Association Between Subjective and Objective Measures of Lip Form and Function: An Exploratory Analysis

CARROLL-ANN TROTMAN, B.D.S., M.A., M.S.
CEIB PHILLIPS, PH.D., M.P.H.
JULIAN J. FARAWAY, PH.D.
KELLY RITTER, D.D.S., M.S.

Objective: In patients with cleft lip and palate, the aims of this study were to generate objective measures of different attributes of lip movement and to explore the utility of these objective measures by examining the association between examiners' subjective assessments with the objective measures.

Patients and Participants: Thirteen patients with unilateral cleft lip and palate with varying degrees of cleft scar severity were selected. All patients had a previously repaired complete unilateral cleft lip and palate.

Interventions: Photographs and videotape recordings were made of the patients with cleft at rest and during smiling. Measurements of lip movement were obtained by means of a motion analysis system.

Main Outcome Measures: The study sought to obtain rankings of cleft scar severity and impairment on a 6-point Likert scale by a lay and professional panel and measurements of displacement, asymmetry, speed, and velocity of upper lip during smiling.

Results: Displacement was the most consistent and valid objective measurement. An objective analysis of the entire upper lip provided the most information. In general, there was a decrease in the objective measures of upper lip movement as examiners' perceptions of facial appearance or disfigurement at rest and impairment during movement became worse. This relationship was stronger for the at-rest perceptions, implying that subjective assessments should be made with the face at rest.

Conclusions: Objective measures provided the promise for differentiation of the components of movement and should be used to supplement subjective evaluations of lip appearance at rest and during movement.

KEY WORDS: *cleft lip and palate, lip movement, scarring*

Subjective assessments of static lip form and, to a lesser degree, dynamic lip form or function represent the current standard of care upon which decisions are made regarding lip-revision surgeries (Marsh, 1990; Asher-McDade et al., 1991; Morrant and Shaw, 1996; Trotman et al., 2000). Recent research suggests that these assessments tend to be biased, and may not provide a sensitive enough measure for this decision-making process (Ritter et al., 2002). Marsh (1990) contends that because of the individualized nature of subjective assess-

ments, a child with a cleft may undergo unnecessary procedures because of the parents' or surgeon's unrealistic search for perfection. Objective measures of lip function could supplement subjective assessments and play a major role in the evaluation of patients who may need lip revision surgeries. Additionally, with current advancements in surgical techniques, it can be expected that surgical corrections are likely to result in incremental improvements rather than dramatic changes (Morrant and Shaw, 1996). The use of more precise objective measures that are sensitive enough to assess these incremental changes because of surgery would impact future clinical practices regarding the need for, timing, and nature of lip revisions (Trotman et al., 2000).

Development of objective measures to evaluate facial form (appearance) and function has been slow, in part because neither the utility nor the superiority of many objective measures considered to date has been convincing. In treatment planning of patients, linear and angular measures of facial form may be made from two-dimensional still photographs or cephalometric radiographs or directly on the patient (Riolo et al., 1974;

Broadbent et al., 1975; Coghlan et al., 1987; Roberts-Harry et al., 1991; Hurwitz et al., 1999). Although such measurements are objective, they do not measure those fine esthetic attributes upon which clinical decisions are based. For example, Roberts-Harry et al. (1991) found that the degree of nasal deviation, assessed by a constructed nose-displacement angle, did not correlate with observers' ratings of facial attractiveness. Given that surgical decisions for the patient with cleft are based largely on a subjective assessment of facial appearance and that subjective assessments are affected by professional experience, the need for better objective measures to supplement subjective assessments is unequivocal (Trotman et al., 2000; Ritter et al., 2002). Coghlan et al. (1987) suggest that it is only by the use of objective outcome measures that it will be possible "to show beyond personal opinion whether one (surgical) technique has demonstrable advantages over another."

In a previous study (Ritter et al., 2002), we investigated the utility of subjective assessments for evaluating facial form and movement in patients with cleft lip and palate. Intra- and interexaminer reliability was found to be good for the evaluations of faces at rest but not during movement; professionals gave evaluations of greater severity and impairment than laypersons, and professionals agreed when evaluating faces at rest more so than during movement. Additionally, the presence of an artificial lip scar affected perceptions of impairment during movement by viewers in both professional and lay panels.

In the present study, we focused on two separate issues. The first was a simplistic assessment of consistency associated with objective measures of lip movement that is an important concern when any new measurement is proposed. The second was to explore the relationship between subjective evaluations of lip appearance and objective measures of dynamic circumoral lip activity in patients with cleft. The specific aims, therefore, were twofold: to generate objective measures that describe different attributes of circumoral movement or function and to further explore the utility of objective assessments to evaluate lip form and function by examining the association between examiners' subjective assessments of lip form and function with the objective measures so generated. The research hypothesis was that the examiners' subjective evaluation of static and dynamic lip form are correlated with specific objective measures of lip function.

