A New Index for Assessing Surgical Outcome in Unilateral Cleft Lip and Palate Subjects Aged Five: Reproducibility and Validity

Nikki E. Atack, B.D.S., F.D.S.R.C.S. (Engl)
Iain S. Hathorn, B.D.S., D.Orth., F.D.S.R.C.S. (Engl)
Gunvor Semb, D.D.S., Ph.D.
Tom Dowell, B.D.S., M.Sc., D.D.P.H.R.C.S.
Jonathan R. Sandy, M.Sc., Ph.D., M.Orth., F.D.S.R.C.S. (Engl), F.D.S.R.C.S. (Edin)

Objective: **This study assessed the reproducibility, reliability, and predictive validity of a previously developed index by the authors for assessing surgical outcome in unilateral cleft lip and palate (UCLP) children aged 5.**

Methods: **Sixty randomly selected study models of 5- to 6-year-old complete UCLP subjects were obtained and the index was used to assess their surgical outcomes.**

Results: **Assessment of these study models using the new index demonstrated excellent intra-examiner agreement. The inter-examiner agreement was shown to be good. The corresponding longitudinal models at 16 to 18 years of 54 of the initial 5- to 6-year-old sample were also acquired. These subjects had undergone orthodontic treatment but not orthognathic surgery. The need for osteotomy amongst these models was assessed. Between 13% and 18% (depending on examiner) of 5-year-olds' models were scored in the groups likely to require orthognathic surgery. In the corresponding 16- to 18-year-olds' models, 9% were assessed as likely to benefit from an osteotomy. However, on an individual basis, it was not possible to predict future growth from study models at age 5.**

Conclusions: **This study has provided a reliable and reproducible index for assessing the outcome of surgery in UCLP subjects earlier than indices already available. True validation of the index was not possible but it appears that it relies on face validity.**

KEY WORDS: *index, pediatric, surgical outcome, unilateral cleft lip and palate*

Low-incidence anomalies such as clefts of the lip and palate present major difficulties in monitoring the benefits or drawbacks of different treatment regimens. The methods available have advanced the care of these children to a point where any new techniques or developments are likely to bring about small changes. Consequently, more sophisticated research methodology is required to detect any improvements (Roberts et al., 1991). Natural variation means a large number of cases need to be assessed to detect clinically and statistically significant changes. One facet of care that has been examined is the effect of primary surgery on the maxilla and facial growth. It is frequently suggested that the full effects of surgery cannot be determined until late adolescence when the majority of facial growth has ceased. More recently, the development of a robust and reproducible index to assess dental arch relationships (Mars et al., 1987) has shown that the detection of differences in outcome as early as 10 years of age is possible (Mars et al., 1992; Shaw et al., 1992).

Clearly the ability to detect differences at an earlier stage would be advantageous, because surgeons would be able to relate any changes in technique or timing to outcome. Attempts to measure differences in younger age groups have suggested that the soft-tissue profile outline may be a useful indicator (MacKay et al., 1994). The Goslon Yardstick (Mars et al., 1987), although originally established to assess 10-year-old study models, has been used for longitudinal assessment of dental arch relationships in children with UCLP (Noverraz et al., 1993). It appears that the mean Goslon Yardstick score remains similar throughout all stages of occlusal development. We have recently developed an index for dental study models of 5-year-old children with unilateral cleft lip and palate (UCLP) (Atack et al., 1996). The purpose of this article is to report the reproducibility, reliability, and validity of this index.

Dr. Atack is Registrar; Dr. Hathorn is a consultant orthodontist; Dr. Dowell is a clinical lecturer in Statistics; and Dr. Sandy is a consultant senior lecturer, University of Bristol, Bristol, United Kingdom. Dr. Semb is a consultant senior lecturer, University of Manchester, Manchester, United Kingdom.

Supported by a Grant from the United Bristol Healthcare Trust Medical Research Committee number: P19307.

Submitted April 1996; Accepted December 1996.

Reprint requests: Dr. J.R. Sandy, Department of Child Dental Health, Bristol Dental School, Lower Maudlin Street, Bristol BS1 2LY U.K.

