OPEN

PRS GLOBAL OPEN

ORIGINAL ARTICLE
Pediatric/Craniofacial

Validation of a Unilateral Cleft Lip Surgical Outcomes Evaluation Scale for Surgeons and Laypersons

Alex Campbell, MD*†
Carolina Restrepo, MD*†
Gaurav Deshpande, MDS*†‡
Caroline Tredway, MD§
Sarah M. Bernstein, MD, MHA§
Rachel Patzer, PhD, MPH¶‖
Lisa Wendby*
Bjorn Schonmeyr, MD, PhD***

Background: A standardized evaluation tool is needed for the assessment of surgical outcomes in cleft lip surgery. Current scales for evaluating unilateral cleft lip/nose (UCL/N) aesthetic outcomes are limited in their reliability, ease of use, and application. The Unilateral Cleft Lip Surgical Outcomes Evaluation (UCL SOE) scale measures symmetry of 4 components and sums these for a total score. The purpose of this study was to validate the SOE as a reliable tool for use by both surgeons and laypersons.

Methods: Twenty participants (9 surgeons and 12 laypeople) used the SOE to evaluate 25 sets of randomly selected presurgical and postsurgical standardized photographs of UCL/N patients. Interrater reliability for surgeon and laypeople was determined using an intraclass correlation coefficient (ICC).

Results: Individual surgeons and laypeople both reached an ICC in the "fair to good" range (ICC = 0.42 and 0.59, respectively). Averaging 2 evaluators in the surgeon group improved the ICC to 0.58 and in the laypeople group to 0.74, respectively. Averaging 3 evaluators increased the ICC for surgeons to the "good" range (ICC = 0.71) and the ICC for laypeople to the "very good" range (ICC = 0.82).

Conclusions: Surgeon and layperson raters can reliably use the SOE to assess the aesthetics results after surgical repair of UCL/N, and improved reliability and reproducibility is achieved by averaging the scores of multiple reviewers.
(Plast Reconstr Surg Glob Open 2017;5:e1472; doi: 10.1097/GOX.0000000000001472; Published online 13 September 2017.)

INTRODUCTION

The goal of multidisciplinary cleft care is complete rehabilitation and normalization of a child. The objective of cleft lip repair is to minimize all stigmata of the cleft-related deformity and to free the patient from the psychological implications of impaired facial appearance. Many techniques and protocols exist in cleft surgery, and there is a central need for institutions, hospitals, and organizations involved in cleft care to systematically measure and compare outcomes to guide best practices.

A central element of the overall outcome after cleft lip repair is the aesthetic appearance of the lip and nose. Previous studies assessing outcomes in cleft patients have focused on satisfaction with treatment, dental arch relationship, and speech, but evaluation of facial appearance has received limited attention.[1–7] The majority of studies have focused on assessing and differentiating among different surgical techniques or assessing the treatment outcome among different centers.[8–11]

Assessment of the appearance of the cleft-related deformity and the impact of surgical treatment is a critical

*From *Operation Smile, Virginia Beach, Va.; †Plastic Surgery, Penn State Hershey, Hershey, Pa.; ‡MGM Dental College and Hospital, Navi Mumbai, India; §School of Medicine, Emory University, Atlanta, Ga.; ¶Department of Epidemiology, Emory University School of Medicine, Atlanta, Ga.; ‖Department of Surgery, Emory University School of Medicine, Atlanta, Ga.; and **Plastic Surgery, Skane University Hospital, Malmo, Sweden.*

Received for publication June 21, 2017; accepted July 11, 2017.

Presented at the 10th European Craniofacial Congress, June, 2015, Gothenburg, Sweden and at the 13th International Congress for Cleft Lip and Cleft Palate and Related Craniofacial Anomalies, February, 2017, Chennai, India.

DOI: 10.1097/GOX.0000000000001472

Disclosure: *The authors have no financial interest to declare in relation to the content of this article. The Article Processing Charge was paid for by the authors.*

Supplemental digital content is available for this article. Clickable URL citations appear in the text.

component of the quality of life outcome for this patient pool.[12] Researchers have created and tested scales, but proposed methods vary widely in their study design, patient population, treatment stage, type of cleft, evaluators' education, and familiarity with the patients and techniques.[13–16] Limitations of reliable, valid, and meaningful ways of measuring aesthetic results inhibit our ability to determine the best strategies for treatment.[17] To make any accurate determination of the efficacy of the numerous treatment methods currently used, we need a uniform and simple assessment tool that provides objective evaluations of the results of treatment.[18]

Al-Omari et al.[19] undertook a review of grading systems of the appearance of cleft deformities, and the review concluded that, because existing study designs vary extensively, "an internationally agreed objective method of assessment … is required".

