Evaluation of a modified Huddart/Bodenham scoring system for assessment of maxillary arch constriction in unilateral cleft lip and palate subjects

D. Gray and P. A. Mossey

University of Dundee Dental School, Scotland, UK

SUMMARY The aim of this retrospective study was to describe the use and reliability of a numerical scoring system (modified Huddart/Bodenham) for the measurement of maxillary arch constriction in patients born with unilateral cleft lip and palate (UCLP), and to compare and contrast the new scoring system with the current methods of measuring treatment outcome, the Goslon and 5-year-old indices.

Dental study models of 50 patients aged 5 years, and 50 patients aged 10 years, were scored using the 5-year-old and Goslon indices, respectively. Four examiners scored each set of models using the modified Huddart and Bodenham system, and repeated the scoring one month later. The intra- and inter-rater reliability of the numerical scoring system was assessed using the Kappa (κ) statistic. The scores using the new method were correlated with the 5-year-old and Goslon scores using Spearman's (ρ) and Kendall's (τ) rank correlation coefficients.

There was a high level of intra-rater reliability for both the 5 (0.87: incisors, 0.91: canines, 0.88: molars) and 10 (0.9: incisors, 0.84: canines, 0.78: premolars/molars) year models. The weighted κ values measuring inter-rater reliability were above 0.85 and 0.74, respectively, for all examiners. There was a statistically significant correlation between the modified Huddart/Bodenham scores and both the 5-year-old and Goslon scores in all cases (P < 0.001).

It is concluded that the modified Huddart/Bodenham system provides an objective and reliable assessment of maxillary arch constriction. It has a high degree of correlation with the recommended standards, but is more versatile and sensitive to inter-arch discrepancies.

Introduction

It is evident that there are a wide range of surgical techniques used for the correction of cleft lip and cleft palate, but with no clear-cut guidelines for optimal timing or method. Recent studies indicate that poorly performed primary surgery is likely to compromise facial growth, dental development (Mars *et al.*, 1992) and speech (Wyatt *et al.*, 1996). These and various other outcome measures such as facial aesthetics, nasal shape, and the presence of a fistula are criteria used to determine the quality of care. A European study involving six centres demonstrated that it is possible to detect differences in outcome (Shaw, 1992), and there is currently interest in the development of optimum strategies to improve treatment outcome.

Mars *et al.* (1987) introduced a standardized method of measurement of treatment outcome based on analysis of the skeletal and dental relationship using study models. This is described as the Great Ormond Street London, Oslo Norway (Goslon) ranking system and can be used to examine surgical outcome in the late mixed and/or early permanent dentition. In 1997 a similar index, the 5-year-old index (Atack *et al.*, 1997) was developed. The authors were able to demonstrate good correlation between both indices on longitudinal study models taken at 5 and 10

years of age in the same patient sample. A disadvantage of both of these indices is that there is an element of subjectivity in making the assessment and a calibration course is required for those who wish to use the indices for outcome assessment.

Among alternative systems for the measurement of the dental occlusion in studies to evaluate the results of cleft lip and palate (CLP) treatment is a system described by Huddart and Bodenham (1972). This system was applied to the primary occlusion and was used to examine the three segments of interest in unilateral cleft lip and palate (UCLP) cases, i.e. the labial, greater (non-cleft) and lesser (cleft) segments. As described, the buccal segments contain the canine and primary molars and in the labial segment only the central incisors were measured.

Each tooth was awarded a score depending on its position relative to its opponent in the lower arch and a total score obtained for each of the three segments. Figure 1 illustrates the elements and scoring system for the Huddart/Bodenham index.

A pilot study was carried out (Mossey *et al.*, 2003) to explore the possibility of a more objective system than the Goslon and 5-year-old indices for measurement of maxillary arch constriction in orofacial clefting. That



Figure 1 The Huddart/Bodenham system. (a) Segmental divisions of the maxillary arch. (b) Incisor scoring. (c) Canine scoring. (d) Molar scoring. Reproduced from the Cleft Palate Journal 1972; 9: 194–209 with kind permission.

study found the new scoring system to be a valid and reliable alternative to existing methods, with a number of perceived advantages.

The drawbacks of the Goslon and 5-year-old indices are:

Subjectivity. An element of subjectivity based on experience is inherent in the scoring system, which in turn is likely to adversely affect the intra- and inter-rater reliability. Both the 5-year-old and Goslon indices are 5 point scales and therefore a one or two category difference in score would result in a very significant level of error. Since the modified Huddart/Bodenham scoring system uses a cumulative score that is derived from six or eight separate categorical assessments, the effect of random operator error is minimized.