Methods

Selection of the Test Sample

Sixty randomly selected study models of 5- to 6-year-old subjects with repaired complete UCLP were obtained from the

TABLE 1 General Features of Study Models in the 5-Year-Olds' Index

Group	*General Features*	*Predicted Long-Term Outcome*
1	Positive overjet with average inclined or retroclined incisors No crossbites/openbites Good maxillary arch shape and palatal vault anatomy	Excellent
2	Positive overjet with average inclined or proclined incisors Unilateral crossbite/crossbite tendency ± Open bite tendency around cleft site	Good
3	Edge-to-edge bite with average inclined or proclined incisors; or reverse overjet with retroclined incisors Unilateral crossbite ± Open bite tendency around cleft site	Fair
4	Reverse overjet with average inclined or proclined incisors Unilateral crossbite, ± bilateral crossbite tendency ± Open bite tendency around cleft site	Poor
5	Reverse overjet with proclined incisors Bilateral crossbite Poor maxillary arch form and palatal vault anatomy	Very Poor

Oslo CLP Growth Archive (n = 54) and from a recent South-West regional audit record collection initiative (n = 6). All subjects were at a similar stage of dental development. The corresponding longitudinal study models at ages 16 to 18 years for 54 of the test sample subjects were obtained from the Oslo CLP Growth Archive. None of these subjects had undergone orthognathic surgery, although most had undergone orthodontic treatment and secondary alveolar bone grafting.

Testing the 5-Year-Olds' Index

An index had been developed previously that allowed the assessment of primary surgical outcome in UCLP subjects aged 5 (Atack et al., 1996). The general features of that study model index are described in Table 1. This new index was assessed for reproducibility, reliability, and predictive validity.

Assessing Reliability and Reproducibility

The dental-arch relationships of the test sample were assessed using the 5-year-olds' index. The assessment was performed twice in 1 day by four examiners: three orthodontic consultants, A, B and D, and one registrar, C. A suitable period of time was allowed between each scoring, as well as the random reallocation of the position of the study models for the second scoring, to minimize the possible influence of memory on the results. Random numbers were digitally generated and then used to reassign the study model positions.

TABLE 2 Relevance of Weighted κ Values to Levels of Agreement*

κ Value	*Strength of Agreement*
≥ 0.20	Poor
0.21 to 0.40	Fair
0.41 to 0.60	Moderate
0.61 to 0.80	Good
0.81 to 1.00	Very Good

*As described by Altman, 1991.

TABLE 3 Demographic Data of Test Sample

Feature	*Female*	*Male*
Number of 5-year-old subjects	23 (38%)	37 (62%)
Number of 16- to 18-year-old subjects	20 (37%)	34 (63%)
Age (yr)		
5-year-old subjects		
Mean ± SD	5.8 ± 0.7	5.8 ± 0.7
Range	5.0 to 7.5	4.7 to 8.0
16- to 18-year-old subjects		
Mean ± SD	17.5 ± 1.2	17.6 ± 1.3
Range	15.4 to 19.4	16.0 to 20.4

Assessing Validity

The study models of the 16- to 18-year-olds from the Oslo sample were assessed as to whether or not the subjects would benefit from an osteotomy. This assessment was performed by the same examiners who were involved in the reliability and reproducibility study. A unanimous consensus view was recorded as to whether surgery was thought to be required. These findings were then compared with the scores assigned to the equivalent study models at age 5.

Statistical Analysis

Intra- and inter-examiner agreement was determined using the weighted kappa (κ) statistic, which takes into account the degree of disagreement within the results (Table 2). Percentage assessments and contingency tables were used to describe the predictive validity. Statistical calculations were undertaken using Survey Plus, Survey Analysis-Statistics, Version 4.50 (Providence Software Services, U.K.).