Subjective or qualitative assessments using standardized photographs analyzed by a panel of judges have been shown by Asher-McDade et al.[20] to provide valid, reliable, and reproducible ratings of cleft patients. This scale uses a standardized method to evaluate 4 nasolabial components (nasal form, nose symmetry, nasal profile, and vermilion border) and grades each component on a 5-point scale. The Asher-McDade scale introduced numerous advances in the evaluation of surgical outcomes. These include a validated scale based on subjective evaluation of standardized 2D images, summing scores of the individual components for a final evaluation score, and improving reliability by averaging scores of multiple reviewers. However, numerous weaknesses limit its relevance for measuring the aesthetic outcome after a primary cleft lip repair on a scale that is necessary to provide comprehensive outcomes evaluations in global cleft care. The scale is designed to evaluate patients in late childhood/early adolescence after years of multidisciplinary treatment, rather than focusing on the aesthetic results of a single surgical intervention. The scale looks at 3 separate components of outcomes on the nose and does not place significant emphasis on critical elements of the lip repair such as the philtrum, white roll, lateral lip, and balance of cupid's bow. For "very good" reliability, individual patients must each be scored by multiple calibrated expert raters, reviewers, which is inconvenient and impractical.

The Asher-McDade scale has been utilized in the Eurocleft, Americleft, Scandcleft, and the CSAG studies, with each patient was evaluated by multiple experts in cleft care and interrater reliability in the "moderate to good" range (Tables 1, 2). Very little was learned about aesthetic outcomes in general, and no insights were offered regarding merits and pitfalls of various surgical techniques.[21–24] Authors in the Americleft study summarized that "ideally, more standardized and objective assessment methods will be developed to improve the accuracy and reliability of evaluations of nasolabial aesthetic outcomes."

Responding to the need for improved outcomes measures, a group of surgeons at Operation Smile experienced in cleft care participated in development of the Unilateral Cleft Lip Surgical Outcomes Evaluation (UCL SOE) scale (Figs. 1, 2). The UCL SOE scores symmetry of 4 individual anthropomorphic components of the cleft repair (Cupid's bow, lateral lip, nose, and free vermillion). Each element is scored on a 3 point scale: 2 (excellent), 1 (mild asymmetry), 0 (unsatisfactory). The scores of the 4 individual scores are then summed for a total score of 0 (lowest) to 8 (highest). The purpose of this study was to validate the UCL SOE as a reliable tool for use by both surgeons and laypersons.

Table 1. Interpretation of Intraclass ICC

ICC	Strength of Agreement
< 0.20	Poor
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Good
0.81–1.00	Very good

MATERIALS AND METHODS

All study subjects were drawn from patients at the Guwahati Comprehensive Cleft Care Center in Assam, India, between 2011 and 2014 admitted for primary cheiloplasty. All patients' parents or guardians signed an informed consent allowing for the use of their medical records and photographs for research. All required forms and signatures were submitted to the institutional review board, and ethical approval was granted.

Patients ranged in age from 6 to 24 months, and exclusion criteria included prior surgeries, known congenital syndromes, or other craniofacial abnormalities. From this group, 25 sets of standardized frontal and basal photographs taken with a Nikon digital single-lens reflex camera were randomly selected and deidentified. Photographs were cropped to minimize the portion of the face or body not affected by the cleft. All photographs were formatted to be of uniform length and width.