MATERIALS AND METHODS

The sample consisted of 13 patients with unilateral cleft lip and palate (UCLP), eight boys and five girls and women (age range: boys 9 to 17 years, girls and women 8 to 40 years). All patients were recruited from the University of North Carolina School of Dentistry Craniofacial Clinic. The selection criteria included: the presence of a previously repaired complete unilateral cleft lip and palate; no other craniofacial anomalies; and lip scars that ranged in severity from very mild to very severe. Of the 13 patients with UCLP, six had lip scars on the right side of the face and seven on the left; seven patients were in

orthodontic treatment at the time of the study; and six had completed some phase of orthodontic treatment.

Movements of the circumoral region of each patient during smiling were measured by means of a motion analysis system that is described below. At a separate sitting, the patients were videotaped repeatedly with the face at rest and during smiling (Ritter et al., 2002). From these video recordings, one at-rest view and one smiling view of each patient was selected and randomly combined on videotape for viewing. Two groups of panelists, a professional and lay group, were instructed to view and rate the circumoral regions of the patients. The panelists rated the appearance of scarring of the lip or the degree of disfigurement "at rest" on a 6-point Likert scale that ranged from "1 = not visible" to "6 = very severe"; then they rated the appearance of the lip during smiling or the degree of impairment on a 6-point Likert scale that ranged from "1 = no impairment" to "6 = very severely impaired." Specific details of the methods for this aspect of the study have been described in a previous manuscript (Ritter et al., 2002). The study protocol was approved by the UNC Institutional Review Board, and written consent was obtained from all patients prior to their participation in the study.

Recording Circumoral Movements

A video-based tracking system (Motion Analysis, Motion Analysis Corp., Santa Rosa, CA) was used to measure the circumoral movements of each patient during smiling. This system tracks retroreflective markers secured to the face. The movement of each marker was captured in real time by the tracking system (Trotman et al., 1998a, 1998b). Four analog video cameras were positioned in front of the patient to record the spatial positions of the markers as a function of time at a rate of 60 frames/second for 3 seconds. To obtain three-dimensional coordinate data for a marker, two cameras must record the marker position in space. Because markers on the face may be carried outside the field of view of the two primary cameras, two additional cameras were used to ensure that data from at least two cameras were recorded for all facial markers. Camera optics consisted of lenses with a focal length of 25 mm.

Prior to recording facial movements of each patient, the space within which the head was positioned was first calibrated. This calibration was completed with a cube-shaped metal space frame (20 cm on each edge) fitted with an array of 12 markers (Dimensional Inspection Laboratories had previously certified the position in space of the markers to an accuracy of $\pm$ 7.6 nm), and a 21.5-cm-long wand fitted with three markers (one marker on each end and one marker 4.5 cm from one end). Lens distortion was corrected automatically. Under conditions of the study, lens distortion determined by a 3-cm object positioned at the center and corners of the measurement space produced a mean error of 0.53 mm ($\pm$ 0.45).

Ten spherical retroreflective markers, each with a diameter of 2 mm, were attached by means of eyelash adhesive to specific sites on the facial skin of the upper lip of each patient
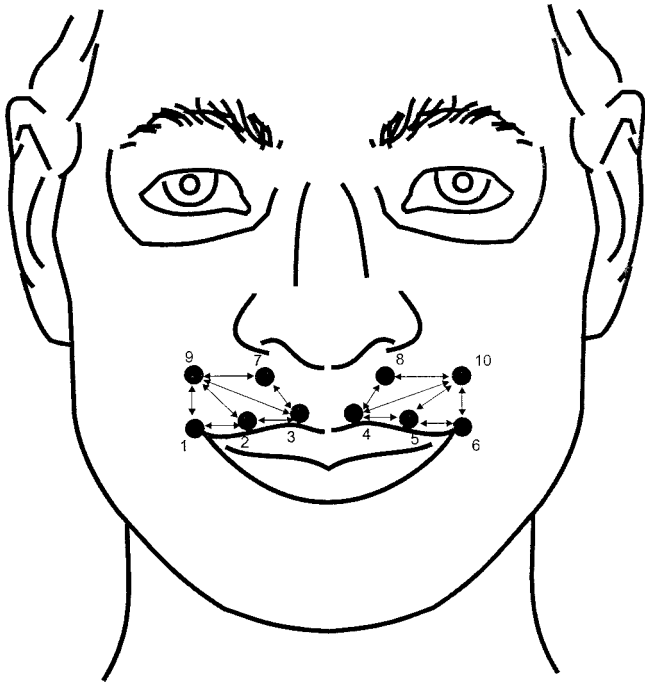
FIGURE 1 Facial landmarks and interlandmark distances. 1,6, right and left commissure points located at the mouth corners; 3,4, right and left upper lip points located on the peak of Cupid's bow; 2,5, right and left mid/upper lip points located midway between points 1 and 3 and 4 and 6, respectively; 7,8, right and left infranasal points located below the lateral alar rims; and 9,10, right and left supracommissure points located 1 cm above points 1 and 6 and 1 cm lateral to points 7 and 8, respectively. The interlandmark distances are represented by the double-ended arrows.