Results

Descriptive Data

The test sample 5-year-olds comprised 23 (38%) females and 37 (62%) males. The overall mean age of these samples was 5.8 ± 0.7 (SD) years; with a range of 4.7 to 8.0 years. Of the 54 subjects used in the validity exercise, there were 20 (37%) females and 34 (63%) males. The mean age of these records was 17.6 ± 1.2 years, with a range of 15.4 to 20.4 years. Table 3 shows the mean ages and ranges of the different sexes in the age group.

Intra- and Inter-examiner Agreement

The weighted κ values demonstrating the intra-examiner agreement using the index are shown in Table 4. Overall,

TABLE 4 Intra-examiner Agreement (κ Values) for the Index Scores of the 5-Year-Old Study Models

Examiner	*Kappa Value*
A	.73 (.59)*
B	.86 (.76)
C	.96 (.90)
D	.87 (.77)

*The one-sided lower 95% confidence limits are in parentheses.

TABLE 6 Percentages of 5-Year-Old Subjects Assessed as Likely to Benefit from Orthognathic Surgery and the Required Need for Surgery of the 16- to 18-Year-Old Study Models*

Examiner	*5-Year-Olds Assessed in Groups 4 and 5 (n = 54)*	*16- to 18-Year-Olds Assessed as Requiring Orthognathic Surgery (%)*
A	13% (7)	
B	15% (8)	9% (5)
C	15% (8)	(consensus view)
D	18% (10)	

*The absolute figure is given in parentheses.

there was good to very good agreement within examiners. The weighted κ values demonstrating the inter-examiner agreement for the two separate assessments are shown in Table 5. Although the κ values ranged from .49 to .76 across the two assessments, there was, in general, good agreement between examiners using the 5-year-olds' index to assess the test sample. There was complete agreement between examiners for 47% (28) of cases across both assessments.

Predictive Validity

The relationship between numbers of subjects who were assessed as requiring an osteotomy at 16 to 18 years of age and their assigned index category at 5 years of age indicated little association between the two. Between 13% and 18% (depending on examiner) of 5-year-olds models were scored in the groups likely to benefit from orthognathic surgery (i.e., groups 4 and 5). Of the corresponding 16- to 18-year-old study models, 9% were assessed as likely to benefit from an osteotomy (Table 6). On an individual basis, however, the results demonstrate that it is not possible to predict future long-term growth from study models at age 5.

Discussion

The 5-year-olds' index tested in this study was shown to be reproducible and reliable with high levels of intra- and inter-examiner agreement. Over two assessments, the intra- and the inter-examiner agreement were unanimous in almost half of the cases. This was a subjective assessment of outcome, and not all patients could be tightly categorized. Where there was disagreement, it was never by more than one group. The development of methods to detect levels of outcome early is necessary if surgeons are to have a sound basis on which they can justify modifications of their timing or techniques. Early predictors of outcome provide a means to reduce the length of research studies without increasing sample size (Roberts et al., 1991). To date, the earliest stage at which CLP outcome has been detected is 10 years of age. This has been achieved with the Goslon Yardstick (Mars et al., 1987), which has proved reproducible and reliable in cross-center studies (Mars et al., 1992; Shaw et al., 1992). Its validity has not been investigated, and it is recognized that this is difficult since it requires a cohort of adults with CLP treated only with primary surgery. Dental study models at the age of 10 would also have to be available. It would then be possible to compare the adult and 10-year-old's dentition to see if Goslon ratings at 10 give a valid prediction in the adult dentition. To our knowledge, this sample does not exist since most patients undergo orthodontics, restorative procedures, and usually alveolar bone grafting (Shaw and Semb, 1990), all of which can mask the effects of primary surgery. Noverraz et al. (1993) maintained that the power of the Goslon Yardstick is in its face validity, which is said to apply when the relevance of a measurement appears obvious to the investigator (Abrahamson, 1990). A validation of the 5-year-olds' index was attempted by examining the corresponding adult dentition for each of 54 subjects. A panel consensus was used to decide whether or not adult subjects would benefit from an osteotomy. This decision making process has been successful and valid for other outcome measures in CLP (Asher-McDade et al., 1991; Roberts-Harry et al., 1991; Cussons et al., 1993). In some cases, orthodontics and restorative procedures may well have provided adequate dental compensation for skeletal discrepancy. Furthermore, the decision to perform an osteotomy would normally require facial assessment and other records such as a cephalometric radiograph. Within any population, there is variation in individual growth patterns, most follow the average but there are extremes. Growth prediction on an individual basis is currently not possible (Houston, 1979; Friede et al., 1988). In the sample examined, individual genetic variability meant that some subjects would have had small or large mandibles, which in turn would have influenced their final adult skeletal patterns. Any case with a large mandible automatically falls into a poor-outcome group, regardless of the quality of primary surgery. Facial growth is not fully expressed until after the pubertal growth spurt. Therefore, unless the subject

