The nonrevision and noncleft participants, who do not have surgery, were tested at corresponding times, thus maintaining fairly consistent intervals between testing sessions for the nonrevision and noncleft participants (i.e., sessions at -3, 0, 3, and 12 months; Fig. 1) to provide time intervals of testing comparable to that for the revision participants. Participants were asked to allow 1 full day of testing for each of the four sessions.

The change in lip function from prerevision to postrevision in the participants who had lip revision included the treatment effect as well as the change due to maturation across sessions at 0, 3, and 12 months. Given that we did not know whether participants in the revision and nonrevision groups would demonstrate normal maturational change, the nonrevision participants allowed an assessment of maturational change in children who were expected to be impaired. The noncleft group provided information on normal maturational change over the 15-month period of follow-up. It was expected, however, that any systematic difference in the maturational change or initial functional impairment in lip function between the revision and nonrevision participants was minimal.

Surgery

The Craniofacial Team at UNCSOD met weekly to plan the coordinated treatment for all patients enrolled with the team. Prior to the final recommendations being made for the treatment of a patient, team members discussed the recommendations made by the different disciplines on the team. Generally, recommendations for lip revision were made by the plastic surgeon, and discussions occurred with the orthodontist and

TABLE 1 Description of Quantitative Outcomes

<i>Measure</i>	<i>Description</i>
Facial movement	Three-dimensional measures of facial tissues during facial animations such as smile, lip purse, cheek puff, grimace, and mouth opening (Trotman et al., 2000; Weeden et al., 2001; Trotman et al., 2003)
Lip force	Force measures made at the midline of the upper and lower lips (Barlow, 1998)
Sensory	
Perception thresholds	Two-point perception thresholds measured on the right and left sides of both the upper and lower vermillion (Chen et al., 1995; Feldman et al., 1997; Essick et al., 2001)
Temperature thresholds	Warmth-detection and cool-detection thresholds measured on the right and left sides of both the upper and lower vermillion (Martinez, 1996; Essick et al., 2004)
Sensorimotor integration	
Perioral reflex activity	Punctate (pressure-like) low-level mechanical stimulus to each right and left, upper and lower vermillion; mechanically evoked electromyography activity is sampled bilaterally from sites over the orbicularis oris superior and inferior muscle before, during, and after the stimulation (Barlow et al., 1993; Barlow and Bradford, 1996)

oral surgeon regarding the desired esthetic outcomes of the revision surgery as well as the need for, and timing of, bone-grafting procedures.

The revision participants had surgery within a few days of the testing at time 0 (Fig. 1). The surgery was performed by the study surgeon, who was experienced in cleft care. All surgeries were secondary correction of a cleft lip and fell into two categories: (1) a full-thickness lip revision or (2) a partial-thickness lip revision. In a full-thickness lip revision, the lip was revised by re-creation of the defect and reclosure. The scar from the previous surgery was excised completely. If the orbicularis oris muscle was not repaired originally, the muscle was realigned and repaired. In all patients, the mucosa, muscle, dermis, and skin were repaired. All unilateral repairs were by the rotation-advancement technique. If the original repair was by a different technique, the original scars were excised in such a way as to allow rotation-advancement closure. For bilateral clefts, the secondary repair was in the configuration of a modified Manchester repair, with a narrower philtrum. For the partial-thickness lip revision, the full thickness of the skin was divided with a partial division of the orbicularis oris muscle. This approach was used in patients who have discrepancies in vertical lip length and symmetry. For either category of lip revision, concomitant nasal correction, when felt to be indicated by the surgeon, was performed by a standard open rhinoplasty with a V-shaped columellar incision. If a cartilage graft was needed, it was harvested from either or both ears via a postauricular incision.

Quantitative Outcomes

Table 1 gives the quantitative outcome measures that are the best surrogates of facial functional disability. The primary quantitative outcome measure is facial (lip) movement; thus, sample size calculations were based on this measure. Secondary outcomes were measures of lip force, sensation, and sensorimotor integration.

