European Journal of Orthodontics 34 (2012) 762–767
doi:10.1093/ejo/cjr092
Advance Access publication 5 October 2011

Calibration of the modified Huddart and Bodenham scoring system against the GOSLON/5-year-olds' index for unilateral cleft lip and palate

L. M. Dobbyn*, J. T. Weir**, T. V. Macfarlane*** and P. A. Mossey**
*Orthodontic Department, Glasgow Dental Hospital and School, UK, **Orthodontic Department, School of Dentistry, University of Dundee and ***Division of Applied Medicine, School of Medicine and Dentistry, University of Aberdeen, UK

Correspondence to: L. Dobbyn, Orthodontic Department, Glasgow Dental Hospital and School, 378 Sauchiehall Street, Glasgow G2 3JZ, UK. E-mail: lornadobbyn@hotmail.com

SUMMARY The GOSLON/5-year-old scoring systems have been used in various national and international studies. This study aims to identify the range of Modified Huddart and Bodenham (MHB) scores, which correlate with each GOSLON and 5-year-old category and thereby create a new scoring system that allows comparison with historical data. Two hundred and eighty-three unilateral cleft lip and palate study models from England and Scotland that had all been previously scored using the 5-year-old and GOSLON indices by calibrated examiners were scored using MHB on two separate occasions a month apart by two examiners. Reliability analysis using intraclass correlation and Bland Altman plots were performed. Ordinal regression was used to define the categories of MHB that correspond to the 5-year-old and GOSLON categories. The results revealed a high level of repeatability for both the 5 and the 10 year old models. The MHB scale was grouped into the five categories of the GOSLON and 5-year-old indices with more precision for the 5 year than the 10 year scores. In conclusion, there was high intra-examiner and inter-examiner agreement using the MHB scoring system. It also correlates well with the 5-year-old and GOSLON indices and has been shown to be a much more sensitive scoring system.

Introduction

Background

Studies indicate that poorly performed primary cleft surgery is likely to compromise facial growth, dental development, and speech. Unfortunately, the optimum timing or technique for this surgery is not known. In the search to find this ideal, early predictors of growth and dental arch form are essential. Precise and reproducible outcome measures are a necessity as evidence-based treatment and guidelines are an essential part of modern clinical practice.

There are many previously described methods for scoring arch relationships and dental irregularities in orofacial clefting. Examples include Pruzansky and Aduss (1964), Matthews *et al.* (1970), and Huddart and Bodenham (1972). Mars *et al.* (1987) introduced a system for measuring treatment outcome based on the analysis of dental relationships of study models of children with unilateral cleft lip and palate (UCLP) at the age of 10 years called the Great Ormond Street, London, and Oslo, Norway (GOSLON) index. The scoring is done by comparing with a set of standardized models and each model is given a score from 1 to 5, where 1 = excellent and 5 = very poor outcome. The factors considered when scoring are the antero-posterior arch relationships, the vertical labial segment relationships, and the transverse relationships. In 1997, a similar index, the 5-year-olds' index, was developed by Atack *et al.* (1997) for 5 year olds. A perceived disadvantage of both of these indices is their inherent subjectivity. A calibration course and reference models are a prerequisite to using the 5-year-old and GOSLON scoring systems. Longitudinal incompatibility between the 5-year-old and GOSLON indices was demonstrated by Mars *et al.* (2006).

An alternative system for the measuring cleft affected dental occlusions was described by Huddart and Bodenham (1972). Maxillary tooth positions or arch attributes are scored relative to the mandibular arch in maximum inter-cuspation. Each maxillary tooth or midpoint of the arch (if the tooth is missing) is scored according to its relationship with the corresponding tooth in the mandible (Figure 1). A cumulative score is derived from 8 categorical assessments in the primary dentition and 10 in the mixed or permanent dentitions. The more negative the score, the more severe the arch constriction. The advantages of the Modified Huddart and Bodenham (MHB) system introduced by Mossey *et al.* (2003) are its objectivity, relative simplicity, and avoidance of the need for a calibration course, the last being an important feature in the developing world. It is also versatile as it can be applied to any cleft subgroup at any age (Tothill and Mossey, 2007). The MHB uses a 40 point ordinal quasi-continuous scale, which allows differentiation of outcome between cases that would be identified as the same with the 5-year-old or