TABLE 5 Inter-examiner Agreement (κ Values) for the Index Scores of the 5-Year-Old Study Models*

Assessment	*Examiner*	*A*	*B*	*C*
First assessment	B	0.49 (0.31)	—	
	C	0.72 (0.58)	0.50 (0.34)	—
	D	0.65 (0.49)	0.66 (0.50)	0.65 (0.51)
Second assessment	B	0.67 (0.53)	—	
	C	0.76 (0.62)	0.56 (0.40)	—
	D	0.72 (0.58)	0.73 (0.59)	0.61 (0.45)

*The one-sided lower 95% confidence limits are in parentheses.

has a severe skeletal discrepancy (i.e., at the extremes of the normal range), the outcome measured at 5 years of age is assessing the influence of primary surgery. Furthermore, a 5-year-old sample is clean as far as orthodontic, restorative, or other surgical procedures are concerned. None of these are usually carried out at this early age; consequently, there is no masking of the skeletal pattern. Nevertheless, it was surprising to us that the percentage requiring osteotomy was similar in both the 5-year-old and adult groups. It must be emphasized that not all of the 5-year-old cases in groups 4 and 5 were in the adult group who required surgery. One further difficulty arises in that the records were from Oslo, a center with good results compared to others in Europe (Mars et al., 1992; Shaw et al., 1992), and is likely to have few individuals in the poor-outcome group, which might skew the predictive validity of the 5-year-olds' index.

An unconfirmed observation was that consultants who were most experienced in the treatment of clefts (B and D) appeared to score study models in lower groups than both the registrar (C) and one consultant (A), both of whom do not undertake cleft care. This would suggest there may be a systematic bias (i.e., experience with treatment of CLP subjects altered the perception of the results). This latter point may alter the face-validity claim that is made against the Goslon Yardstick (Noverraz et al., 1993) and, in turn, the present index. Reduction of this systematic bias and consistency between examiners might be aided by calibration prior to use of the index. This would aid the standardization (i.e, the reproducibility and reliability) of individual examiners (Richmond et al., 1993).

The development of early predictors of surgical outcome is timely and relevant. Since it is recommended that study models be obtained at 5, 10, 15, and 20 years of age for individuals with CLP as part of an internationally agreed protocol (Lee et al., 1993), the records should be available to assess surgical outcome at the age of 5. If standards of care can be reliably assessed at this age, it may prove easier to identify the quality of primary surgery. In turn, this would indicate techniques or surgical units that provide the best or worst results and allow the appropriate alterations in cleft care to be implemented (Shaw et al., 1992). We believe that prediction on an individual basis is not possible, but the index should allow inter-center and inter-technique comparisons. On a group basis, the individual variability in growth should be averaged, thus enabling group comparisons rather than individual predictions.

Previously, soft-tissue form has been investigated as a means of identifying differences in outcome between centers at 5 years of age (MacKay et al., 1994). Although this measure is a useful guide, the use of soft-tissue outline analysis is limited, since it is a two-dimensional picture of a three-dimensional structure. There are also inherent problems in achieving a standardized and accurate representation of soft tissues.