Sample Size

Estimates of variability and mean differences for facial movement during the smile and cheek-puff animations be-

tween children with and without cleft lip were obtained from nine children with complete unilateral and bilateral cleft lip and 50 noncleft subjects. These children were studied prior to the initiation of this clinical trial (Trotman et al., 2000). The pilot data indicated considerable variability in asymmetry and magnitude of displacement among the participants with cleft lip and palate; however, the primary focus was the difference in the change in movement measures between participants who had a lip revision and the noncleft healthy children. It was hypothesized that there would be large mean facial movement changes on the order of 50% following lip revision and a much smaller mean change (approximately 10%) resulting from maturation in both the noncleft and the nonrevision cleft lip groups.

Given this background, a first approximation of a sample size per group that would be sufficient to detect a large effect size ($\geq .80$; Cohen, 1998) between the primary comparison groups of interest and the revision and noncleft control groups, with a level of significance of .05 and 90% power, was calculated using an unpaired *t* test approach (NQuery version 5; Elashoff, 2002) to compare the average 12-month change in the facial movement measures. The estimated sample size of 34 per group was used to generate power curves for a one-way analysis of variance with three groups and a single one-way between-means contrast (comparing two of the three groups assuming the overall test is significant) when the effect size or sample size was varied. The inspection of the power curves (Fig. 2) and the likelihood that power would be improved by the inclusion of covariates (e.g., age, sex) in the analyses that would explain a portion of the variation in the facial movement summary measures suggested that an effective sample size goal of 34 children in each of the three groups would be appropriate.

Statistical Analysis of the Primary Outcome Measures

The study used a three-group, parallel design, and participants were followed for 15 months. All enrolled participants with at least the -3 visit data were included in the full analysis set. The overall alpha level for each analysis was set at .05. No pairwise group comparisons were performed unless the

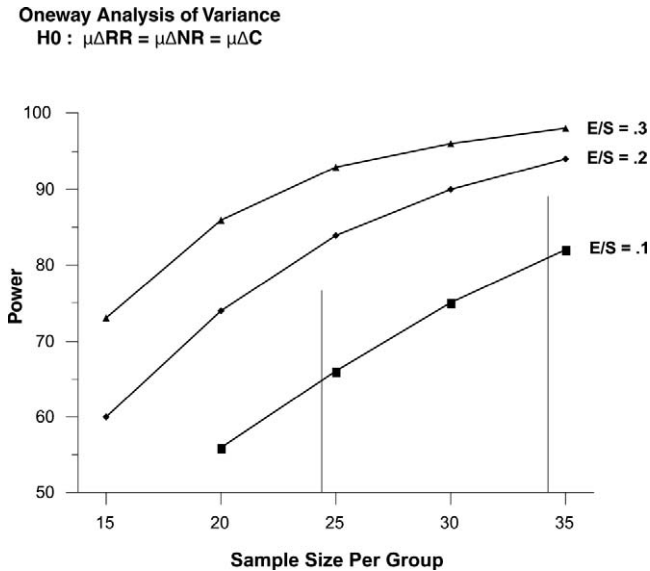


FIGURE 2 Power curves for a one-way analysis of variance with three groups and single one-way between-means contrast (comparing two of the three groups assuming the overall test is significant) generated by varying the effect size or sample size.

overall F value for fixed effects was statistically significant ($p < .05$). Based on closed-testing principles, if the overall hypothesis of equality among the three groups was rejected, each of the three hypotheses for equality of pairs of treatments could be conducted at the .05 significance level (Westfall et al., 1999).

The general analytical approach for all the outcome measures was linear mixed-effect modeling. Data from different participants were considered independent, whereas data within a participant (multiple sites tested on the face per visit or multiple visits) were expected to be correlated. Thus, a subject was considered to have a cluster of correlated response data for each outcome. In the linear mixed model, modeling of the variances and covariances was achieved through specification of random effects and/or specification of the variance matrix of the error vector for a subject.

SURGEON RATINGS OF LIP DISABILITY AND NEED FOR REVISION SURGERY

Method of Subjective Evaluation

Eight plastic surgeons from different craniofacial centers across the United States who were experienced in cleft care viewed and rated photographs and videotapes of 11 revision and 10 nonrevision study participants. The revision and nonrevision participants were selected from all the participants in the clinical trial to represent a wide range of lip scars. For each revision participant, photographic and videotaped views recorded at baseline and at 12 months postsurgery were selected. For the nonrevision participants, similar views recorded at corresponding times were selected. The two views for each of the 21 participants were compiled in random order on a