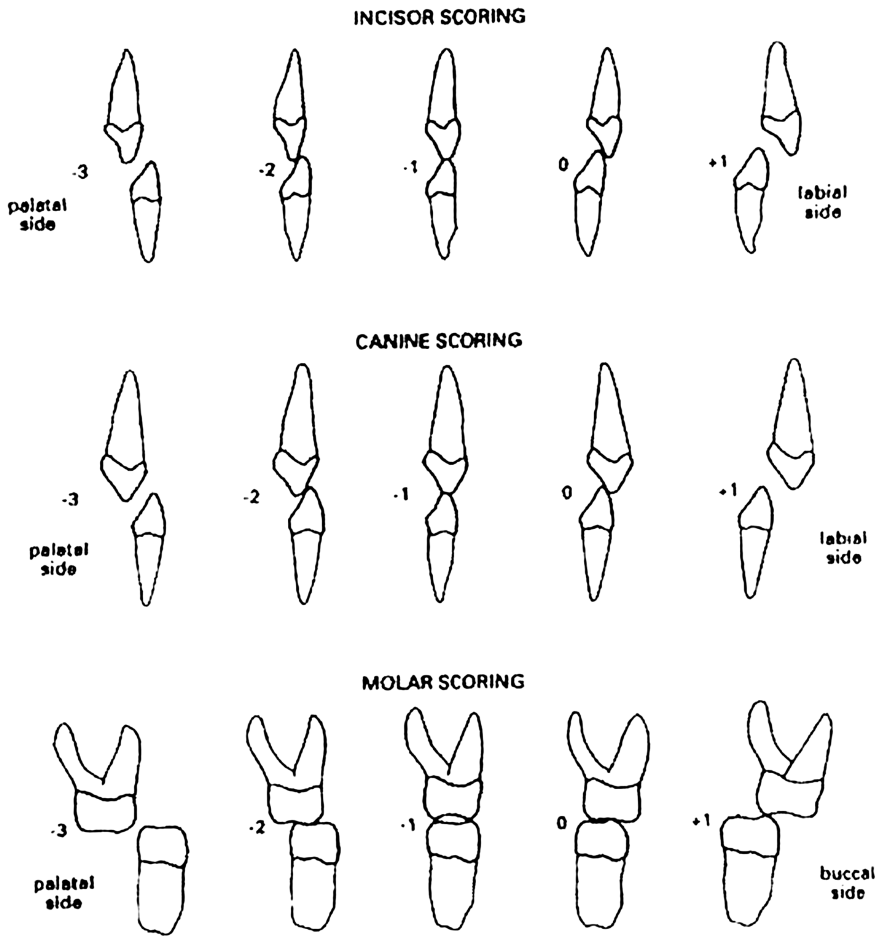

Figure 1 Modified Huddart and Bodenham scoring instructions.

GOSLON 5 point indices. Potential disadvantages of the MHB system are that it does not take into account the underlying antero-posterior skeletal discrepancy, incisor inclination, and that is based on crossbites some of which may be treatable. However, many are extremely difficult to treat with nearly 50 per cent of patients with UCLP requiring orthognathic surgery in the CSAG study (Williams *et al.*, 2001). There is also a very high correlation between Class III skeletal discrepancy and Class III incisor relationship.

Aims

1. To create a system in study model scoring that allows the MHB score to be converted into categorical scores for the purpose of comparison with historical data. In the past, GOSLON/5-year-old scores have been used in various national and international studies.
2. Identify the range of MHB scores, which correlate with each GOSLON and 5-year-old category.
3. Identify the inter- and intra-examiner reliability of the MHB scoring system applied to models previously scored and categorized using the GOSLON and 5-year-olds' indices.