At present, there is some debate regarding the best methods of primary surgical treatment (Cabre et al., 1995; Markus and Ward Booth, 1995a,b; Pigott, 1995; Sell, 1995; Timmons, 1995). Although these authors disagree on the type of primary surgery, who should be performing this surgery, and where this surgery should be undertaken (i.e., small or regional units), all agree that continual assessment and audited results are essential to advance the management of CLP patients. The development of a rating system to examine surgical outcome at the age of 5 years should advance CLP research and help influence the improved management of CLP care. It is hoped to continue the development of the index with the inclusion of other clinical parameters such as an assessment of profile.

Conclusions

- The 5-year-olds' index tested in this study was reproducible and reliable.
- True validation of the index is not possible and it relies on face validity.
- The index is designed for inter-center comparison rather than individual prediction.
- Experience in treating clefts may be an important factor when applying the new index.
- Calibration of examiners in the use of the index may reduce systematic errors.
- This index has been developed to assess complete UCLP subjects, and like the Goslon Yardstick, its use is not appropriate to assess other types of clefts.

References

Abrahamson JH. Survey methods in community medicine. 4th Ed. London: Churchill Livingstone, 1990:152–153.

Altman DG. Practical statistics for medical research. London: Chapman & Hall, 1991:285–288, 403–409.

Atack NE, Hathorn IS, Mars M, Sandy JR. The use of study models of 5-year-old children as predictors of surgical outcome in unilateral cleft lip and palate. Eur J Orthod 1997 (in press).

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;28:385–391.

Cabre A, Giggin KP, Henley M, Nashed RN. Specialised multidisciplinary team can solve problems of distance (letter). BMJ 1995;311:1431–1432.

Cussons PD, Murison MS, Fernandez AE, Pigott RW. A panel based assessment of early versus no nasal correction of cleft lip nose. Br J Plast Surg 1993;46:7–12.

Friede H, Enocson L, Lilja J. Features of maxillary arch and nasal cavity in infancy and their influence on deciduous occlusion in unilateral cleft lip and palate. Scand J Plast Reconstr Surg Hand Surg 1988;22:69–75.

Houston WJB. The current status of facial growth prediction: a review. Br J Orthod 1979;6:11–17.

Lee ST, Cole O, McNanamny D, et al. International Committee on Cleft Documentation and Measurement: recommendations. Proceedings of 7th International Conference on Cleft Palate and Related Craniofacial Anomalies, Australia, 1993.

MacKay F, Bottomley J, Semb G, Roberts C. Dentofacial form in the 5-year-old child with unilateral cleft lip and palate. Cleft Palate Craniofac J 1994;31:372–375.

Markus T, Ward Booth P. Managed cleft lip and palate (editorial). BMJ 1995a;311:765–766.

Markus T, Ward Booth P. Author's reply. BMJ 1995b;311:1432–1433.

Mars M, 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 3. Dental arch relationships. Cleft Palate Craniofac J 1992;29:405–408.

Mars M, Plint DA, Houston WJB, et al. 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.

Noverraz AEM, Kuijpers-Jagtman AM, Mars M, Van't Hof MA. Timing of hard palate closure and dental arch relationships in unilateral cleft lip and palate patients: a mixed-longitudinal study. Cleft Palate Craniofac J 1993;30:391–396.

Pigott RW. Trials of management are inevitably long (letter). BMJ 1995; 311:1431.

Richmond S, Turbill EA, Andrews M. Calibration of non-dental and dental personnel in the use of the PAR Index. Br J Orthod 1993;20:231–234.

Roberts CT, Semb G, Shaw WC. Strategies for the advancement of surgical methods in cleft lip and palate. Cleft Palate Craniofac J 1991;28:141–149.

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

Sell DA. An invitation for Markus to present his results (letter). BMJ 1995; 311:1432.

Shaw WC, Dahl E, Asher-McDade C, et al. A six-center international study of treatment outcome in patients with cleft of the lip and palate: Part 5. General discussion and conclusions. Cleft Palate Craniofac J 1992;29: 413–418.

Shaw WC, Semb G. Current approaches to the orthodontic management of cleft lip and palate. J R Soc Med 1990;83:30–33.

Timmons MJ. Editorial proves controversial (letter). BMJ 1995;311:1431.