Two sets of evaluators were recruited to participate in the study. The first set consisted of 9 experienced cleft surgeons. All surgeons hold current plastic surgery board certification,

Table 2. Summary of Various Studies Using Asher-McDade Methodology

Study	Reviewers	Profession	Patients	Ages (y)	ICC
Asher-Mcdade (1991)	5	Orthodontists	32	9–13	0.6 (single rater), 0.82 (3 raters), 0.90 (6 evaluations)
Eurocleft (1992)	6	Orthodontists	115	8–11	0.6
Americleft (2011)	5	4 Orthodontists, 1 plastic surgeon	124	5–13	0.59–0.71
Scandcleft (1992)	6	3 Orthodontists, 3 plastic surgeons	61	16	0.26–0.35
CSAG (1998)	3	1 Orthodontist, 1 plastic surgeon, 1 speech therapist	391	5–12	062–0.71

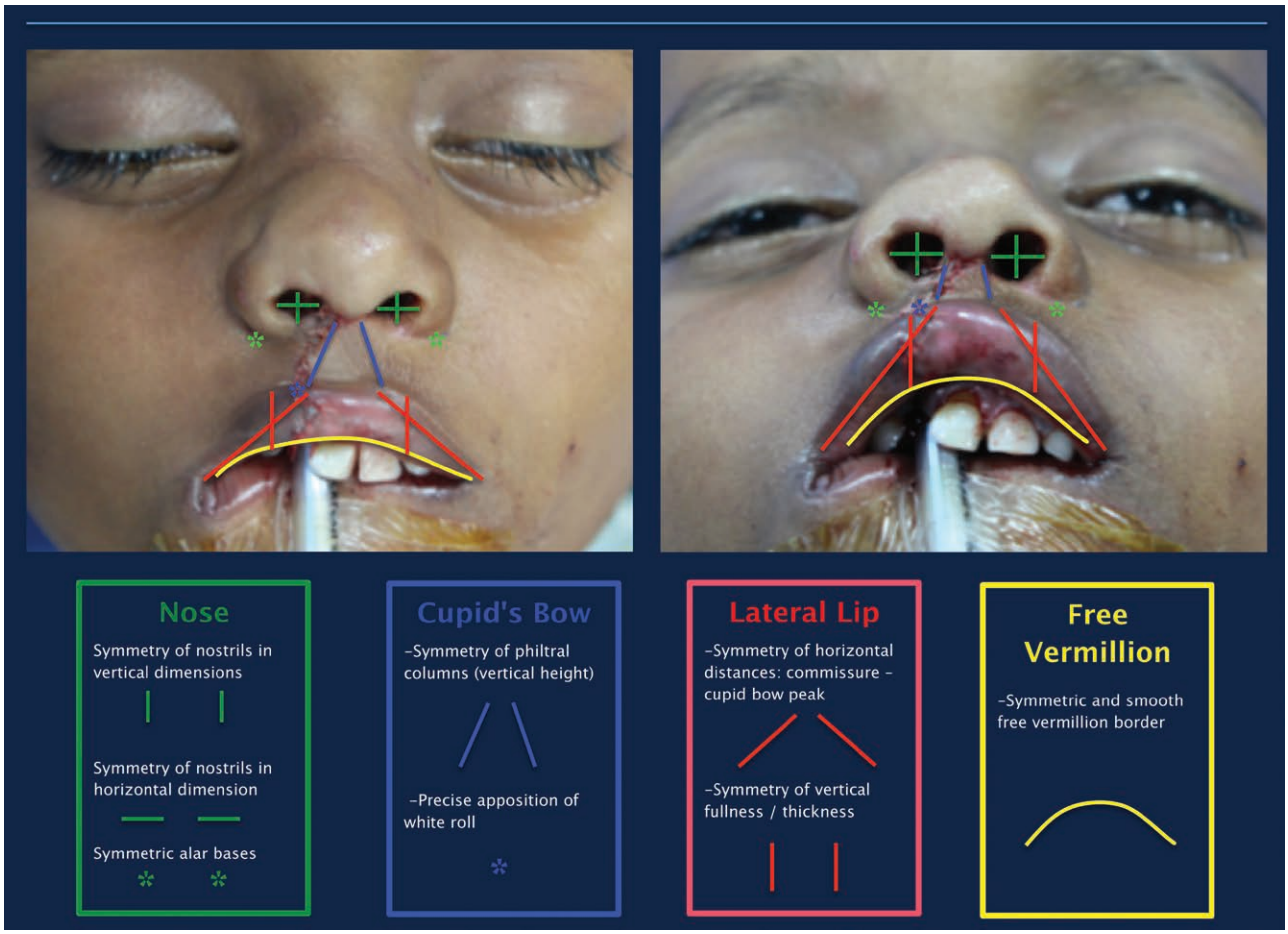

Fig. 1. The UCL SOE scores symmetry of 4 individual anthropomorphic components of the cleft repair (Cupid's bow, lateral lip, nose, and free vermillion).

international credentialing with Operation Smile, and have extensive experience in cleft lip and palate repair. The second set of 12 laypersons were recruited from nonsurgical staff at the Guwahati Comprehensive Cleft Care Center and medical students at Emory University. Demographic information of the evaluators can be found in Table 3.