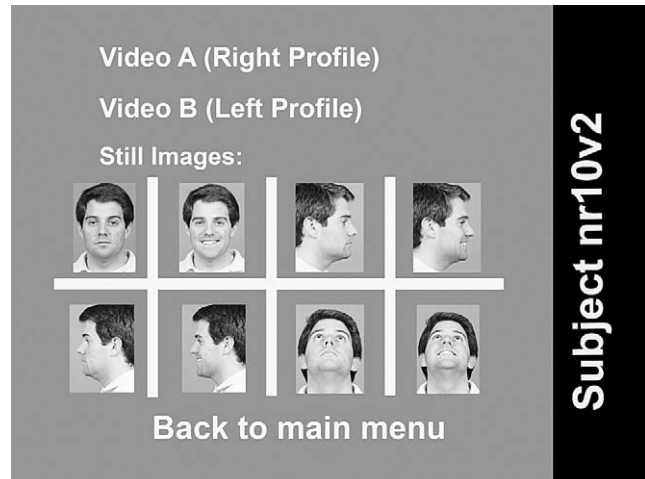


FIGURE 3 Example of the screen for viewing participant photographs and videotapes. Selecting a still image will enlarge it. Selecting video will play the different animations with frontal and profile images combined.

DVD for viewing. In addition, to determine consistency in ratings by individual surgeons, either the baseline or 12-month view of eight participants was repeated on the DVD. Thus, each surgeon viewed a total of 50 sets of photographs and videotapes: eleven baseline and eleven 12-month views for the revision participants; ten baseline and ten 12-month views for the nonrevision participants; and eight repeated views.

For optimal and consistent viewing quality, all the views (Fig. 3) were displayed to each surgeon independently on a computer monitor with a 17-inch screen. Each surgeon was blinded to the participants' identity, group membership (revision versus nonrevision), and surgical history (baseline versus 12 months). Surgeons were shown the photographic still images first, followed by the video images. By selecting a still image (Fig. 3) for a participant, that photograph was magnified for further evaluation. By selecting either video A or video B (Fig. 3), the surgeon was able to view either a frontal and right profile image (video A) or frontal and left profile image (video B) of the participant performing a smile, lip purse, cheek puff, mouth opening, and natural smile. In addition, surgeons could view the participants speaking the phrase "put the baby in the buggy."

Prior to the start of the viewing, the research assistant read the same set of instructions to each surgeon explaining the process of viewing the DVD. The surgeons were not told which participants had received surgery, but they were told that participants would be viewed more than once. On viewing each participant, the surgeon was asked the following questions: (1) Do you think this person would benefit from a lip revision surgery? (2) If yes, would the surgery be a minor revision (involving skin only), full-thickness revision (involving the entire depth of skin and muscle), or partial-thickness revision (skin with partial muscle involvement)? and (3) Which view (still or dynamic video images) gave you the most information to make your decision?

Each surgeon was allowed to proceed with the viewing at

TABLE 2 Number of Participants for Whom There Was Agreement Among the Surgeons on the Recommendation for Revision or No Revision for the 21 Subjects at Baseline and at 12 Months

Percentage of Surgeons in Agreement	Nonrevision Recommendations (No. of Patients)	Revision Recommendation (No. of Patients)
Baseline		
100% (8 of 8 surgeons agree)	0	1
~88% (7 of 8 surgeons agree)	0	3
75% (6 of 8 surgeons agree)	4	0
~63% (5 of 8 surgeons agree)	3	3
12 months		
100% (8 of 8 surgeons agree)	1	0
~88% (7 of 8 surgeons agree)	1	0
75% (6 of 8 surgeons agree)	5	1
~63% (5 of 8 surgeons agree)	2	4

his or her own pace. The responses to the questions were recorded by the research assistant directly by hand on a form and tape recorded with a handheld recorder for later verification. After verification, the responses were entered into a spreadsheet for data processing. For each surgeon, the entire session of viewing and rating the participants lasted approximately 2½ hours.

Statistical Analysis

Standard descriptive statistics were generated from the ratings of the surgeons and included the following. (1) The percentage agreement for the repeated views was calculated for each surgeon to assess the consistency in the replicate recommendations. (2) The number of surgeons who agreed on a recommendation for a participant, and the Kappa statistics for agreement between pairs of surgeons, was calculated to assess the consistency among the surgeons in the recommendations made. (3) For those participants who had a revision surgery, the number of surgeons who categorized participants into one of the four possible combinations of recommendation (revision at presurgery and no revision at postsurgery, revision at presurgery and revision at postsurgery, no revision at presurgery and no revision at postsurgery, and no revision at presurgery and revision at postsurgery) was calculated to assess the consistency among the surgeons in providing a revision recommendation and to assess whether the surgeon's recommendations suggested that, in the surgeon's perception, the revision improved, did not change, or worsened the participant's condition.