with CLP (Fig. 1). Each patient then was positioned within the calibrated measurement field and instructed to ''bite on his/her back teeth'' and ''smile as much as possible and then relax.'' This instructed smile from rest was a maximum smile. Recent work in our laboratory demonstrated that the instructed maximum smile is the most consistent animation when repeated

(Weeden et al., 2001) and is very similar in characteristics to the natural smile (Méndez, 1999). Additionally, impairments in movement appear to be most evident during smiling (Trotman et al., 2000). Before data collection, the smile animation was practiced with each patient, and five smiles were recorded at the same sitting. The data obtained for each patient were stored for later digitization and processing.

Data Processing

For each landmark pair, the relative change from rest over time y(t) then was calculated as follows (see Trotman et al., 1998a for a discussion of related measures):

$$y(t) = \frac{\text{interlandmark distance at time t}}{\text{interlandmark distance at rest.}}$$

Figure 2 displays this change for one landmark pair (e.g., 1-11) on the upper lip. Rest position is indicated by the ''0'' line on the y-axis, and time is on the x-axis. In this instance, the relative distance between the landmark pairs decreased during the smile so that the relative change in distance is in a negative direction as opposed to a positive direction if this distance had increased.

The y(t) then was averaged over the landmark pairs in the following upper lip regions: the entire upper lip, cleft or affected side of the lip, and noncleft or unaffected side of the lip. Because for any particular landmark pair, this measurement was made relative to the at-rest distance, this measurement was dimensionless. As such, y(t) has the advantage of factoring out differences because of facial size and is insensitive both to irrelevant whole-head motion and minor variations in marker placement. Subsequently, the following four variables were calculated.
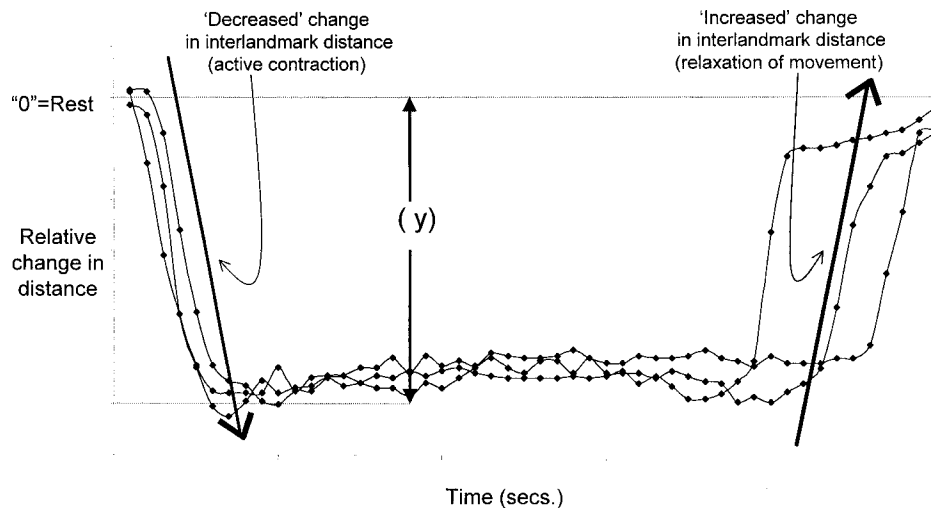

FIGURE 2 Relative change in distance from rest to maximum displacement for one landmark pair. The relative distance between the landmarks is decreasing so the change is below the rest or ''0'' line. The position is held by the patient for several frames at the maximum displacement and then increases to the ''rest'' line at the end of the movement when the patient relaxes.

Maximum Displacement From Rest (D)

For each of the five smile animations for each patient, the maximum value (D) of the averaged absolute values of the y(t)s over time was calculated separately for the entire lip, cleft or affected side of the lip, and noncleft or unaffected side of the lip.

Asymmetry (A)

For symmetrical landmark pairs on the upper lip, one on the cleft side and one on the noncleft side, the cleft versus noncleft absolute difference in the corresponding y(t)s was calculated. These differences then were averaged over the landmark pairs, and the maximum difference over time was calculated. Because (A) is based on movement relative to the rest position, it is a measure of asymmetry in motion, not asymmetry at rest, which is effectively subtracted out. If the movement were symmetrical, (A) would be zero.