Materials and methods

Ethical approval was obtained from the Tayside Committee on Medical Research Ethics (LREC reference no. 06/S1402/54), and Research and Development approval was received from NHS Tayside Research Consortium (Project ID: 2006OR01).

A total sample of 280 dental study models including 184 five year and 96 ten year old UCLP patients was gathered from the CLEFTSiS database in Scotland and a database at Liverpool Dental Hospital. Calibrated examiners had previously scored these casts using the 5-year-olds' index and the GOSLON index.

1. For this study, two examiners scored the casts independently on two occasions 1 month apart using the MHB index.
2. The MHB results at 5 and 10 years were then compared with the 5-year-old and GOSLON index scores.

Statistical analysis was carried out using SPSS 16.0 for Windows (SPSS, 2008) statistical package. Descriptive statistics included means, standard deviations, and boxplots. Intra- and inter-examiner reliability was calculated using

the intraclass correlation coefficient (ICC), reliability coefficients (Cronbach's Alpha), and Bland Altman plots (Bland and Altman, 1986). The ICC is a measure of reliability, correlation, and consistency. It assesses rating reliability by comparing the variability of different ratings of the same subject to the total variation across all ratings and subjects. Cronbach's alpha is a measure of internal consistency, based on the average inter-item correlation.

Ordinal regression was used to define the categories of MHB that correspond to the 5-year-olds' index and GOSLON categories, which are ordinal variables with values ordered from 1 to 5 from excellent to very poor outcome.

Table 1 Intraclass correlation coefficient for intra- and inter-examiner scoring of 5 year old models.

	Intraclass correlation coefficient	Cronbach's alpha	Mean difference	Standard deviation (SD)	Limits of agreement (mean difference ± 1.96 SD)	
Intra-examiner A	0.971	0.985	−0.43	1.301	−3.032	2.172
Intra-examiner B	0.978	0.989	0.054	1.090	−2.126	2.234
Inter-examiner	0.982	0.991	0.102	0.905	−1.672	1.876

Table 2 Intraclass correlation coefficient for intra- and inter-examiner scoring of 10 year old models.

	Intraclass correlation coefficient	Cronbach's alpha	Mean difference	Standard deviation (SD)	Limits of agreement (mean difference ± 1.96 SD)	
Intra-examiner A	0.994	0.997	0.083	0.854	−1.625	1.791
Intra-examiner B	0.993	0.996	−0.198	0.902	−2.002	1.606
Inter-examiner	0.996	0.998	0.036	1.424	−1.388	1.46