RESULTS

A surgeon's classification of a patient's need for lip revision was categorized as nonrevision (either no or only minor revision recommended) or as revision (a partial- or full-thickness lip revision recommended). On viewing the participants, the surgeons commented that they welcomed the opportunity to systematically view the facial photographs and animated video

TABLE 3 Surgery Recommendations for the Randomly Presented Presurgery and Postsurgery Viewings of the 11 Revision Participants Made by Each of the Eight Surgeons (A to H) and Possible Surgeon Perceptions*

Surgeon	Condition Improved	No Perceptible Difference	Surgery Not Needed	Condition Worsened
A	2 (22%)	4 (44%)	2 (22%)	1 (11%)
B	4 (44%)	3 (33%)	2 (22%)	0 (0%)
C	4 (36%)	7 (64%)	0 (0%)	0 (0%)
D	1 (9%)	4 (36%)	2 (18%)	4 (36%)
E	3 (27%)	1 (9%)	7 (64%)	0 (0%)
F	2 (18%)	5 (45%)	3 (27%)	1 (9%)
G	3 (27%)	1 (9%)	7 (64%)	0 (0%)
H	2 (18%)	2 (18%)	5 (45%)	2 (18%)

* Condition improved: revision at presurgery and no revision at postsurgery. No perceptible difference: revision at presurgery + revision at postsurgery. Surgery not needed: no revision at presurgery + no revision at postsurgery. Condition worsened: no revision at presurgery + revision at postsurgery.

images, and they all felt that the video images provided useful additional diagnostic information.

Consistency for Each Surgeon

Five of the eight surgeons demonstrated 100% agreement on their decisions for repeated ratings. Two of the eight surgeons had an 87% agreement, in which only one participant was rated differently after the second viewing. One surgeon had a 75% agreement for repeated ratings, with two participants rated differently after the second viewing.

Consistency Among Surgeons

All eight surgeons made the same recommendation for only two of the 42 viewings (~5%): baseline views of one participant (equivalent to ~5% of the baseline views) and 12-month views of one participant (~5%). Seven of the surgeons, although not consistently the same seven, agreed on the recommendation for 3 (14%) of the 21 baseline viewings and 1 (5%) of the 12-month viewings. Six of the surgeons, although not consistently the same six, agreed on the recommendation for 4 (19%) of the 21 baseline viewings and 6 (29%) of the 12-month viewings. Five of the surgeons, although not consistently the same five, agreed on the recommendation for 6 (29%) of the 21 baseline viewings and 6 (29%) of the 12-month viewings (Table 2). As an additional assessment of agreement, Kappa statistics were calculated for all possible pairs of surgeons (28 pairs): Kappa values ranged from .0 to .57 for the baseline views and from .0 to .44 for the 12-month views. All but one of the Kappa values were less than .45, implying poor agreement among the surgeons.

Revision Participants Only: Surgery Recommendations at Baseline and 12 Months

Table 3 provides the presurgery and 12-month follow-up recommendations made by each surgeon for each participant who had a revision during the study. These recommendations can be categorized into the following combinations.

1. Revision at baseline and no revision at 12 months. This recommendation implies that in the surgeon's perception, the revision improved the participant's lip form and/or function.
2. Revision at baseline and a revision at 12 months.
3. No revision at baseline and no revision at 12 months. Both recommendations 2 and 3 imply that the surgeon did not perceive a substantial enough difference to differentiate the baseline viewing recommendation from the 12-month viewing.
4. No revision at baseline and a revision at 12 months. This recommendation implies that in the surgeon's perception, the revision either had no effect or the surgery may have worsened the participant's facial form and/or function.