Calibration. Calibration courses are a prerequisite for using the 5-year-old and Goslon indices, and even with regular use re-calibration is necessary to ensure accuracy. Another potential weakness of the 5-year-old and Goslon scoring systems is that reference models must always be incorporated into the samples being analysed, adding to the complexity of the exercise. With the use of reference models there is also a tendency to look for similar features as opposed to similar severity. The aims of the present study were to:

- 1. Apply the modified Huddart/Bodenham scoring systems to the study models of patients with UCLP, previously scored using the Goslon and 5-year-old indices.
- Identify the inter- and intra-examiner variability of the modified Huddart/Bodenham scoring system applied to previously scored and categorized Goslon and 5-year-old models.
- 3. Determine the degree of correlation between the modified Huddart/Bodenham scoring system and Goslon and 5-year-old indices.
- 4. Provide further evidence of the reliability of an alternative, more objective, yet valid and versatile tool for measurement of surgical outcome using study models.

Materials and methods

Materials

Dental study models of 50 patients aged 5 and 50 patients aged 10 years, all with non-syndromic complete UCLP, were used in the study. The models were all of good quality, accurately trimmed, and distinguished only by a consecutive number to allow a repeatable ordered sequence over two separate episodes. They were both randomly chosen to be representative of the full spectrum of possible surgical outcomes, and mixed to eliminate bias. The developers of the 5-year-old and Goslon indices had previously scored the 5 and 10 year study models, respectively, to provide what could be regarded as 'reference' scores.

Examiners

Four examiners scored each set of models independently using the modified Huddart/Bodenham system. Two were consultant orthodontists calibrated in the 5-year-old and Goslon indices (examiners A and B), one a specialist registrar in orthodontics (examiner C), and one a qualified orthodontic technician with no experience of cleft patients (examiner D). Examiner A differed for each set of study models. The scoring of the 5 and 10 year models was undertaken by the developers of the 5-year-old and Goslon indices, respectively. Examiners B, C, and D were the same throughout the study. The same examiners repeated the scoring under similar conditions one month later, to allow calculation of inter- and intra-examiner reliability and minimize the possible influence of memory on the results.

Modifications to the scoring system

The Huddart/Bodenham system was designed for use in the primary dentition and it was therefore necessary to modify it for use in the mixed dentition. This was undertaken by scoring premolars in the same way as primary molars. The sum of the modified Huddart/Bodenham scores for a given model was described as the 'total arch constriction score'.

Rules were also drawn up for situations where there might be ambiguity. For example, when one central incisor was missing the other central incisor was scored. When a canine was unerupted, its score was determined by the midpoint of the maxillary alveolar ridge. Where a premolar was absent due either to non-eruption or hypodontia, a score was allocated equivalent to the adjacent premolar, if erupted. If no premolars were erupted, the same rule as for the canine was applied, i.e. the score was determined by the mid-point of the maxillary alveolar ridge.

The modified system requires that all teeth from first permanent molar forward be given a score to reflect the maxillary arch constriction. The number of teeth scored changes when a patient reaches 6 years of age. Before the age of 6, the first permanent molars were not scored, even if erupted (maximum range of scores: -18 to +2). After and including 6 years of age, the first permanent molars were scored if present, or the mid-point of the maxillary alveolar ridge used in a similar way as described previously (maximum range of scores: -22 to +2).

Statistical analysis

Intra- and inter-examiner variability were calculated using Cohen's weighted and unweighted Kappa (κ) statistics (Altman, 1991). The Huddart/Bodenham scores were correlated with the 5-year-old and Goslon scores using Spearman's(ρ) and Kendall's(τ) rank correlation coefficients. The κ statistic relates the actual measure of agreement obtained with the degree of agreement that would have been attained by chance. It is suggested that κ greater than 0.8 indicates good agreement, greater than 0.6 substantial agreement and greater than 0.4 moderate agreement.

Results

Intra-rater reliability

Table 1 shows the aggregate weighted and unweighted κ values derived from comparing the first and second ratings of the four examiners (intra-rater reliability).

Huddart/Bodenham scores of 5-year-old calibration models. For the intra-rater reliability, the weighted κ values indicated a high level agreement for the incisors (0.87), canines (0.91), and molars (0.88). The mean unweighted κ value for the incisors (0.75) indicated substantial agreement. The canine (0.87) and molar (0.83) mean unweighted κ values indicated a high level of agreement.

Huddart/Bodenham scores of Goslon calibration models. The intra-rater reliability weighted κ values (0.9 for the incisors, 0.84 for the canines, and 0.78 for the premolars and molars) indicated a high level of reliability. The mean unweighted κ values (0.78 for the incisors, 0.75 for the canines, and 0.7 for the premolars and molars), indicated substantial intra-rater

reliability. The reduced weighted and unweighted canine premolar and molar values for the Goslon compared with the 5-year-old models, was probably due to the buccal occlusion in the mixed dentition being less well defined.