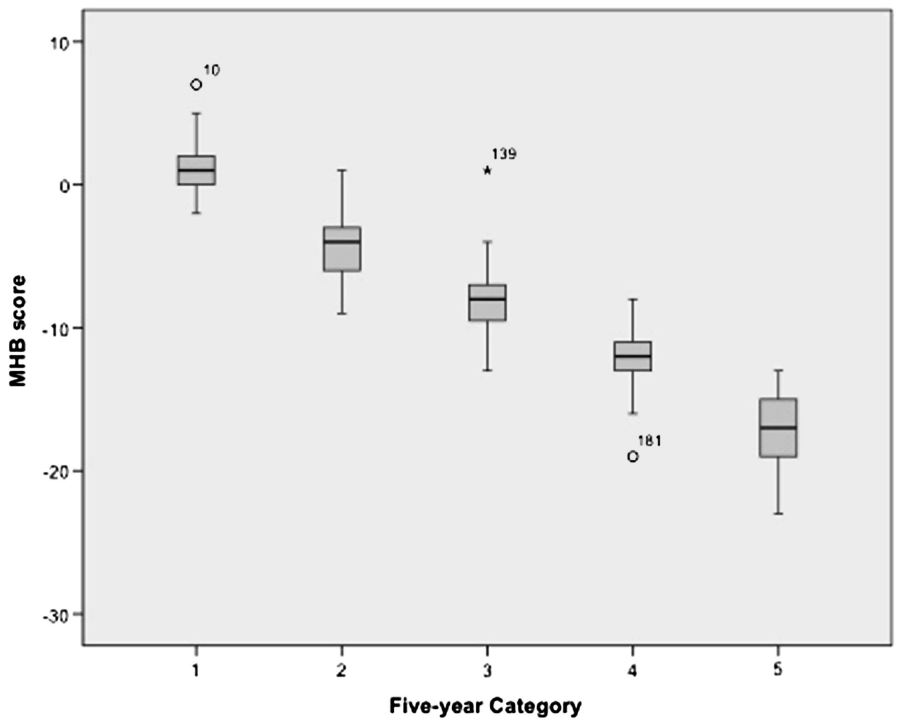

Figure 2 Boxplot of distribution of modified Huddart and Bodenham scores according to 5-year-olds' index category.

The explanatory variable (predictor) was MHB score, which was a mean MHB score of two examiners. It requires the assumption of MHB parallel lines across all levels of 5-year-old/GOSLON scores. Five different link functions are available in the ordinal regression tool in SPSS: Logit, complementary log-log, negative log-log, Probit, and Cauchiti (SPSS, 2008). Goodness of fit was assessed using the likelihood-ratio test and pseudo R-squared (proportion of the total variability explained by the model). Test of parallel lines was used to test the null hypothesis that the slope coefficients were the same across 5-year-old/GOSLON categories, which is the assumption for ordinal regression. Negative log-log was chosen for further analysis as it had the best fit. The predicted values were compared with the actual 5-year-old/GOSLON scores. Kappa statistics (Landis and Koch, 1977) was used to measure the agreement between observed and predicted GOSLON scores.

Table 3 Predicted response category crosstabulation for 5-year-olds' index.

Five year score	Predicted response category, *N* (%)				
	1	2	3	4	5
1	19 (79.2)	5 (20.8)	0	0	0
2	4 (7.3)	43 (78.2)	8 (14.5)	0	0
3	2 (3.3)	7 (11.7)	45 (75)	6 (10)	0
4	0	0	13 (39.4)	18 (54.5)	2 (6.1)
5	0	0	0	4 (33.3)	8 (66.7)
Total	25	55	66	28	10

Results

The results (Tables 1 and 2) show that the ICC and the reliability coefficients are all high (close to 1). The results also demonstrate small mean differences. This all equates to excellent examiner reliability both between the measurements of the two examiners on both scorings and between examiners. This shows that the MHB scoring index is consistently reliable and reproducible in both age categories.

Figure 2 illustrates the distribution of MHB scores in the five categories described in the 5-year-olds' index. This demonstrates a progressive trend of a relationship towards increasing severity in both scales. The grouping of a continuous numerical scale (MHB) into the five groups of the 5-year-olds' index demonstrates that there is a trend from best to worst in both illustrating negative correlation. As the 5-year-old score increases, this correlates with a more negative (decreasing) MHB score.

There is inevitably some overlap between the five categories showing that each 5-year-old category contains a range of MHB scores. However, this overlap is contained within the whiskers of the plot, which are beyond the 25th and 75th percentiles. The boxes in the plot are all separate and the medians are moderately well distributed, about four or five MHB points apart.

Relationship between 5-year-old/GOSLON and MHB score

Table 3 shows the predicted response for the Negative log-log function for 5-year-old scores. It estimates the percentage of MHB results that are predicted in the correct 5-year-old

