

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

P. A. Mossey*, J. D. Clark* and D. Gray**

*Unit of Dental and Oral Health, Dundee Dental School and **Department of Orthodontics, Victoria Hospital, Kirkcaldy, UK

SUMMARY The objective of this study was to describe a numerical scoring system for the measurement of maxillary arch constriction in patients born with unilateral cleft lip and palate (UCLP). A modification of the Huddart/Bodenham scoring system was compared and contrasted with the current methods of measuring treatment outcome, the GOSLON and 5-year indices.

The GOSLON and 5-year indices are represented by 10 sets of study models grouped into five categories representing the range of possible outcomes in terms of dental arch relationship, with two sets of models in each of the five categories, whilst the modified Huddart/Bodenham method uses the frequency and severity of crossbite of the dental occlusion to evaluate maxillary arch constriction.

The latter system was found to be more objective and reliable, and to correlate well with current recommended standards. It was also more versatile and more sensitive to interarch discrepancies. However, further work is required to refine it to reflect the potential for orthodontic treatment to mask interarch discrepancy following surgery.

Introduction

Non-syndromic orofacial clefting is a heterogeneous group of disorders, with unilateral cleft lip and palate (UCLP) being most representative of the spectrum of surgical and non-surgical interventions. Management of UCLP requires a multidisciplinary approach, and involves primary surgical repair of both the lip (usually at around 3 months) and the palate (any time between 6 and 14 months). 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).

A wide range of surgical techniques are used for repairing cleft lip and cleft palate, but as yet there are no 'clear cut' protocols for timing or technique. A recent European study involving six centres demonstrated that it is possible to detect differences in outcome (Shaw *et al.*, 1992) and this has given impetus to the development of strategies to improve the quality of treatment. Furthermore, a precise and reproducible outcome measure is a necessity, at a time when evidence-based medical care and treatment guidelines regarding best practice are becoming an integral part of contemporary clinical practice.

Mars *et al.* (1987), introduced a standardized method described as the Great Ormond Street, London, and Oslo, Norway (GOSLON) ranking system for measuring treatment outcome based on an analysis of dental relationships using study models of children with UCLP, in the late mixed and/or early permanent dentition. In

1997 a similar index, the 5-year index (Atack *et al.*, 1997) was developed for 5-year-old children with UCLP. The authors were able to demonstrate good correlation between the two indices on longitudinal study models taken at 5 and 10 years of age in the same patient sample. Both the 5-year and GOSLON indices use five categories, from 1 (excellent) through 2 (good), 3 (fair), 4 (poor), and 5 (very poor). These categories are based on criteria that are used to describe the dental arch relationship.

The use of the GOSLON and 5-year indices require a degree of professional judgement with regard to the possibility of orthodontic correction, which introduces an element of subjectivity. Also reference models have to be used for comparison (10 models for the GOSLON and 10 for the 5-year index) and a calibration course is necessary for competent use. An alternative system for measuring upper arch constriction in the primary dentition of patients with repaired cleft lip and palate (CLP) was described by Huddart and Bodenham (1972). This system uses the frequency and severity of crossbite of the dental occlusion to evaluate maxillary arch constriction in the labial segment and the greater (non-cleft) and lesser (cleft) buccal segments. The buccal segments comprise the canine and primary molars, and the labial segment only the central incisors (Figure 1). Each maxillary tooth is scored according to its relationship with the corresponding tooth in the mandible (Figure 2). Individual scores are summated to give a total score for each set of models.

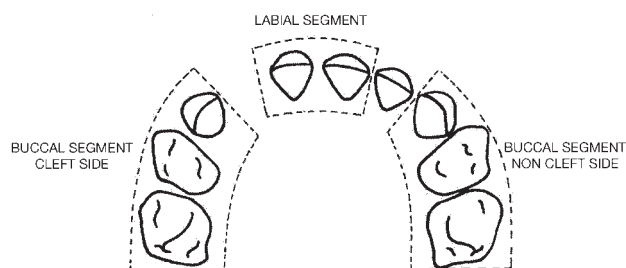


Figure 1 Segmental divisions of the maxillary arch as used in the Huddart/Bodenham scoring system (adapted from Huddart and Bodenham, 1972).