Time (T)

For the averaged absolute y(t), the time (T) at which the maximum displacement occurred was determined. Thus, this value was a measure of how long the patient took to reach maximum displacement. As for (D), (T) was calculated for each of the five smile animations for each patient and regions that included the entire upper lip, cleft or affected side of the lip, and noncleft or unaffected side of the lip.

Velocity (V)

Velocity (V) was calculated as D/T for each of the five smile animations for each patient and in the entire upper lip, cleft or affected side of the lip, and noncleft or unaffected side of the lip.

STATISTICS

Consistency

A desirable attribute of a variable is consistency. In this study, the consistency of the objective measures was compared using four different descriptive approaches. First, the intrinsic measurement variability (i.e., the variability related to fluctuations in the patient's production of the smile and the motion analysis) was assessed by calculating the coefficient of variation (CV) for each patient and each measure from the five smile animations. The CV ($100\% \times$ [SD/mean]) is a simple measurement to compare the variability of measures independent of the magnitude of their means. The larger the CV, the more imprecise the measure was. As a within-patient descriptive statistic, the CV provided a comparison of patients as well as measures. Second, the intraclass correlation (ICC) statistic was calculated using the five smile replicates obtained from each patient as a global reliability or consistency estimate.

Third, patient variability (i.e., the intrinsic biological variability) was assessed by the descriptive statistics (SDs) of the aggregated patient value (the average of the five replicates) for each measure. Fourth, the variability of the measures was compared by the CV calculated from the descriptive statistics of the aggregated patient values.

Criterion-Related Validity

How well a variable actually represents the phenomenon of interest is difficult to quantify. In this study, agreement between the objective measures and subjective assessment of the 13 patients, made by the lay and professional panels, was evaluated as an indication of the criterion-related validity (convergence validity) of the objective measures. No attempt was made to assess predictive validity, and the face validity of the objective measures was assumed. Convergence validity was quantified by calculating Spearman correlation coefficients for each member of the lay and professional panels for each objective measure. These calculations were as follows: each judge's subjective rating of the disfigurement of the upper lip at rest was correlated separately with the patient's aggregated value for each of the objective measures, and each judge's subjective rating of the impairment of the upper lip during smiling was correlated separately with the patient's aggregated value for each of the objective measures. Thus, these calculations yielded eight coefficients per rater.

Descriptive statistics for the correlation coefficients were calculated separately for the lay and professional group ratings. A Wilcoxon signed rank test was used to assess whether the average individual correlation coefficient was significantly ($p < .001$) different from zero. Because the objective measures quantified "dynamic" activity of the upper lip, our expectation was that an individual rater's correlation between the smile ratings and objective measures would be stronger than that for the scar ratings. To quantify the difference between the "during smile" and "at rest" correlations with a given objective measure for each rater, a difference score was calculated for each rater ($smile_{correlation} - scar_{correlation}$), and descriptive statistics were generated separately for the two panels.

The panel-wide relationship between the subjective and objective measures was explored in two ways. First, the mean subjective "collective" rating was calculated for each patient separately by panel. The mean patient scar and smile "collective" ratings then were correlated separately with each of the objective measures. Second, a linear regression analysis was conducted, with the patient's mean subjective rating as the dependent variable, and displacement, time, and symmetry for the upper lip as the explanatory variables.

RESULTS

Consistency

The consistency of maximum displacement was better overall than asymmetry, time, or velocity (Tables 1 and 2). Within-

TABLE 1   Coefficient of Variation (Medians and Quartiles) for the Cleft Upper Lip (Total), Affected Side Only (Aff.), Unaffected Side (Unaff.), and Difference Between Affected and Unaffected Sides (Diff.) for the D, ASS, T, and V Measures

	Displacement (D)			Asymmetry (ASS)			Time (T) sec			Velocity (V) sec^{-1}		
	25%	Median	75%	25%	Median	75%	25%	Median	75%	25%	Median	75%
Total	3.56	4.90	11.43	9.94	13.77	28.32	16.54	28.65	38.09	14.20	24.95	39.37
Aff.	4.24	7.76	10.86	—	—	—	27.18	34.65	51.25	23.24	37.95	42.13
Unaff.	4.06	9.28	12.43	—	—	—	17.67	29.51	47.07	15.77	24.12	42.64
Diff.	11.53	17.68	49.41	—	—	—	−253.23	−130.86	69.16	25.79	60.16	116.20

patient maximum displacement quartile values for the CV for all regions of the face were substantially smaller than for the other measures, indicating less measurement error because of the production of the animations and motion analysis (Table 1). Not only was the within-patient consistency better overall for displacement, but also the smaller interquartile range for the within-patient CV values for all regions of the face indicated a greater level of consistency across all patients. The entire upper lip region tended to have better consistency than the other regions of the face for all measures (Tables 1 and 2).