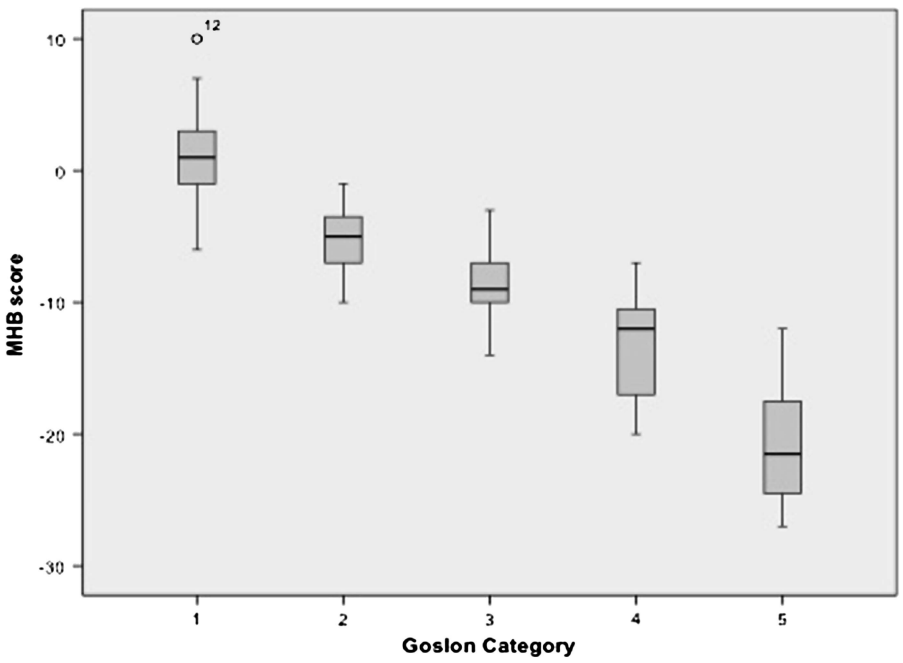

Figure 3 Boxplot of distribution of modified Huddart and Bodenham scores according to GOSLON category.

category. As can be seen, the lower categories (1–3) are more accurate, predicting between 75 and 79 per cent of results correctly. Kappa = 0.628 ($P < 0.001$), which indicates substantial agreement. The results however are still helpful as more and more patients are being classified in the lower categories due to improvements in cleft surgery and outcomes. There is therefore a greater need to distinguish between patients in these lower categories.

Figure 3 illustrates the distribution of MHB scores in the GOSLON categories for the 10 year old data. This demonstrates a progressive trend of a relationship towards increasing severity in both scales. The grouping of a continuous numerical scale (MHB) into the five groups of the GOSLON index demonstrates that there is a trend from best to worst in both illustrating negative correlation. As the GOSLON category increases, this correlates with a more negative (decreasing) MHB score.

Table 4 shows the predicted response for the Negative log-log function. It estimates the percentage of MHB results that are predicted in the correct GOSLON category. As can be seen, categories 1, 3, and 5 are most accurate (between 65.4 and 78.9 per cent). It is shown to be more difficult to predict categories 2 and 4 (55 and 42.1 per cent, respectively). Kappa = 0.509 ($P < 0.001$), which shows moderate agreement.

It was noted that when the 5-year-old and GOSLON indices were analysed separately in this study, the cut offs for the five categories of each index differ slightly (Table 5).

Discussion

Many scoring systems to assess arch constriction in UCLP exist. Presently, the 5-year-old and GOSLON indices are regarded as the best available objective measures for assessing surgical outcome. However, the dilemma here 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 (Shaw *et al.*, 1992). This calculation assumes that differences are detectable at 5 years of age. As surgical outcomes improve and protocols become more standardized, there will be an ongoing tendency towards convergence of scores in categories 2, 3, and 4 thus limiting discrimination and ability to detect more subtle differences in surgical outcome.

The MHB system is capable of reliably categorizing the severity of outcome measured on models into similar categories as the 5-year-old and GOSLON indices, with a high degree of precision particularly in the former. However, the relationship between the two scales was less predictable in 5 year groups 4 and 5 as well as GOSLON groups 2 and 4. This is probably related to the inherent differences in what the two systems are measuring in relation to incisor decompensation and lesser segment collapse in unilateral clefts.

The MHB scoring system provides a simple, yet sensitive method of assessing total arch constriction and therefore assessment of surgical outcome. The larger range of the MHB scoring system compared with the 5-year-old and GOSLON indices would provide a more sensitive assessment of surgical outcome. Yet, if cleft units still wished to compare their results on the 5 point scale that has become a feature of outcome measurement in clefts, the corresponding score on the GOSLON/5-year-old categorical index can still be used. The sensitivity of MHB is required to differentiate between outcomes that are so close together that 5-year-old/GOSLON does not distinguish between them. MHB clearly grades within the bands identified when using 5-year-old/GOSLON.