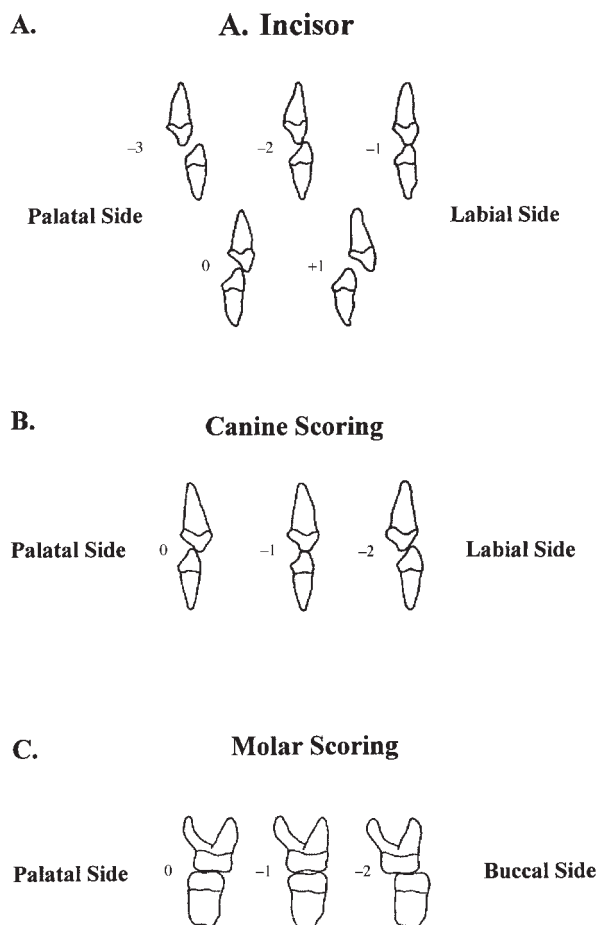


Figure 2 Occlusal scoring of the maxillary arch as used in the Huddart/Bodenham scoring system: scoring of (A) the anteroposterior relationship of the central incisors; (B) the buccopalatal relationship of the canines; (C) the buccopalatal relationship of the molars (adapted from Huddart and Bodenham, 1972).

The GOSLON and 5-year indices reference models formed the basis of this study. The models used were carefully chosen to be representative of the five categories in their respective scoring systems. They were therefore ideal to validate the modified Huddart/Bodenham system. Validation should be carried out against an accepted reference and to use a sample where the

GOSLON or 5-year index scores might be ambiguous would diminish the validity of the study.

The aims of this study were to:

1. modify the Huddart/Bodenham scoring system so that it can be used at any stage of the developing occlusion.
2. apply the modified scoring system to the GOSLON and 5-year index reference models and identify the inter- and intra-examiner variability.
3. determine whether the modified Huddart/Bodenham scoring system sufficiently discriminates between the reference models to rank them in their categories.

Materials and methods

Materials

The 10 GOSLON and 10 5-year index reference models, two from each category of both indices were used. All models were of patients with UCLP, aged 10 and 5 years, respectively.

Examiners

Three examiners scored the models independently using the modified Huddart/Bodenham system. Two examiners were calibrated in the GOSLON and 5-year indices but the third, although on an orthodontic training programme, had no experience of cleft treatment outcome measures. Scoring forms suitable for use with both the GOSLON and 5-year index reference models were designed and distributed to the examiners. The same examiners repeated the scoring under similar conditions, one month later.

Modifications to the scoring system

As the Huddart/Bodenham system was designed for use in the primary dentition it was necessary to modify it for use in the mixed dentition. This was undertaken by scoring premolars in the same way as primary molars, i.e. normal occlusion scored 0, cusp to cusp scored -1, and buccal crossbite scored -2.

The modified system requires that all teeth from first permanent molar forward be given a score to reflect the maxillary arch constriction. Therefore, in all cases where there was an absent or unerupted tooth, the score was determined by the midpoint of the maxillary alveolar ridge at the location of the missing tooth.

Statistical analysis

Intra- and inter-examiner variability was calculated using Cohen's weighted and unweighted Kappa statistic (Altman, 1991). The unweighted Kappa (κ) statistic for intra- and inter-examiner variability detects correlation

between ordinate variables, i.e. it matches exact scores and if they do not match they are not correlated. The weighted κ statistic on the other hand takes into account the degree of mismatch, i.e. the seriousness of the discrepancy in the scoring of the same model between examiners. Altman (1991) suggested that a κ value greater than 0.8 indicates good agreement, greater than 0.6 indicates substantial agreement, and greater than 0.4 moderate agreement. The modified Huddart/Bodenham scores were correlated with the 5-year index and GOSLON scores using the Spearman rank correlation coefficient.

Results

Intra-rater reliability

Table 1 displays the inter-rater reliability mean kappa scores. The first and second ratings of each examiner were compared, and a mean value calculated for the incisors, canines, and molars.

Modified Huddart/Bodenham scores: 5-year index reference models

The intra-rater reliability mean weighted κ scores were 0.74 for the incisors, 0.95 for the canines, and 0.89 for the molars. The mean unweighted κ statistic for the incisors (0.49) was less impressive. This can be explained by the use of a five-point scale for the incisors, and the difficulty in being certain about the incisor relationship in the mixed dentition when the incisors are worn. By comparison there is a three-point scale for the canines and molars with less scope for ambiguity.

Modified Huddart/Bodenham scores: GOSLON reference models

The mean weighted κ scores for the GOSLON models were 0.91 for the incisors, 0.91 for the canines, and 0.88 for the molars. The mean unweighted κ statistic also showed high intra-rater reliability, 0.78 for the incisors, 0.86 for the canines, and 0.79 for the molars. The more definite incisor occlusion at the age of 10–12 years

explains the improved reliability in the scoring of incisors in this age group.

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).

