Comparison of the Modified Huddart/Bodenham and GOSLON Yardstick Methods for Assessing Outcomes Following Primary Surgery for Unilateral Cleft Lip and Palate

Montian Manosudprasit DDS, MDS*, Tasanee Wangsrimongkol DDS, MS, PhD*,
Somsak Kitsahawong DDS, MS*, Thanarat Thienkosol DDS*

** Department of Orthodontics, Faculty of Dentistry, Khon Kaen University, Khon Kaen, Thailand*

Objective: *To test the measure of agreement between the modified Huddart/Bodenham scoring system and the GOSLON Yardstick for assessing the dental occlusions of patients with unilateral complete cleft lip and palate (UCLP) and to test the time taken for each assessment.*

Material and Method: *60 sets of study models of 8-10-year-old UCLP patients who attended the Department of Orthodontics at the Faculty of Dentistry, Khon Kean University were evaluated. All subjects had undergone their cleft lip and palate repairs, but no alveolar bone grafts or any orthodontic treatment were done. The judgments of two trained examiners were used to place the modified Huddart/Bodenham score for each set of models into one of 5 categories corresponding to the GOSLON ratings to test for agreement between the two methods. The strength of agreement of ratings was analyzed by weighted kappa statistics. A paired t-test was carried out to compare the time taken in assessment with each index.*

Results: *There was good agreement between the two methods with a kappa value of 0.73. The GOSLON assessment took significantly less time than the modified Huddart/Bodenham assessment.*

Conclusion: *The modified Huddart/Bodenham scoring system can be used as an alternative to the more commonly used GOSLON Yardstick for diagnostic purposes. Although the numerical scoring system takes more time, it provides more information about the sites of occlusal discrepancy than does the GOSLON Yardstick.*

Keywords: *Modified Huddart/Bodenham scoring system, GOSLON Yardstick, Cleft lip and palate*

J Med Assoc Thai 2011; 94 (Suppl. 6): S15-S20
Full text. e-Journal: *http://www.jmat.mat.or.th/journal*

Assessment of dental arch relationships for patients with cleft lip and palate can reflect their surgical outcomes which are advantageous in giving surgeons feedback that they may use in seeking possible improvements to their surgical protocol. To do this, several assessment indices for dental occlusions have been developed, such as the GOSLON (Great Ormond Street, London and Oslo, Norway) Yardstick[1] and the modified Huddart/Bodenham scoring system[2] which are two of those commonly used. The former index uses an ordinal scale of five categories, excellent, good, fair, poor, very poor, to identify an individual's occlusal status with a consequent general prediction about the relative complexity of future correction of malocclusion. The modified Huddart/Bodenham system provides an additive score for malpositions of right and left maxillary teeth recorded as a continuous range from +2 (equivalent to an excellent condition) to -22 (exceedingly severe malocclusion). A potential problem for those assessors who wish to use the GOSLON Yardstick is that they are required to have special pre-assessment training for their findings to be accepted[3]. This is not required for the use of the modified Huddart/Bodenham scoring system so making the latter a more accessible application. The GOSLON assessment is a subjectively ordered and a broadly categorical classification of dental occlusions which seems to be less powerful than the more objective numerical grading scale of the modified Huddart/Bodenham assessment system which also identifies sites of malocclusion of the maxillary dentition.

Correspondence to:
Manosudprasit M, Department of Orthodontics, Faculty of Dentistry, Khon Kaen University, Muang, Khon Kaen 40002, Thailand.
Phone & Fax: 043-202-863
E-mail: monman@kku.ac.th

The aim of the present study was to compare the assessment of socclusal outcomes, in repaired unilateral cleft lip and palate (UCLP) patients evaluated by the modified Huddart/Bodenham scoring system with use of the GOSLON Yardstick, in terms of the level of agreement, reliability of the assessment and time taken for model assessment.