Written and video instructions on how to use the UCL SOE were provided to all evaluators, and a 1-hour teaching session was provided to describe the scale and perform practice cases. Evaluators provided with frontal and basal photographs for each of the 25 cases. On-table, immediate postoperative photographs were utilized. Evaluators recorded their impressions or each of the 4 components of the scale in the provided table adjacent to the pictures. The evaluations were then e-mailed back to the researchers for analysis.

Interexaminer reliability of the scale was calculated using the intraclass correlation coefficient (ICC). The ICC was calculated for the ratings of individual surgeons, for the ratings of individual laypeople, and for the ratings of all individual evaluators combined. The ICC was also calculated for the ratings of averages of 2 and 3 randomly grouped surgeons and for the ratings of averages of 2 and 3 randomly grouped laymen. For the ICCs for averages of 2, random sets of 2 were averaged, and then the ICC of these sets of 2 was calculated; this was done 4 times to ensure there was little variation. The same method was used for calculating the ICC for averages of 3.

Means of the scores and their components were calculated for both groups of surgeons and laypeople. Differences in average scoring pattern were assessed using a paired t-test. Regressions were calculated for average surgeon versus average layman score (total and each of the 4 components) for each of the 25 pictures.

Distributions of the total scores and each of the 4 components were calculated for surgeons and laymen. To compare the distributions between the 2 groups, a correction was made to account for the differing numbers of surgeons and laymen completing the evaluations. The surgeons' distributions were multiplied by the number of laymen[13] divided by the number of surgeons.[10]

RESULTS

The ICCs for individual and averaged ratings in each group were calculated and are listed in Table 4.

Statisticians consider an ICC of greater than 0.40 to be a "moderate" correlation for health care research, ICC above 0.60 to be a "good" correlation, and above 0.80 to be a "very good" correlation.[25,26]

Fig. 2. Each element is scored on a 3-point scale: 2 (excellent), 1 (mild asymmetry), 0 (unsatisfactory). The scores of the 4 individual components are then summed for a total score of 0 (lowest) to 8 (highest).

Table 3. Demographic Information of the Surgeon and Layperson Evaluators

Group	Total Individuals	Males	Females
Surgeons	9	8	1
Laymen	12	6	6

Table 4. The Intraclass ICCs for Individual and Averaged Ratings in Each Group

Group	Total	Nose	Cupid's Bow	Lateral Lip	Free Vermillion
Individual					
Surgeons	0.42	0.42	0.39	0.34	0.22
Laymen	0.59	0.43	0.33	0.4	0.43
Combined	0.48	0.41	0.36	0.35	0.31
Average of 2					
Surgeons	0.58	0.57	0.57	0.5	0.37
Laymen	0.74	0.58	0.51	0.56	0.58
Average of 3					
Surgeons	0.71	0.67	0.54	0.61	0.44
Laymen	0.82	0.71	0.59	0.67	0.69

Individual surgeons and laymen both reached an ICC in the "moderate" range for the total score (ICC = 0.42 and 0.59, respectively). ICCs were lower among both surgeons and laymen for individual components. ICCs of all the evaluators combined did not decrease below the ICCs of either group, which suggests that the laymen and the surgeons exhibited similar patterns of scoring.

Averaging 2 evaluators in the surgeon group improved the ICC for the total score to 0.58 and in the laypeople group to 0.74, respectively. Averaging 2 evaluators in both the surgeon and the layman groups also produced ICCs in the "moderate to good" range for individual components of scoring (ICCs = 0.50–0.74).

Averaging 3 evaluators increased the ICC for the total score judged by surgeons to the "good" range (ICC = 0.71) and the ICC for laypeople to the "very good" range (ICC = 0.82). Averaging 3 evaluators in both the surgeon and the layman groups produced ICCs in the "moderate to very good" range in all individual components of scoring (ICCs = 0.44–0.82).