The results showed that there was considerable variability among the surgeons with regard to their perception of the need for revision from the presurgery views (Table 3). This variability ranged from 36% (surgeons E, G, and H) to 100% (surgeon C) of the participants. The percentage of participants who, in the independent, masked ratings by the surgeons, were judged to need a revision at baseline but no revision at 12 months, implying that a perceptible improvement was apparent, was low (9% to 44%). For all surgeons, the recommendation of either no revision at both baseline and 12 months or a revision at both baseline and at 12 months was given for most of participants (from 45% to 73%). The combination of no revision at baseline but a revision at 12 months, implying a perceptible worsening, was not given to any of the participants by four of the surgeons. The other four surgeons chose this latter combination of recommendations for 9% to 36% of the participants. The percentage agreement among pairs of surgeons ranged from 0% to 64%. None of the pairs of surgeons agreed on more than 75% of the recommendations, again implying a low to moderate agreement among the surgeons.

DISCUSSION

The purpose of this trial was to assess the efficacy of lip revision surgery on lip and perioral function in a controlled clinical environment. The participants were recruited from the UNCSOD clinics and Craniofacial Center. As a result, the findings may not be generalizable to a broad population; however, should the quantitative approaches used in this study be sufficiently sensitive to detect effects related to lip revision surgery, then these methodologies could be used to compare the effects of different surgical techniques on function in future randomized clinical trials and/or in clinical settings. In support of our approach to the assessment of perioral function are the results of the Eurocleft Intercenter studies, in which the findings of subjective ratings of nasolabial appearance in patients with cleft lip and palate, based on static photographic images, were found to be invalid as surgical outcome measures (Brattström et al., 2005). These researchers noted that "the rating of nasolabial outcome is a key area for further research." This

TABLE 4 Demographics of Subjects Recruited: Age, Gender, and Cleft Type by Group

Group	Total n	Age, Mean \pm SD		Female, n		Unilateral	Bilateral
		Male, n	n	Cleft Lip, n	Cleft Lip, n		
Noncleft	37	13.1 \pm 3.6	20	17	—	—	
Nonrevision	32	12.4 \pm 3.3	21	11	28	4	
Revision	31	12.1 \pm 4.0	18	13	24	7	

clinical trial has been totally devoted to outcomes of perioral surgery in patients with cleft lip and palate.

Recruitment

Recruitment goals for noncleft control participants were met (n = 37). As of July 2006, more than 90% of the recruitment goal had been achieved for the nonrevision (n = 32 of 34) and revision groups (n = 31 of 34). The refusal and dropout rates were low in all groups. In the revision group, one patient was dropped at the first visit because the child would not perform the tests. Four potential revision candidates declined participation in the study: one child refused, one child did not show for the first testing appointment, and the parents of two children declined to give parental permission. Recruitment in the revision and nonrevision groups continues. The mean ages of the participants by group (noncleft, nonrevision, and revision) and cleft lip status (unilateral and bilateral) are shown in Table 4. The DSMB assigned to this study met with the study investigators at UNCSOD prior to the start of recruitment and held yearly meetings with the investigators during the grant tenure. Also, teleconferences with the DSMB were held every 6 months to monitor all aspects of the trial. As stated earlier, the nonrevision participants may have been judged to not need revision surgery or they may have been judged by the surgeon to be candidates for lip revision but may have elected not to proceed with the surgery. Of the 32 nonrevision participants, 8 were judged to need a revision and elected not to proceed with the surgery.

Subjective Assessments

The subjective assessments for or against lip revision surgery in the participants with a cleft lip indirectly addressed the rating of esthetic outcome of lip revision surgery in these individuals. Our findings demonstrated that the surgeons' consistency in rating repeated views of the same participant was moderate to excellent; however, the overall agreement among the surgeons when rating individual participants was low to moderate. Any recommendation for revision surgery by a surgeon depends on the surgeon's assessment of the participant's lip form and function at the time of the clinical evaluation and the surgeon's own clinical expertise related to whether she or he can improve on the lip form and/or function. Therefore, it was not surprising that different surgeons had different thresholds for revision and that there was poor interobserver reliability. These findings are similar to those of other studies in

which there was poor agreement among evaluators (Asher-McDade et al., 1991; Morrart and Shaw, 1996; Ritter et al., 2002) when subjectively rating individuals' facial features. Thus, substantially more work is needed in developing consensus among surgeons, and this work is one aspect of our current research.