Inter-rater reliability

The inter-rater reliability takes into account the scores from each examiner averaged over the two scoring episodes compared with the other examiners (Table 2).

For all six examiner pairs using the 5-year-old models, the weighted κ value was above 0.85 and the unweighted κ value above 0.75, indicating high inter-rater reliability. It should be noted that the two uncalibrated 5-year-old examiners (C and D) recorded the highest agreement (0.88 weighted and 0.8 unweighted) κ values.

For the Goslon models the inter-rater reliability was above 0.7 for the weighted, and 0.6 for the unweighted κ values. These values were lower compared with the 5-year-old models. A possible reason for this variation may be that as the Goslon models reflect the mixed dentition, a larger number of teeth are scored, some of which may be unerupted or partially erupted, which means a greater chance of variation.

Correlation between total arch constriction and 5-year-old/Goslon score

Results from Spearman's (P) and Kendall's tau (τc) rank correlation coefficients are recorded in Table 3. They

 Table 1
 Weighted and unweighted kappa values for intraexaminer reliability.

Tooth type	5-year-old models		Goslon models	
	Weighted kappa	Unweighted kappa	Weighted kappa	Unweighted kappa
Incisors	0.871	0.749	0.897	0.78
Canines	0.912	0.866	0.836	0.75
Premolars/molars	0.884	0.827	0.783	0.696

 Table 2
 Weighted and unweighted kappa values for interexaminer reliability.

Examiners	5-year-old models		Goslon models	
	Weighted kappa	Unweighted kappa	Weighted kappa	Unweighted kappa
AvB	0.853	0.769	0.768	0.641
AvC	0.882	0.812	0.746	0.61
A v D	0.887	0.817	0.827	0.684
B v C	0.877	0.796	0.861	0.799
ΒvD	0.864	0.772	0.865	0.789
C v D	0.878	0.796	0.849	0.771

Examiner	5-year-old models	5-year-old models		Goslon models	
	Spearman's P	Kendall's tau (τc)	Spearman's P	Kendall's tau (τc)	
A	-0.914	-0.876	-0.82	-0.736	
В	-0.921	-0.888	-0.861	-0.769	
С	-0.911	-0.874	-0.874	-0.782	
D	-0.902	-0.856	-0.864	-0.771	

 Table 3
 Correlation coefficients for both the 5-year-old and Goslon models.

P < 0.001 for all values.

indicated a statistically significant correlation in all cases (P < 0.001), consistently greater than 0.8 (range 0.8 to 0.91), for those who scored the 5-year-old models, and greater than 0.7 (range 0.7 to 0.87), for all examiners who scored the Goslon models.

The modified Huddart/Bodenham system was therefore both reliable and capable of categorizing the severity of the models into similar categories as the 5-year-old and Goslon indices.

Both uncalibrated examiners (C and D), were comparable with the calibrated examiners (A and B), and in fact achieved higher correlation values (Spearman's P value of 0.87 and 0.86, respectively) between the total arch constriction scores and the Goslon models. This suggested a high level of reliability.

Discussion

The development of early predictors of outcome in cleft care is timely and relevant. In some countries, including the UK, the standard of care is perceived to be below that of the best European centres, and the problem of surgical outcome is compounded by surgeons who perform few operations each year (Williams *et al.*, 1994). In a survey of CLP primary surgical services in Europe, 175 cleft teams and almost as many different cleft treatment protocols were identified (Shaw, 1992). It was also observed that small caseloads make meaningful statistical analysis almost impossible for an individual surgeon. It has been suggested that larger centres could act as a reference norm against which smaller units could measure their success (Ward-Booth, 1995).

The modified Huddart/Bodenham scoring system could be used for this comparison and the percentiles of each possible total arch constriction score would be an important parameter in the assessment of surgical outcome. The larger range of the modified Huddart/Bodenham scoring system compared with the 5-year-old and Goslon indices would provide a more sensitive assessment of surgical outcome.

A report from the Clinical Standards Advisory Group (1998) suggested a minimum caseload of 30 new patients annually for primary surgery in CLP. This number is a reflection of the need to be able to audit the outcome so that effects of surgical protocols can be monitored objectively.

Currently, the Goslon index (Mars *et al.*, 1987) and 5year-old index (Atack *et al.*, 1997) are widely regarded as the best available measures for assessing surgical outcome. The dilemma, however, is that with a 5 point scale, such as used in the 5-year-old and Goslon indices, in order to detect a difference of 0.5 at 5 per cent probability and with 80 per cent power, an annual case load of some 60 patients over a period of 8.5 years is required. This calculation assumes that differences are detectable at 5 years of age (Shaw, 1992). There is a substantial difference in the level of sensitivity using a 25 point modified Huddart/Bodenham scale (-22 to +2), as opposed to the 5 point scales of the 5-year-old and Goslon indices, and therefore for the same surgeon caseload, the same effect size can be detected in a shorter timescale.