The plot of the average surgeon versus the average layman total score for each picture, along with the regression for this plot, is found in Figure 3. The R^2 for the regression is 0.86, showing a strong correlation between surgeon and layman scoring patterns. The X variable of 0.82 and the intercept of 1.48 suggest that as the outcome score increases the surgeons and the laymen show more agreement.

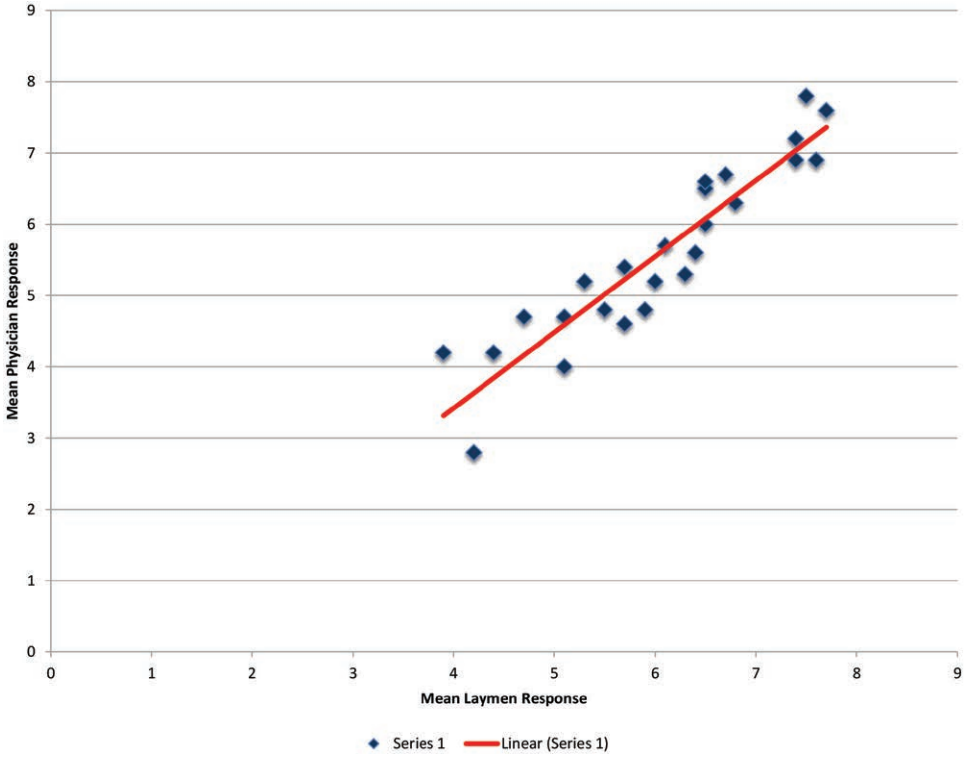

Fig. 3. The plot of the average surgeon versus the average layman total score for each picture, along with the regression. An R^2 of 0.86, shows a strong correlation between surgeon and layman scoring patterns.

Plots were also created of the average surgeon versus the average layman component scores, along with the regressions for those plots (**see figure, Supplemental Digital Content 1**, which displays plots and regressions of the average surgeon versus the average layman total score for Cupid's bow, lateral lip, nose, and free vermillion, ***http://links.lww.com/PRSGO/A525***). The R^2 for the Cupid's bow, lateral lip, and nose regressions show fairly strong correlation between surgeon and layman scoring patterns (R^2 = 0.80, 0.72, and 0.79, respectively), but a poorer correlation between surgeon and layman scoring patterns for the free vermillion (R^2 = 0.56). As with the total score, all regressions show an increase in surgeon and layman score agreement as the outcome score increases (X variables = 0.64–0.82; intercepts = 0.33–0.66). The distributions of the total scores for each group of evaluators (Fig. 4) and of each of the 4 components for each group of evaluators also demonstrate strong correlation between surgeon and layman (**see figure, Supplemental Digital Content 2**, which displays distributions of Cupid's bow, lateral lip, nose, and free vermillion for each group of evaluators, ***http://links.lww.com/PRSGO/A526***).