The ICC values (Table 2) represent the within-patient consistency for the measures. In general, the displacement measurement demonstrated the greatest consistency. The ICC for the total upper lip displacement was 0.67. All of the other regions had slightly lower ICC values, although all were within the acceptable range. Velocity, but not time, was reasonably reliable for the entire upper lip as well as the affected and unaffected sides.

In terms of between-patient variability, displacement for the entire upper lip was the most consistent of all the measure and region combinations. Although the distribution of responses for maximum displacement appears limited, compared with time and velocity (Table 2), the distribution is likely to be sensitive enough to detect moderate differences between groups or changes over time because of the very small SD. For example, using the descriptive statistics as estimates (Table 2), 40 patients would be needed to detect a 20% difference in displacement between two independent groups (t test, $\alpha = .05$, $\beta = .20$), but 68 would be needed if sample size were based on time or velocity.

Criterion-Related Validity

Tables 3 through 6 show the associations (Spearman rank correlation coefficients) between the subjective individual and collective ratings of attractiveness and impairment given the 13 patients by lay and professional panel members and each of the objective measures for each of the four regions of the face. The general trend was that, despite considerable individual variation in the associations, the greater the perception of disfigurement at rest and the greater the perception of impairment during smiling, the less the movement (displacement, asymmetry, velocity) that occurred during the animation. The small p values indicate that, on average, these within-rater correlations are statistically different from zero. These association patterns were consistent for all "at rest" and "during smile" ratings except for the association with time. The "at rest" rating correlations with time to maximum displacement were negative, on average, in both groups, indicating that patients who were rated as more disfigured at rest took less time to produce the smile; however, "during smile" correlations with time to maximum displacement were reversed: the longer the animation took, the greater the perception of impairment during movement. For the displacement and asymmetry measures, the at-rest correlations were stronger than the during-smile correlations, but for velocity the during-smile correlations were stronger. The discrepancy between the at-rest and during-smile correlations with time was quite large because of the change in sign of the correlation values.

The associations between the objective measures and the mean subjective collective rating given each patient are also given in Tables 3 through 6. As expected, the "collective" correlation values are slightly stronger than the average of the "individual" rater's correlations, but the pattern is consistent: correlations were strongest for the entire upper lip; and the absolute value of the time, velocity, and asymmetry correlations with "during smile" ratings were slightly larger than displacement in both panels, but displacement was slightly larger with "at rest" ratings. The regression model predicting the patient's average rating during the smile movement from

TABLE 2   Descriptive Statistics (Mean, SD, Intraclass Correlation [ICC], Coefficient of Variation [CV]) for the Aggregate Patient Level Raw Objective Measures of the Cleft Upper Lip (Total), Affected Side Only (Aff.), Unaffected Side (Unaff.), and Difference Between Affected and Unaffected Sides (Diff.) for the D, ASS, T, and V Measures

	Displacement (D), mm				Asymmetry (ASS)				Time (T), sec				Velocity (V) sec^{-1}			
	Mean	SD	ICC	CV (%)	Mean	SD	ICC	CV (%)	Mean	SD	ICC	CV (%)	Mean	SD	ICC	CV (%)
Total	0.17	0.05	0.67	29.41	0.11	0.07	0.53	63.64	0.45	0.16	0.28	35.56	0.45	0.17	0.50	37.78
Aff.	0.16	0.05	0.43	31.25	—	—	—	—	0.49	0.13	0.09	26.53	0.39	0.16	0.43	41.03
Unaff.	0.19	0.06	0.86	31.58	—	—	—	—	0.47	0.20	0.29	42.55	0.52	0.21	0.46	40.38
Diff.	0.03	0.05	0.56	166.67	—	—	—	—	0.02	0.16	0.24	800.00	0.13	0.16	0.61	123.08

TABLE 3 Professional Group Results with Scar at Rest†

| | Spearman Rank Correlation Coefficient | | | |
	25% Q‡	Median Individual Correlation	75% Q	Collective Correlation
Upper lip (total)				
Displacement	−0.46	−0.33**	−0.25	−0.42
Time	−0.31	−0.21**	−0.14	−0.28
Velocity	−0.25	−0.09	0.02	−0.11
Symmetry	−0.44	−0.34**	−0.23	−0.37
Cleft side (affected)				
Displacement	−0.44	−0.31**	−0.13	−0.31
Time	−0.32	−0.24**	−0.02	−0.22
Velocity	−0.39	−0.30**	−0.08	−0.29
Difference (affected − unaffected)				
Displacement	−0.04	0.05	0.14	0.05
Time	−0.45	−0.35**	−0.28	−0.46
Velocity	0.04	0.08*	0.17	0.10