Material and Method

Sixty sets of study models of patients aged 8 to 10 years with a complete UCLP from files of the Orthodontic Department of the Faculty of Dentistry, Khon Kaen University (KKU), Thailand were available. The patients had completed all their primary surgery, with or without pre-surgical orthopedics, but no orthodontic or surgical treatment had been performed before study models were taken. Syndromic patients or patients with other congenital malformations were excluded. The present study was granted approval by the Institute Review Board (IRB) Committee at Khon Kaen University. A number in random order was assigned to each set of models by a non-examiner to ensure examiner blinding. Two examiners participated in the present study, one was an expert orthodontist who had attended the GOSLON Yardstick calibration course (examiner A); another examiner was an orthodontic postgraduate student but had been trained in the use of GOSLON (examiner B). Analysis of patient records is composed of 2 parts: the first part was patient's general characteristics (gender, age and cleft side); the second one was the study model assessment. Each model was independently rated according to the GOSLON and modified Huddart/Bodenham methods twice with a two-week interval between each assessment to minimize effects of memory bias on the results.

The intra-rater and inter-rater reliability of GOSLON Yardstick were evaluated by the weighted kappa statistic and the interpretation of kappa value was based on data according to Table 1[4]. The intra- and inter-examiner reliability of the modified Huddart/ Bodenham index were evaluated using Intraclass Correlation Coefficient (ICC)[5], its interpretation being similar to kappa. Moreover, the total time used in evaluation all models was recorded to compare the amount of time spent for each index and compared by paired t-test.

To attempt matching the modified Huddart-Bodenham score with the GOSLON categorization for each subject, the former scores were arbitrarily separated into 5 groups as in Table 2 and weighted kappa statistic was used to test agreement between the two methods.

Table 1. Interpretation of kappa values[4]

Kappa Value	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

Table 2. Categorization of the modified Huddurt/Bodenham scoring system into 5 groups

Group	Range of the Modified Huddart/Bodenham Score
1 (Excellent)	+2 to 0
2 (Good)	-1 to -5
3 (Fair)	-6 to -10
4 (Poor)	-11 to -16
5 (Very poor)	-17 to -22

Results

The reliability of the intra- and inter-examiner agreement

From Table 3, the intra-examiner agreement for the GOSLON Yardstick was very good agreement (kappa = 0.95 for both examiners). The inter-examiner reliability for the GOSLON Yardstick revealed very good agreement (kappa = 0.86 for both assessments).

Table 3 also shows very good reliability of intra-examiner agreement in using the modified Huddart/ Bodenham scoring system (ICC = 0.98 for both examiners). The inter-examiner agreements in evaluation with the modified Huddart/Bodenham scoring system were very good (ICC = 0.95 and 0.97 for the first and second measurements, respectively).

Table 3 also gives kappa values for the modified Huddart/Bodenham scoring for each subject which was transformed into one of 5 categories as for the GOSLON assessment. The intra-examiner agreements were very good for both examiners (kappa = 0.81 and 0.86 for examiners A and B, respectively). For the agreement between assessors, the first assessment showed good reliability between the two examiners (kappa = 0.77) whereas the second assessment presented very good reliability (kappa = 0.86).

The measurement of agreement between the modified Huddart/Bodenham scoring system and the GOSLON Yardstick

There was good agreement of the modified Huddart/Bodenham scoring system with the GOSLON Yardstick with weighted kappa of 0.73 and 95% confidence interval of 0.57 to 0.90 and the p-value less than 0.001. This indicates a good relationship between the modified Huddart/Bodenham scoring system and the GOSLON Yardstick.

The time taken for assessment with the modified Huddart/Bodenham scoring system and the GOSLON Yardstick

The time taken for evaluation with each index was compared using a paired t-test. Table 4 shows that the GOSLON assessment took significantly less time than the modified Huddart/Bodenham system.