DISCUSSION

Oral clefts are among the most widely known and common craniofacial anomalies, yet there is currently no widely utilized assessment tool to evaluate the aesthetic result after primary UCL repair. Thus, there is relatively little evidence on aesthetic results after UCL repair to guide best techniques and protocols. Many techniques and protocols exist in cleft surgery, and there is a central need for institutions, hospitals, and organizations involved in cleft care to systematically measure and compare outcomes to guide best practices. The UCL SOE builds on the concepts and work of prior authors to quickly, efficiently, and consistently score the aesthetic result of a primary UCL repair.[27,28] The intent for the UCL SOE is to be an intuitive, easy to use, and reliable outcomes evaluation tool for use by experts and laypersons. In structuring the scale, researchers utilized numerous concepts and components that had been validated through prior studies.

Precise measures that accurately capture aesthetic results after surgical treatment for cleft lip are an immediate necessity. Without these, accurate comparison among surgeons, techniques, presurgical interventions, and treatment protocols cannot be made. Long-term assessments that include 3-dimensional data from computed tomographic scans, animated recordings from videos, and clinical examination by experts undoubtedly allow superior discrimination in assessing results. However, these are expensive, labor intensive, and require significant patient participation. Furthermore, the vast majority of cleft surgery takes place in developing regions, with charities treating cleft lip and cleft palate performing more than 150,000 operations per year.

Standardized photographs are currently the most commonly used medium for outcomes analysis worldwide. They are easy to use, widely available as part of the medical

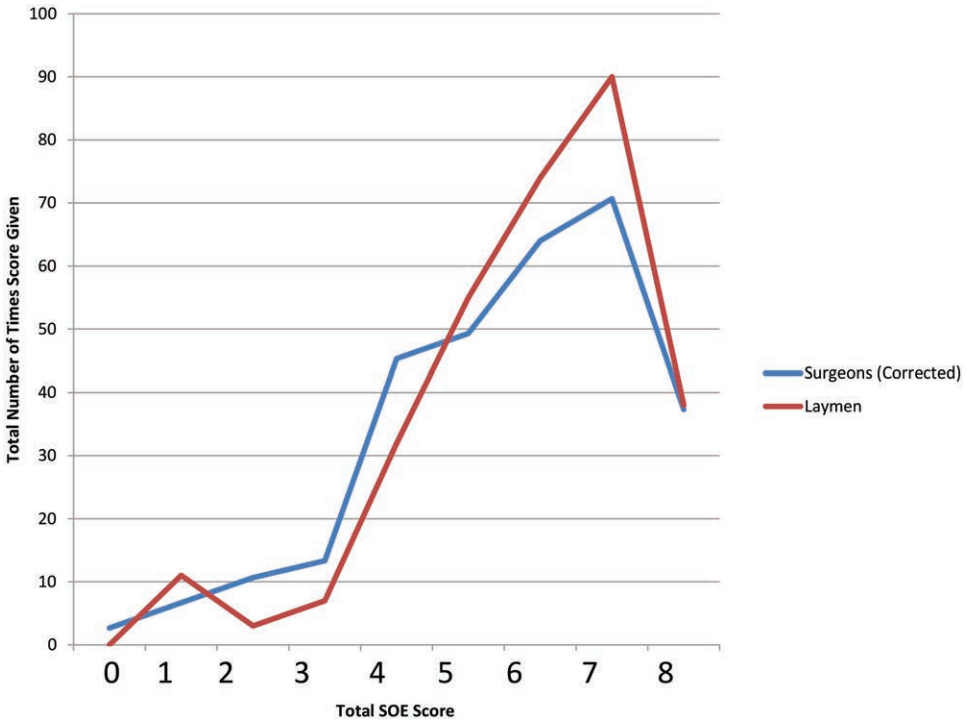

Fig. 4. Distributions of the total scores for each group of evaluators.

record, and reproducible. Numerous studies have successfully used 2-dimensional photographs for assessment of the cleft deformity.[20,29–31] Operation Smile utilizes specialized patient imaging technicians to capture and organize standardized photographs of all patients at each stage in intervention. Complete sets of patient photographs (frontal and basal) were available for all patients in this study, and views were consistent among patients being evaluated.