During the rating sessions in this study, the baseline (before surgery) and 12-month (after surgery) views of the revision participants were presented to the surgeons in a random manner, and the surgeons did not know which were the before or after surgery views. The expectation was that for most of the revision participants, surgeons would recommend a revision at the baseline viewing and would not recommend a revision at the 12-month viewing. The findings show that the percentage of participants who were perceived by the surgeons as needing a lip revision at baseline and then at 12 months varied considerably among the surgeons, suggesting that the surgeons did not process or rank the information presented to them in the same way. In fact, after viewing the 12-month sets of images for the revision participants, all the surgeons made the recommendation of further revision surgery for at least one participant. One conclusion of this finding could be that the outcome of the revision surgery in some participants was ineffective; however, this conclusion would be premature. Other factors could lead to this outcome. For example, in certain cases, a determination may be made by the surgeon that more than one revision may be necessary to achieve a satisfactory surgical outcome. Such a situation may occur with revisions for a bilateral cleft lip. The surgeon may stage the surgery, revising one side of the lip at any one time and the other side at a later date, or the surgeon may revise both sides on the lower part of the lip and defer the upper part that is related to the columella for a subsequent surgery. In this study, however, of the 11 revision participants who were rated by the surgeons, only 3 had a repaired bilateral cleft lip, and all of their revisions were planned and fully completed by the surgeon with one surgery.

Another factor affecting the decision by surgeons for additional surgery on viewing the 12-month images may be related to the length of time at which this judgment was made and the extent of visible scarring at that time. Some patients develop hypertrophic scarring following surgery that requires a considerable period of time to resolve. The 12-month follow-up of this study may not have been long enough for surgeons to make a final decision regarding resolution of scarring. In addition, the presence of a scarring confounds ratings or judgments of lip movement (Ritter et al., 2002), which may have affected the surgeon ratings when they used the video images. The focus of this study, however, was on the agreement among surgeons for or against the need for revision surgery. These additional factors, although interesting, would have had a minimal impact on these results. One other factor, described earlier, that could have affected surgeon ratings and may have confounded the results somewhat was that surgeons may have perceived a need for revision surgery but may not have believed that they could achieve an improvement of the lip form

and/or function, a factor that relates to surgeons' willingness to perform the surgery and their competence.

REFERENCES

- 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.
- Barlow SM. *Handbook of Clinical Speech Physiology Part II*. San Diego, CA: Singular; 1998.
- Barlow SM, Abbs JH. Fine force and position control of select orofacial structures in the upper motor neuron syndrome. *Exp Neurol*. 1986;94:1-15.
- Barlow SM, Abbs JH. Force transducers for the evaluation of labial, lingual, and mandibular motor impairments. *J Speech Lang Hear Res*. 1983;26:616-621.
- Barlow SM, Abbs JH. Orofacial fine motor control impairments in congenital spasticity: evidence against hypertonus-related performance deficits. *Neurology*. 1984;34:145-150.
- Barlow SM, Bradford PT. Comparison of perioral reflex modulation in the upper and lower lip. *J Speech Lang Hear Res*. 1990;39:55-75.
- Barlow SM, Finan DS, Bradford PT, Andreatta RD. Transitional properties of the mechanically evoked perioral reflex from infancy throughout adulthood. *Brain Res*. 1993;623:181-188.
- Barlow SM, Netsell R. Differential fine force control of the upper and lower lips. *J Speech Lang Hear Res*. 1986;29:163-169.
- Barlow SM, Rath EM. Maximum voluntary closing forces in the upper and lower lips of humans. *J Speech Lang Hear Res*. 1985;28:373-376.
- Brattström V, Mølsted K, Prah-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.
- Chapparo CJ, Yerxa EJ, Nelson JG, Wilson L. Incidence of sensory integrative dysfunction among children with orofacial cleft. *Am J Occup Ther*. 1981; 35:96-100.
- Chen CC, Essick GK, Kelly DG, Young MG, Nestor JM, Masse B. Gender-, side- and site-dependent variations in perioral spatial resolution. *Arch Oral Biol*. 1995;40:539-548.
- Cohen J. *Statistical Power Analysis for Behavioral Sciences*. City, NJ: Lawrence Erlbaum Associates; 1998.
- D'Antonio L, Barlow S, Campanelli J, Finan D, Umeda A. Comparison of fine force control of the perioral muscles in individuals with and without repaired cleft lip. Presented at the Joint Session of the American Cleft Palate-Craniofacial Association and the American Society of Craniofacial Surgeons, 1995; Tampa, FL.
- D'Antonio L, Barlow SM, Umeda A, Finan D, Herber S, Renshaw J. Upper and lower lip force control in individuals with repaired unilateral or bilateral cleft lip. Presented at the Annual Meeting of the American Cleft Palate-Craniofacial Association; 1994; Toronto, Canada.
- Elashoff JD. *nQuery Advisor Version 5.0 User's Guide*. Los Angeles: Statistical Solutions; 2002.
- Essick GK. Invited discussion of Lemke RR, Clark GM, Bays RA, Tiner BD, Rugh JD. Effects of hypesthesia on oral behaviors of the orthognathic surgery patient. *J Oral Maxillofac Surg*. 1998;56:158-160.
- Essick GK, Austin S, Phillips C, Kiyak A. Short-term sensory impairment following orthognathic surgery. *Oral Maxillofac Surg Clin North Am*. 2001; 13:295-313.
- Essick GK, Dorian C, Rumley S, Rogers L, Young M, Trotman CA. Report of altered sensation in patients with cleft lip. *Cleft Palate Craniofac J*. 2005; 42:178-184.
- Essick GK, Guest S, Martinez E, Chen C, McGlone F. Site-dependent and subject-related variations in perioral thermal sensitivity. *Somatosens Mot Res*. 2004;21:159-175.
- Feldman JA, Essick GK, Zuniga JR, Phillips C. Inter-examiner reliability of three subjective clinical neurosensory tests. *Int J Adult Orthod Orthognath Surg*. 1997;12:273-285.