† Spearman rank correlation coefficients for the association between the individual and collective subjective ratings of the appearance of the upper lip with the scar at rest and the objective measures. Results are reported for the upper lip (total), affected side of the cleft lip (cleft side), and the difference between the affected and unaffected sides of the cleft lip (difference).
‡ Q = quartile.
* p ≤ .001.
** p ≤ .0001.

the objective measures of displacement, time, and symmetry was not statistically significant ($p = .15$), but that for the patient's average rating at rest was significant ($p = .01$). Displacement ($p = .02$) and time ($p = .02$) were statistically significant predictors, but asymmetry was not ($p = .54$). The R square for the model was 0.39.

DISCUSSION

In this study, we generated several different objective measures of lip movement (specific aim 1) and examined the re-

TABLE 4 Professional Group Results with the Scar During Movement†

| | Spearman Rank Correlation Coefficient | | | |
	25% Q‡	Median Individual Correlation	75% Q	Collective Correlation
Upper lip (total)				
Displacement	−0.47	−0.33**	−0.11	−0.38
Time	0.25	0.38**	0.55	0.52
Velocity	−0.56	−0.46**	−0.29	−0.53
Symmetry	−0.56	−0.39**	−0.24	−0.51
Cleft side (affected)				
Displacement	−0.44	−0.30*	−0.07	−0.32
Time	0.06	0.22*	0.35	0.32
Velocity	−0.39	−0.27**	−0.10	−0.34
Difference (affected − unaffected)				
Displacement	−0.34	−0.23*	−0.10	−0.21
Time	−0.12	0.01	0.14	0.00
Velocity	−0.32	−0.18*	−0.13	−0.20

† Spearman rank correlation coefficient for the association between the subjective individual and collective ratings of the appearance of the upper lip with the scar during movement and the objective measures. Results are reported for the upper lip (total), affected side of the cleft lip (cleft side), and the difference between the affected and unaffected sides of the cleft lip (difference).
‡ Q = quartile.
* p ≤ .001.
** p ≤ .0001.

TABLE 5 Lay Group Results with Scar at Rest†

| | Spearman Rank Correlation Coefficient | | | |
	25% Q‡	Median Individual Correlation	75% Q	Collective Correlation
Upper lip (total)				
Displacement	−0.53	−0.48**	−0.40	−0.51
Time	−0.30	−0.25**	−0.14	−0.23
Velocity	−0.24	−0.14*	−0.08	−0.20
Symmetry	−0.51	−0.38**	−0.35	−0.42
Cleft side (affected)				
Displacement	−0.41	−0.36**	−0.19	−0.40
Time	−0.33	−0.15**	−0.07	−0.12
Velocity	−0.31	−0.23**	−0.11	−0.35
Difference (affected − unaffected)				
Displacement	−0.18	−0.05	0.04	−0.01
Time	−0.46	−0.32**	−0.24	−0.46
Velocity	−0.12	0.01	0.15	0.09

† Spearman rank correlation coefficient for the association between the subjective 'individual' and collective ratings of the appearance of the upper lip with the scar at rest and the objective measures. Results are reported for the upper lip (total), affected side of the cleft lip (cleft side), and the difference between the affected and unaffected sides of the cleft lip (difference).
‡ Q = quartile.
* p ≤ .001.
** p ≤ .0001.

lationship between the subjective evaluations of both lip form and lip movement, respectively, and the objective measures that were generated (specific aim 2). Because this was an exploratory study, we sought to determine which objective measure was best suited for the analysis of lip function. A good measurement is one that is internally and externally consistent and valid. The individual CV values for the different measures represented the small inherent noise because of both the within-patient differences as well as the measurement variability during the production of the animations. The ICCs provided information on the consistency or reliability of the measures.

TABLE 6 Lay Group Results with Scar in Smile Movement†

| | Spearman Rank Correlation Coefficient | | | |
	25% Q‡	Median Individual Correlation	75% Q	Collective Correlation
Upper lip (total)				
Displacement	−0.38	−0.29*	−0.06	−0.30
Time	0.22	0.34**	0.48	0.45
Velocity	−0.48	−0.39**	−0.26	−0.43
Symmetry	−0.47	−0.41**	−0.35	−0.51
Cleft side (affected)				
Displacement	−0.41	−0.28*	−0.12	−0.28
Time	0.02	0.24*	0.34	0.26
Velocity	−0.39	−0.28*	−0.07	−0.25
Difference (affected − unaffected)				
Displacement	−0.20	−0.11	0.03	−0.17
Time	−0.20	−0.07	0.09	−0.01
Velocity	−0.26	−0.09	0.08	−0.15

† Spearman rank correlation coefficient for the association between the subjective individual and collective ratings of the appearance of the upper lip with the scar during movement and the objective measures. Results are reported for the upper lip (total), affected side of the cleft lip (cleft side), and difference between the affected and unaffected sides of the cleft lip (difference).
‡ Q = quartile.
* p ≤ .001.
** p ≤ .0001.