The advantages of the modified Huddart/Bodenham scoring system are:

Objectivity combined with relative simplicity. The results of the study indicate that no clinical experience is required to reliably use the modified Huddart/Bodenham system. Examiner D, a dental technician, achieved high intra- and inter-rater reliability scores, which compared favourably with calibrated and experienced orthodontists. The use of non-professional auxillary staff to score models would simplify the training of assessors, and improve inter-centre collaboration studies.

Versatility. The Huddart/Bodenham scoring system can be applied to models of any cleft subgroup and at any age. This is important as the recommendation for the appropriate age to obtain study casts varies, and there are differences for UCLP and isolated cleft palate.

Sensitivity. This modified scale is an ordinal continuous scale of severity of arch constriction rather than a categorical scale. This enables differentiation of severity within the categories that would be identified by the 5-year-old or Goslon indices. It lends itself to the more easily interpreted and discriminating non-parametric statistical tests rather than contingency table testing. It is important to note, however, that while the modified system is an ordinal scale, it is not an equal interval scale, but the greater range of scoring improves the sensitivity.

Digital recording. The measurements used in the modified Huddart/Bodenham scoring system lend themselves to calculations based on the assessment of digital images. This would speed up the measurement and analysis of data and allow easy inter-centre comparisons to be made. Arch constriction and buccal crossbite could all be recorded from scanned digital images, and the development of a computer program would enable the assimilation of this data to provide a score. A system that incorporates a subjective element does not allow this computer-based approach.

Conclusions

The following points summarize the findings from this study, and describe the future possibilities for this technique.

- The modified Huddart/Bodenham scoring system is a valid and reliable indicator of outcome of treatment for patients with UCLP.
- 2. Intra- and inter-examiner agreement are both excellent.
- 3. There is a high degree of correlation with the Goslon and 5-year-old indices.
- 4. It provides a sensitive and objective assessment of maxillary arch constriction.
- 5. This scoring system could be used on the study models of any cleft subtype from the time of eruption of the complete primary dentition.
- 6. In the future it may be possible to obtain scanned digital images from models or intra-oral images, and produce an arch constriction score from these.

Address for correspondence

Professor Peter A Mossey Dundee Dental School Park Place Dundee DD1 4HR UK E-mail: p.a.mossey@dundee.ac.uk

Acknowledgements

The authors would like to thank Dr Nicki Atack and Dr Michael Mars, who allowed access to their departmental facilities, and gave their time freely to help with the scoring. We are also grateful for the assistance of John Clark and David Stirrups, both at Dundee Dental School, for their advice and assistance during this study.

References

- Altman D G 1991 Practical statistics for medical research. Chapman & Hall, London
- Atack N, Hathorn I, Mars M, Sandy J 1997 Study models of 5 year old children as predictors of surgical outcome in unilateral cleft lip and palate. European Journal of Orthodontics 19: 165–170
- Clinical Services Advisory Group (CSAG) Report 1998 Cleft lip and/or palate. Royal College of Surgeons of England. ISBN 0 11 322103 77
- Huddart A G, Bodenham R S 1972 The evaluation of arch form and occlusion in unilateral cleft palate subjects. Cleft Palate Journal 9: 194–209
- Mars M, Plint D A, Houston W J B, Bergland O, Semb G 1987 The Goslon Yardstick: a new system of assessing dental arch relationships in children with unilateral clefts of the lip and palate. Cleft Palate Journal 24: 314–322
- Mars M *et al.* 1992 A six-center international study of treatment outcome in patients with clefts of the lip and palate: Part 3. Dental arch relationships. Cleft Palate-Craniofacial Journal 29: 405–408
- Mossey P A, Clark J D, Gray D 2003 Preliminary investigation of a modified Huddart/Bodenham scoring system for assessment of maxillary arch constriction in unilateral cleft lip and palate subjects. European Journal of Orthodontics 25: 251–257
- Shaw W C 1992 Intercentre clinical audit for cleft lip and palate a preliminary European investigation. In: Jackson I T, Sommerlad B C (eds) Recent advances in plastic surgery. Churchill Livingston, Edinburgh, pp. 1–29
- Ward-Booth P P 1995 Services for cleft lip and palate. British Medical Journal 310: 1406
- Williams A, Shaw W C, Devlin H B 1994 Provision of services for cleft lip and palate in England and Wales. British Medical Journal 309: 1552
- Wyatt R, Sell D, Russell J, Harding A, Harland K, Albery E 1996 Cleft palate speech dissected: a review of current knowledge and analysis. British Journal of Plastic Surgery 49: 143–149

Copyright of European Journal of Orthodontics is the property of Oxford University Press / UK and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.