Most studies evaluating facial appearance in cleft patients have involved panels of health care professionals with experience in cleft care. Relatively few studies have used laypersons as evaluators and even fewer have compared evaluations of laypersons to the evaluations of surgeons. Those that have yielded inconsistent results, with some studies indicating laypeople and surgeons agree on grading and others suggesting varying degrees of disagreement between expert and nonexpert reviewers.[30,32,33] The potential to use laypersons as evaluators holds great appeal as surgeons and other medical professionals with expertise in cleft care are relatively limited in their numbers, time, and willingness to perform large numbers of evaluations. A recent study by Tse, et al.[34] demonstrated the power of using crowdsourcing to complete massive numbers of layperson assessments on an unprecedented scale in a convenient, rapid, and reliable means of assessing aesthetic outcome of treatment for UCL.

The UCL SOE overcomes many shortcomings of prior scales and is useful for large-scale outcomes evaluations. This study demonstrates that surgeon and layperson raters can reliably use the UCL SOE to assess the aesthetic result of UCL/N. According to these ICC calculations, individual raters can assess the total aesthetic result, with "moderate" reliability, and improved reliability and reproducibility is achieved by averaging the scores of multiple reviewers. Results also demonstrate that surgical expertise is not necessary to use the UCL SOE reliably and that "very good" reliability (ICC = 0.82) is achieved when the scores of 3 layperson reviewers are pooled and averaged.

Combined with the Unilateral Cleft Lip Cleft Severity Scale (UCL CSS), validated in a parallel study, the UCL SOE allows surgeons and laypersons to grade the preoperative severity of the UCL/N deformity and the final aesthetic result after primary surgical repair.[35] This has significant implications on the ability to conduct outcomes studies evaluating and comparing various surgeons, centers, techniques, and protocols. These tools have additional value to track patient results through time and also to monitor surgical development during training and practice. The ability to objectively measure UCL surgical outcomes will provide insight into the factors that contribute to differences in outcomes among patients. Future studies will be able to use the scale to investigate factors that may contribute to differences in surgical outcomes.

Broad implementation of these tools will gather data necessary to create a "bell curve" of expected aesthetic results after cleft lip repair, allowing significant advancements in quality assurance as well as identification of best practices that lead to superior results.

Alex Campbell, MD
Operation Smile
3641 Faculty Boulevard
Virginia Beach, VA 23453
E-mail: alexcampbellmd@gmail.com

PATIENT CONSENT

Parents or guardians provided written consent for the use of the patients' image.