The aggregate CV values provided information on which measures would be most suitable to better discriminate between groups of patients, and the correlation between the objective measures and subjective evaluations provided information on the convergence validity of the different objective measures. Based on these criteria of consistency and validity, on balance, displacement was the best objective measurement of lip movement.

As opposed to objective measures generated from the cleft and the noncleft sides of the upper lip, the objective measures generated from the entire upper lip region provided the most overall reliability with the best within- and among-patient consistency. Furthermore, for both the lay and professional raters, the objective measures from the entire upper lip region accounted for an overall greater percentage (25%) of the variation in the subjective evaluations than objective measures from the cleft and noncleft sides of the upper lip. One caveat in this regard, was the measure of the CV for the total upper lip, which, as expected, was smaller than the values for the other lip regions because it was based on the average of a larger number of interlandmark distances. It is impossible to say how large this effect might be, but it does undermine slightly the favorable CV values for the entire upper lip. Several investigators have used the noncleft side of the upper lip in patients with unilateral cleft lip (and palate) as a control or comparison for the cleft side. For the analysis of movement, however, this approach is unwise for two reasons. First, the differences in measurements between the cleft or affected side of the lip and the noncleft or unaffected side are small. Second, the impairment in movement because of surgical repair of a unilateral cleft lip would likely affect movement of the noncleft side to varying degrees that would depend on a number of factors such as the extent of the cleft, skill of the surgeon, and effects of scarring.

Our hypothesis that objective measures of lip function are correlated with examiners' subjective evaluation of lip form both at rest and during movement did hold. There was a variable but definite relationship between the objective measures and the subjective evaluations. This relationship was stronger between the "at rest" ratings and the objective measures versus the "during smile" ratings and the objective measures. It appears that for the patients with cleft in this study, the appearance of the scar on the static upper lip provided greater information to the examiners on the expected quality of lip movement that was measured objectively. This unexpected finding implies that subjective evaluations of impairment (animation ability) should be made with the patient at rest. One explanation for this result may be that the examiners were affected by a perceived or real impairment during the smile animation that served to confound the association of the smile ratings with the objective measures. Our previous work supports this contention; we demonstrated that professionals agree more so when evaluating faces at rest than during movement and that the appearance of scarring on the upper lip influenced or confounded examiners' subjective ratings of movement (Ritter et al., 2002). These findings tend to weaken the value of subjective evaluations during movement as a tool to assess the degree of lip impairment.

In general, for both the "at rest" and "dynamic" subjective assessments, as the examiners ratings got worse, the values for the objective measures decreased. Thus, the cue when viewing and rating the upper lip appearance was a decrease in lip activity. This finding may be intuitive for the subjective "dynamic" evaluations during which the examiners actually viewed the appearance of lip movements, but it is less so for the subjective evaluations of lip appearance at rest. Specifically, different objective measures were involved in each instance. The "at rest" ratings were more strongly associated with the displacement and asymmetry measures, but the movement (during smile) ratings were more strongly associated with velocity. It is interesting that velocity is a composite measure of displacement and time, and this finding may imply that the development of global objective measures combining the measurement of different attributes of movement may ultimately improve the association between subjective ratings during movement and the objective measures.

One further finding in this regard was the difference in the association between the objective measurement of "time to displacement" and the subjective "at rest" and "during smile" ratings, respectively. It appears that the examiners' cue when viewing and rating the upper lip appearance at rest was a decrease in the length of time to make the movement. Because the examiners were not actually viewing the smiling face, this can only be an expectation on the part of the examiners as to what would occur during movement (i.e., an expected cue). When rating the lip during movement, however, the examiners' cue was an increase in the length of time taken to make the movement. Intuitively, the latter finding seems plausible and may represent an actual cue because it is highly possible that increased scarring of tissue may not only limit the tissue movement but also lengthen the time it takes to make the movement.