Discussion

Reliability of the modified Huddart/Bodenham scoring system

In the present study, the modified Huddart/Bodenham scoring system was evaluated by the ICC value. The results showed that both assessors had very good intra-examiner reliability as well as inter-examiner reliability. The ICC values of the intra-examiner were 0.98 for both examiners, while the inter-examiner agreement was similarly high for both the first and second assessments.

Gray and Mossey[6] evaluated use of the modified Huddart/Bodenham system and concluded that the modified Huddart/Bodenham scoring system was a reliable indicator since the intra- and inter-rater agreements were both excellent.

Categorization of the modified Huddart/Bodenham scoring system

In the present study, it was necessary to transform the continuous scores of the modified Huddart/Bodenham system into 5 groups in order to compare with the 5 ordinal rankings of the GOSLON Yardstick. With the former system, there are only 3 possible scores (+2, +1 and 0) in the excellent outcome group (Group 1-Table 2) since these values represent cases with positive incisor overjet that are rare among operated cleft patients. The remaining possible scores were evenly split into 4 groups from "good" (scores from -1 to -5) to "very poor" (-17 to -22) representing arbitrary matchings with the remaining 4 GOSLON groups. As in the Eurocleft study, Group 1 subjects were found to be less than 5% in all centers[7]. Likewise, there was a report of only 2.4% of Group 1 subjects in a Malaysian cleft study[8].

Comparisons of the modified Huddart/Bodenham with and its use as an alternative to the GOSLON Yardstick

The modified Huddart/Bodenham scoring system and the GOSLON Yardstick have been developed as indicators of outcome of treatment for patients with UCLP. The present study demonstrated good agreement when the groupings of the modified Huddart/Bodenham scores were tested for agreement

Table 3. Intra- and Inter-examiner agreement of the GOSLON Yardstick and the modified Huddart/Bodenham scoring system

	Intra-examiner agreement		Inter-examiner agreement	
	A	B	1st measurement	2nd measurement
GOSLON Yardstick(Kappa value)	0.95	0.95	0.86	0.86
Modified Huddart/Bodenham (ICC)	0.98	0.98	0.95	0.97
Grouped modified Huddart/Bodenham(Kappa value)	0.81	0.86	0.77	0.86

Table 4. Analysis of time taken for assessment with each index

Time (sec)	Mean (SD)	Difference A-B (SD)	95% CI of Difference	p-value
GOSLON (A)	8.18 (3.50)	-13.84 (4.69)	-15.05 to -12.62	< 0.001
Huddart/Bodenham (B)	22.02 (3.70)			

with the weighted kappa statistics with the GOSLON groupings. A recent New Zealand study[9] reported a similar finding. Fig. 1 illustrates that in the present study, the percentage distributions of all the KKU cleft children among three categories[1] of good (Groups 1 and 2 requiring either straightforward orthodontic treatment or none at all), acceptable (Group 3, complex orthodontic treatment), and poor (Groups 4 and 5, orthognathic surgery to correct skeletal malrelationships) for both assessments were closely comparable. Likewise, Mossey et al[2] and Gray and Mossey[6] found good agreement between the modified Huddart/Bodenham system and GOSLON Yardstick. The authors of both these reports concluded that the modified Huddart/Bodenham scoring system was as reliable and capable of categorizing the models into similar groups as the GOSLON index but did not explain how they related their Huddart/Bodenham scores to the five GOSLON Yardstick categories.

The GOSLON Yardstick has the limitations of the need for a set of reference study models and a qualified assessor to carry out and publish results of using the method. Importantly, the GOSLON method seems to be more subjective than the modified Huddart/Bodenham scoring system which uses defined measures of severity of malpositions of teeth and arch segment discrepancies.