REFERENCES

1. Eckstein DA, Wu RL, Akinbiyi T, et al. Measuring quality of life in cleft lip and palate patients: currently available patient-reported outcomes measures. *Plast Reconstr Surg.* 2011;128:518e–526e.
2. Richman LC. The effects of facial disfigurement on teachers perceptions of ability in cleft palate children. *Cleft Palate J.* 1978;15:155–160.
3. Mars M, Plint DA, Houston WJ, Bergland O, Semb G. The Goslon Yardstick: a new system of assessing dental arch relationships in children with unilateral clefts of the lip and palate. *Cleft Palate J.* 1987;24:314–322.
4. Atack NE, Hathorn IS, Semb G, et al. A new index for assessing surgical outcome in unilateral cleft lip and palate subjects aged five: reproducibility and validity. *Cleft Palate Craniofac J.* 1997;34:242–246.
5. Keuning KH, Wieneke GH, Dejonckere PH. The intrajudge reliability of the perceptual rating of cleft palate speech before and after pharyngeal flap surgery: the effect of judges and speech samples. *Cleft Palate Craniofac J.* 1999;36:328–333.
6. Morrant DG, Shaw WC. Use of standardized video recordings to assess cleft surgery outcome. *Cleft Palate Craniofac J.* 1996;33:134–142.
7. Tanikawa DY, Alonso N, Rocha DL. Evaluation of primary cleft nose repair: severity of the cleft versus final position of the nose. *J Craniofac Surg.* 2010;21:1519–1524.
8. Long RE, Jr, Hathaway R, Daskalogiannakis J, et al. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate part 1. Principles and study design. *Cleft Palate Craniofac J.* 2011;48:239–243.
9. Shaw WC, Asher-McDade C, Brattström V, et al. A six-center international study of treatment outcome in patients with clefts of the lip and palate: part 1. Principles and study design. *Cleft Palate Craniofac J.* 1992;29:393–397.
10. Rajanikanth BR, Rao KS, Sharma SM, et al. Assessment of deformities of the lip and nose in cleft lip alveolus and palate patients by a rating scale. *J Maxillofac Oral Surg.* 2012;11:38–46.
11. Trotman CA, Phillips C, Essick GK, et al. Functional outcomes of cleft lip surgery. Part I: study design and surgeon ratings of lip disability and need for lip revision. *Cleft Palate Craniofac J.* 2007;44:598–606.
12. Bennett ME, Phillips CL. Assessment of health-related quality of life for patients with severe skeletal disharmony: a review of the issues. *Int J Adult Orthodon Orthognath Surg.* 1999;14:65–75.
13. Johnson N, Sandy J. An aesthetic index for evaluation of cleft repair. *Eur J Orthod.* 2003;25:243–249.
14. Kim JB, Strike P, Cadier MC. A simple assessment method for auditing multi-centre unilateral cleft lip repairs. *J Plast Reconstr Aesthet Surg.* 2011;64:195–200.
15. He X, Shi B, Kamdar M, et al. Development of a method for rating nasal appearance after cleft lip repair. *J Plast Reconstr Aesthet Surg.* 2009;62:1437–1441.
16. Kim SC, Nam KC, Rah DK, et al. Assessment of the cleft nasal deformity using a regression equation. *Cleft Palate Craniofac J.* 2009;46:197–203.
17. Henry C, Samson T, Mackay D. Evidence-based medicine: the cleft lip nasal deformity. *Plast Reconstr Surg.* 2014;133:1276–1288.
18. Sitzman TJ, Allori AC, Thorburn G. Measuring outcomes in cleft lip and palate treatment. *Clin Plast Surg.* 2014;41:311–319.
19. Al-Omari I, Millett DT, Ayoub AF. Methods of assessment of cleft-related facial deformity: a review. *Cleft Palate Craniofac J.* 2005;42:145–156.
20. Asher-McDade C, Roberts C, Shaw WC, et al. Development of a method for rating nasolabial appearance in patients with clefts of the lip and palate. *Cleft Palate Craniofac J.* 1991;28:385–90; discussion 390.
21. Shaw WC, Asher-McDade C, Brattstro¨m V, 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;29:409–412.
22. Mercado A, Russell K, Hathaway R, 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;48:259–264.
23. 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;26:313–319.
24. Sandy J, Williams A, Mildinhall S, et al. The Clinical Standards Advisory Group (CSAG) cleft lip and palate study. *Br J Orthod.* 1998;25:21–30.
25. Streiner D, Norman G. Health measurement scales: a practical guide to their development and use. *Oxford University Press.* 1995:167–202.
26. Fleiss J. *The Design and Analysis of Clinical Experiments.* New York: John Wiley & Sons; 1986.
27. Bermudez L, Trost K, Ayala R. Investing in a surgical outcomes auditing system. *Plast Surg Int.* 2013;2013:671786.
28. Bermudez L, Carter V, Magee W, Jr, et al. Surgical outcomes auditing systems in humanitarian organizations. *World J Surg.* 2010;34:403–410.
29. Becker M, Svensson H, Jacobsson S. Clinical examination compared with morphometry of digital photographs for evaluation of repaired cleft lips. *Scand J Plast Reconstr Surg Hand Surg.* 1998;32:301–306.
30. Cussons PD, Murison MS, Fernandez AE, et al. A panel based assessment of early versus no nasal correction of the cleft lip nose. *Br J Plast Surg.* 1993;46:7–12.
31. Tobiasen JM, Hiebert JM, Boraz RA. Development of scales of severity of facial cleft impairment. *Cleft Palate Craniofac J.* 1991;28:419–424.
32. Roberts-Harry D, Stephens CD. Panel perception of facial attractiveness. *Br J Orthod.* 1991;18:152–153.
33. Eliason MJ, Hardin MA, Olin WH. Factors that influence ratings of facial appearance for children with cleft lip and palate. *Cleft Palate Craniofac J.* 1991;28:190–3; 193.
34. Tse RW, Oh E, Gruss JS, et al. Crowdsourcing as a novel method to evaluate aesthetic outcomes of treatment for unilateral cleft lip. *Plast Reconstr Surg.* 2016;138:864–874.
35. Campbell A, Restrepo C, Deshpande G, et al. Validation of a cleft severity scale for grading the severity of primary unilateral cleft lip and nose deformities by surgeons and laypersons. *Plast Reconstr Surg Glob Open.* [Epub ahead of print].