The finding of little difference in the association between the objective measures and the examiners' "individual" versus "collective" ratings is interesting. In a perfect world, decisions based on the average subjective ratings of multiple, well-trained, well-calibrated raters may serve as the single best estimate of the patient's disfigurement or impairment. In the framework of this study, the "collective" correlation tells us that displacement was the best measure. In the real world, however, subjective clinical decisions are most likely to be made by the patient's surgeon, possibly in conjunction with the child, parents, or both. It is of interest, therefore, to determine the variability across individuals in their individual associations between the perception of disfigurement and the objective measure or the perception of impairment and the objective measure and to determine whether that association is different from those who have been professionally trained and those who have not. The descriptive statistics for the "individual" correlations allows us to look at the dispersion of the "individual" correlation between subjective ratings and objective measures. Once again, displacement was the best objective measure. For example, when rating the faces at rest, 50% of

the individual correlation coefficients with displacement that were calculated for the professional raters ranged from -0.46 to -0.25, and 25% were between -1.00 and -0.46; however, for velocity some of the correlation coefficients were positive and some negative, indicating a lack of consistency in the relationship between the subjective and objective measures.

At this juncture, a discussion of the ideal characteristics of objective measures of circumoral movement that would serve our purpose of determining impairment in lip movement seems appropriate. In general, it would be desirable for a measure to be precise, consistent, and externally valid. More specifically, the measure should be sensitive enough to detect differences or changes, have an adequate distribution of responses, be appropriate to study our objectives, enhance the objectivity of the outcome by reducing involvement of the observer, and be efficient in terms of affordability and practicality. In our approach to measure the characteristic of interest, which is impairment, we used several different variables; because impairment is somewhat abstract, incorporating both subjective and objective evaluations should improve the validity of out assessments.

In summary, this exploratory analysis assessed the utility of objective measures of lip function in patients with unilateral cleft lip and palate and the findings were as follows.

- For accuracy, any objective analysis should include the entire upper lip region.
- Measurements of displacement were the most consistent and valid when compared with measures of asymmetry, time, and velocity.
- In general, there was a decrease in the objective measures of upper lip movement as examiners' perceptions of facial appearance (disfigurement) at rest and impairment during movement became worse.
- This relationship between the objective measures of upper lip movement and the examiners' perceptions of facial disfigurement at rest was greater than that between the objective measures of upper lip movement and examiners' perceptions of impairment during movement, implying that subjective assessments should be made with the face at rest.
- There was little difference in the association between the objective measures of upper lip movement and the examiners' ''individual'' versus ''collective'' subjective ratings.

Ultimately, the aim is to use objective measures of lip function for assessment of the outcomes of primary and secondary (revision) lip surgeries. One interesting thought by a reviewer of a previous manuscript (Ritter et al., 2002) is that objective measures may themselves lead to surgery that is unwarranted. We contend that objective measures and subjective evaluations provide a common component as well as separate components and both are necessary for a complete picture.

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;38:385–390.

Broadbent BH Sr., Broadbent BH Jr., Golden WH. *Bolton Standards of Dentofacial Developmental Growth.* St. Louis: CV Mosby; 1975.

Coghlan BA, Matthews B, Pigott RW. A computer-based method of measuring facial asymmetry and facial attractiveness in patients with clefts of the lip and palate. *Br J Plast Surg.* 1987;40:371–376.

Hurwitz DJ, Ashby ER, Llull R, Pasqual J, Tabor C, Garrison L, Gillen J, Weyant R. Computer-assisted anthropometry for outcome assessment of cleft lip. *Plast Reconstr Surg.* 1999;103:1608–1623.

Marsh JL. When is enough enough? Secondary surgery for cleft lip and palate patients. *Clin Plast Surg.* 1990;17:37–47.

Méndez MC. *A Three-Dimensional Analysis of Facial Movements in Normal Adults: Methodological Validation and Characterization of Natural Expressions* Ann Arbor, MI: University of Michigan; 1999. Thesis.

Morrant DG, Shaw WC. Use of standardized video recordings to assess cleft surgery outcome. *Cleft Palate Craniofac J.* 1996;33:134–142.

Riolo ML, Moyers RE, McNamara JA Jr., Hunter WS. *An Atlas of Craniofacial Growth: Cephalometric Standards From the University School Growth Study, The University of Michigan.* Monograph 2. Ann Arbor, MI: Center for Human Growth and Development, University of Michigan; 1974.

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.

Roberts-Harry DP, Evans R, Hawthorn I. Effects of different surgical regimes on nasal asymmetry and facial attractiveness in patients with clefts of the lip and palate. *Cleft Palate Craniofac J.* 1991;28:274–278.

Trotman C-A, Faraway JJ. Sensitivity of a method for the analysis of facial mobility. II. Interlandmark separation. *Cleft Palate Craniofac J.* 1998a;35: 142–153.

Trotman C-A, Faraway J, Essick G. Three-dimensional nasolabial displacement during movement in repaired cleft lip and palate patients. *Plast Reconstr Surg.* 2000;105:1273–1283.

Trotman C-A, Faraway JJ, Silvester T, Greenlee GM, Johnston LE Jr. Sensitivity of a method for the analysis of facial mobility. I. Vector of displacement. *Cleft Palate Craniofac J.* 1998b;35:132–141.

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.