To achieve a modified Huddart/Bodenham score in each study model, occlusal relationship scores for all maxillary teeth up to and including first permanent molars are separately scored. In contrast to the limited diagnostic scope of the GOSLON assessment method, using this scoring system enables a quantitative and descriptive record of the positions of individual teeth and groups of teeth. This provides more specific information of the type of immediate as well as long-term orthodontic treatment needs, such as unilateral or bilateral maxillary arch expansion, while also predicting likely need for orthognathic surgery. Mars et al[1] admit that use of "a system along the lines proposed by Huddart and Bodenham[10]" could be advantageous in making more detailed discriminations of the possible range of occlusal outcomes among UCLP patients than is possible when using the GOSLON Yardstick.

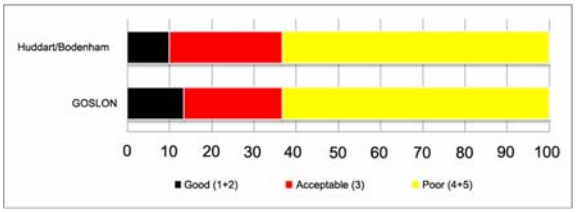

Fig. 1 Comparison of cumulative GOSLON score to the modified Huddart/Bodenham score of KKU sample

It is generally accepted that there is an ability to predict the future treatment requirements with the GOSLON Yardstick. The results of this study indicate that the modified Huddart/Bodenham scoring system is a good alternative GOSLON for diagnostic purposes since the use of weight kappa showed a good relation between both indices.

Comparison of the time taken in assessing of the study models using both methods

The time taken to assess the 60 sets of study models was significantly less using the GOSLON Yardstick (Table 4), thus favoring its use in large comparative surveys such as intercenter comparisons of treatment outcomes. However, the difference between the time taken to assess all these sets of models when comparing use of the GOSLON method (mean 8.18 $\pm$ 3.50 seconds) and the modified Huddart/Bodenham method (mean 22.02 $\pm$ 3.50 seconds) is, arguably, not such an important disadvantage for the latter method.

Conclusion

It is suggested that the GOSLON Yardstick and the modified Huddart/Bodenham scoring system provide equally acceptable means of assessing the primary surgical outcomes in UCLP patients of approximately 9 years of age. Although the GOSLON Yardstick provides a quicker assessment, the modified Huddart/Bodenham scoring system not only scores the incisor relationship as GOSLON does, but it also records the amount and nature of dental arch segment displacements and dental cross bites. Moreover, the GOSLON seems to reflect skeletal pattern rather than outcome following primary cleft repair because it relies on the incisal overjet whereas the modified Huddart/Bodenham scoring system assesses both incisor and buccal segments. The latter index may reflect more accurately arch constriction following surgery than does the GOSLON Yardstick.

Acknowledgement

The authors wish to thank Assoc. Prof. Keith Godfrey for his guidance and support, which enabled us to develop an understanding of the project. Also, the authors wish to thank the Center of Cleft Lip-Cleft

Palate and Craniofacial Deformities, Khon Kaen University in association with the "Tawanchai Project".

Potential conflicts of interest

None.

References

1. 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: 314-22.
2. Mossey PA, Clark JD, Gray D. Preliminary investigation of a modified Huddart/Bodenham scoring system for assessment of maxillary arch constriction in unilateral cleft lip and palate subjects. Eur J Orthod 2003; 25: 251-7.
3. Mars M. Independent blind audit-nothing is credible [letters to the editor]. Summer News Sheet. London: Craniofacial Society of Great Britain and Ireland; 1997.
4. Altman DG. Inter-rater agreement. In: Altman DG, editor. Practical statistics for medical research. London: Chapman and Hall; 1991: 403-9.
5. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979; 86: 420-8.
6. Gray D, Mossey PA. Evaluation of a modified Huddart/Bodenham scoring system for assessment of maxillary arch constriction in unilateral cleft lip and palate subjects. Eur J Orthod 2005; 27: 507-11.
7. Mars M, Asher-McDade C, Brattstrom V, Dahl E, McWilliam J, Molsted K, 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: 405-8.
8. Zreaqat M, Hassan R, Halim AS. Dentoalveolar relationships of Malay children with unilateral cleft lip and palate. Cleft Palate Craniofac J 2009; 46: 326-30.
9. Jack H, Antoun J, Fowler P. Comparison of primary surgical treatment outcomes of unilateral cleft lip and palate patients between two New Zealand cleft centres. Abstracts in 7th International Orthodontic Congress. February 6-9, 2010; Sydney, Australia; 2010.
10. Huddart AG, Bodenham RS. The evaluation of arch form and occlusion in unilateral cleft palate subjects. Cleft Palate J 1972; 9: 194-209.