Cleft Palate–Craniofacial Journal, ??? ???? , Vol. ?? No. ?

- Karas ND, Boyd SB, Sinn DP. Recovery of neurosensory function following orthognathic surgery. *J Oral Maxillofac Surg.* 1990;48:124–134.
- Marsh JL. When is enough enough? Secondary surgery for cleft lip and palate patients. *Clin Plast Surg.* 1990;17:37–47.
- Martinez EI. *Gender-, Side-, and Site-Dependant Variations in Human Perioral Thermal Sensitivity.* Chapel Hill: University of North Carolina; 1996. Thesis.
- Morrant DG, Shaw WC. Use of standardized video recordings to assess cleft surgery outcome. *Cleft Palate Craniofac J.* 1996;33:134–142.
- Ritter K, Trotman C-A, Phillips C. Validity of subjective evaluations for the assessment of lip scarring and impairment. *Cleft Palate Craniofac J.* 2002; 39:587–596.
- Stranc MF, Fogel M, Dische S. Comparison of lip function: surgery vs radiotherapy. *Br J Plast Surg.* 1987;40:598–604.
- Tobiason JM, Hiebert JM, Boraz RA. Development of scales of severity of facial cleft impairment. *Cleft Palate Craniofac J.* 1991;28:419–424.
- Trotman C-A, Faraway JJ. Sensitivity of a method for the analysis of facial mobility: II. Inter-landmark separation. *Cleft Palate Craniofac J.* 1998;35: 142–153.
- Trotman C-A, Faraway JJ, Essick GK. 3-D nasolabial displacement during movement in repaired cleft lip and palate patients. *Plast Reconstr Surg.* 2000;105:1273–1283.
- Trotman C-A, Faraway JJ, Silvester KT, Greenlee GM, Johnston LE Jr. Sensitivity of a method for the analysis of facial mobility: I. Vector of displacement. *Cleft Palate Craniofac J.* 1998;35:132–141.
- 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;40:241–248.
- Van Boven RW, Johnson KO. A psychophysical study of the mechanisms of sensory recovery following nerve injury in humans. *Brain.* 1994;117:149.
- Weeden JC, Trotman C-A, Faraway JJ. Three-dimensional analysis of facial movement in normal adults: influence of sex and facial shape. *Angle Orthod.* 2001;71:132–140.
- Westfall PH, Tobias RD, Rom D, Wolfinger RD, Hochberg Y. *Multiple Comparisons and Multiple Tests Using the SAS System.* Cary, NC: SAS Institute Inc.; 1999.
- Yoshida T, Nagamine T, Kobayashi T, Michimi N, Nakajima T, Sasakura H, Hanada K. Impairment of the inferior alveolar nerve after sagittal split osteotomy. *J Craniomaxillofac Surg.* 1989;17:271–278.