Table 4 Predicted response category crosstabulation for GOSLON scores.

GOSLON score	Predicted response category, *N* (%)				
	1	2	3	4	5
1	15 (78.9)	4 (21.1)	0	0	0
2	3 (15)	11 (55)	6 (30)	0	0
3	0	4 (15.4)	17 (65.4)	5 (19.2)	0
4	0	0	8 (42.1)	8 (42.1)	3 (15.8)
5	0	0	0	4 (33.3)	8 (66.7)
Total	18	19	31	17	11

Conclusions

Overall, intra- and inter-examiner agreement is excellent and there is a high degree of correlation with the GOSLON and 5-year-olds' indices. MHB provides a sensitive and objective assessment of maxillary arch constriction and as it is a continuous numerical scale, it lends itself to statistical analysis.

This study therefore describes a reliable indicator of surgical outcome, a viable alternative to the GOSLON, and 5-year-olds' indices for patients with UCLP, but one that

Table 5 Estimated modified Huddart and Bodenham scores for each 5-year-olds' index category and each GOSLON index category.

Category	1	2	3	4	5
Estimated scores for each 5-year-olds' index category	+8 to −1	−2 to −6	−7 to −10	−11 to −14	−15 to −24
Estimated scores for each GOSLON index category	+10 to −2	−3 to −7	−8 to −10	−11 to −17	−18 to −30

can also be used in conjunction. This appears to have some advantages over previously used methods in objectivity, versatility, and simplicity; but also the improved sensitivity means that smaller more subtle changes in outcome can be reliably measured. This ultimately will improve the profession's approach to outcome measurement in cleft lip and palate inter-centre studies where variations in surgical techniques, operator skill, patient craniofacial morphology, and malocclusion all increase the variation, and demand more sensitivity in the outcome measurement.

Funding

NHS Tayside (Research and Development Project ID: 2006OR01).

Acknowledgements

Ms Mairi Milne, CleftSiS MCN Manager, Perth Royal Infirmary, UK and Dr Joyce Russell, Consultant Orthodontist, Alder Hey Hospital, Liverpool, UK for assistance with access to Scotland and Liverpool UCLP study model archives, respectively.

References

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

Bland J M, Altman D G 1986 Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1: 307–310

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

Landis J R, Koch G G 1977 The measurement of observer agreement for categorical data. Biometrics 33: 159–174

Matthews D, Broomhead I, Grossman W, Goldon H 1970 Early and late bone grafting in cases of cleft lip and palate. British Journal of Plastic Surgery 23: 115–129

Mars M, Plint D A, Houston W J, Bergland O, Semb G 1987 The GOSLON yardstick: a new system of assessing dental arch relationships in children with unilateral cleft lip and palate. Cleft Palate Journal 24: 314–322

Mars M, Batra P, Worrell E 2006 Complete unilateral cleft lip and palate: validity of the five-year index and the Goslon yardstick in predicting long-term dental arch relationships. Cleft Palate-Craniofacial Journal 43: 557–562

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 (UCLP). European Journal of Orthodontics 25: 251–257

Pruzansky S, Aduss H 1964 Arch form and the deciduous occlusion in complete unilateral clefts. Cleft Palate Journal 1: 411–418

Shaw W C *et al.* 1992 A six-center international study of treatment outcome in patients with clefts of the lip and palate: part 5. General discussion and conclusions. Cleft Palate-Craniofacial Journal 29: 413–418

SPSS. 2008 Statistical package for the social sciences. 16.0 edn. Chicago, SPSS Inc.

Tothill C, Mossey P A 2007 Assessment of arch constriction in patients with bilateral cleft lip and palate and isolated cleft palate: a pilot study. European Journal of Orthodontics 29: 193–197

Williams A C *et al.* 2001 Cleft lip and palate care in the United Kingdom—the Clinical Standards Advisory Group (CSAG) Study. Part 2: dentofacial outcomes and patient satisfaction. Cleft Palate-Craniofacial Journal 38: 24–29