การเปรียบเทียบวิธีโมดิฟายด์ฮัดดาร์ทโบเดนแฮมและดัชนีมาตรฐานกอสลอนในการประเมินการสบฟันภายหลังการผ่าตัดปฐมภูมิในผู้ป่วยปากแหว่งเพดานโหว่ข้างเดียว

มนเทียร มโนสุดประสิทธิ์, ทัศนีย์ วังศรีมงคล, สมศักดิ์ กิจสหวงศ์, ธนรัตน์ เธียรโกศล

วัตถุประสงค์: *เพื่อทดสอบความสอดคล้องระหว่างค่าระบบคะแนนโมดิฟายด์ฮัดดาร์ทโบเดนแฮม และดัชนีมาตรฐานกอสลอนในการประเมินการสบฟันผู้ป่วยปากแหว่ง และเพดานโหว่ข้างเดียว และเพื่อทดสอบเวลาที่ใช้ในการประเมินด้วยแต่ละวิธี*

วัสดุและวิธีการ: *ทำการประเมินแบบจำลองการสบฟันของผู้ป่วยอายุ 8-10 ปี ที่มีปากแหว่งเพดานโหว่ข้างเดียว จำนวน 60 คู่ ที่เข้ามารับการรักษาที่ภาควิชาทันตกรรมจัดฟัน คณะทันตแพทยศาสตร์ มหาวิทยาลัยขอนแก่น กลุ่มตัวอย่างทั้งหมดเคยได้รับการผ่าตัดซ่อมแซมปากแหว่งและเพดานโหว่มาแล้ว แต่ไม่เคยได้รับการปลูกถ่ายสันกระดูกเบ้าฟันหรือจัดฟันมาก่อน ใช้ผู้ทำการวัด 2 ราย ที่ได้รับการฝึกอบรมมาแล้วประเมินแบบจำลองการสบฟันแต่ละคู่โดยให้ค่าคะแนนโมดิฟายด์ฮัดดาร์ทโบเดนแฮม และคะแนนกอสลอนสำหรับค่าคะแนนโมดิฟายด์ฮัดดาร์ทโบเดนแฮมจะถูกจำแนกออกเป็น 5 กลุ่มเพื่อทดสอบ ความสอดคล้องระหว่างสองวิธีโดยใช้สถิติ weighted kappa และทำการเปรียบเทียบเวลาที่ใช้ในการประเมินของแต่ละดัชนีโดยใช้ paired t-test*

ผลการศึกษา: *พบว่าวิธีทั้งสองมีความสอดคล้องกันในระดับดีด้วยค่าแคปปา 0.73 และดัชนีมาตรฐานกอสลอนใช้เวลาในการประเมินน้อยกว่าวิธีโมดิฟายด์ฮัดดาร์ทอย่างมีนัยสำคัญ*

สรุป: *ระบบค่าคะแนนฮัดดาร์ทโบเดนแฮมสามารถใช้เป็นทางเลือกแทนดัชนีมาตรฐานกอสลอนที่ใช้กันโดยทั่วไปในการวินิจฉัยได้ ถึงแม้ว่าระบบค่าคะแนนจะใช้เวลาในการประเมินมากกว่าแต่สามารถให้ข้อมูลเกี่ยวกับการสบฟันที่ผิดปกติได้มากกว่าดัชนีกอสลอน*