Modified Huddart/Bodenham scores: 5-year index/GOSLON reference models

For all three examiner pairs using the 5-year index models, the weighted κ statistic was above 0.8 and unweighted κ above 0.7, indicating high inter-rater reliability (Table 2). For the GOSLON models the inter-rater reliability was above 0.8 for both unweighted and weighted κ , which can be explained by the more definite occlusion present in the permanent dentition (Table 2).

Correlation between modified Huddart/Bodenham and GOSLON/5-year index

For each model, a mean score was calculated for each examiner over both scoring episodes. The correlation between the mean Huddart/Bodenham and the 5-year and GOSLON scores was assessed using the Spearman rank correlation coefficient. The results indicated a statistically significant correlation in all cases ($P < 0.01$), consistently greater than 0.8 for all three examiners (range 0.83–0.92).

The relationship between the 5-year index categories, the GOSLON index categories, and the mean Huddart/Bodenham scores is illustrated in Figures 3 and 4. There appears to be a definite linear association for all 5-year models, and all GOSLON cases except for 1j and 2ac, which appear worse than the GOSLON score suggests.

The modified Huddart/Bodenham system is therefore just as reliable and capable of categorizing the models into similar categories as the GOSLON and 5-year indices. The scores of the examiner who had no

Table 1 Weighted and unweighted mean Kappa scores for intra-examiner reliability.

	5-year models		GOSLON models	
	Unweighted Kappa	Weighted Kappa	Unweighted Kappa	Weighted Kappa
Incisors	0.485	0.739	0.775	0.909
Canines	0.937	0.952	0.859	0.912
Molars	0.838	0.89	0.79	0.879

Table 2 Weighted and unweighted mean kappa scores for inter-examiner reliability.

	5-year models		GOSLON models	
	Unweighted Kappa	Weighted Kappa	Unweighted Kappa	Weighted Kappa
Examiner A–examiner B	0.747	0.847	0.826	0.896
Examiner A–examiner C	0.892	0.904	0.864	0.917
Examiner B–examiner C	0.722	0.832	0.876	0.921

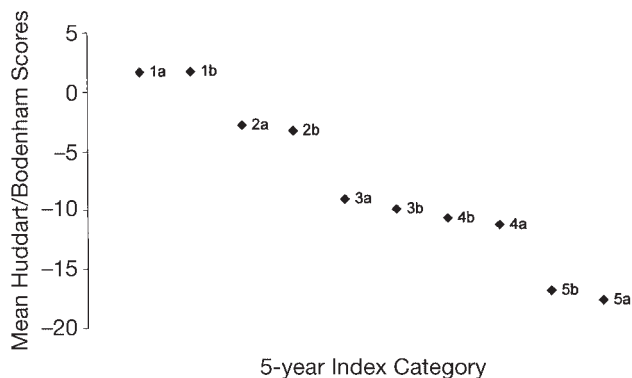


Figure 3 Relationship between 5-year index categories and mean Huddart/Bodenham scores.

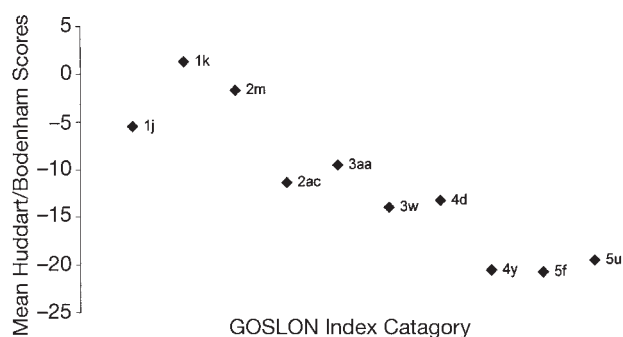


Figure 4 Relationship between GOSLON index categories and mean Huddart/Bodenham scores.

previous exposure to cleft care were comparable to the others and in fact achieved a correlation score of 0.91 for the comparison between the modified Huddart/Bodenham scores and 5-year index. This suggests a high level of reliability.

Discussion

In some countries, including the UK, the standard of orthodontic care is perceived to be below that of the best European centres, and the problem of measuring surgical outcome is compounded by surgeons who perform few operations each year (Williams *et al.*, 1994). Shaw *et al.* (1992) identified 175 cleft teams, and almost as many different cleft treatment protocols. It is therefore extremely important that valid and reliable tools are used to measure treatment outcomes for research and audit.

A recent report from the Clinical Standards Advisory Group (1998) suggested a minimum caseload of 30 new patients annually for primary CLP surgery. 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. The GOSLON and 5-year indices are regarded as the best available objective measures for measuring treatment outcome, but with a 5-point scale an annual case load of some 60 patients over a period of 8.5 years is required in order to detect a difference of 0.5 at 5 per cent probability and with 80 per cent power.

Conceptual differences between the three scoring systems

Unlike the Huddart/Bodenham system, the GOSLON and 5-year index scoring systems take into account the potential for orthodontic management to mask any interarch discrepancy following surgery. This is illustrated by cases 1j and 2ac (Figure 5), which score low (1 and 2, respectively) on the GOSLON index and high (-6 and -12) on the modified Huddart/Bodenham index. Although the Huddart/Bodenham scores more

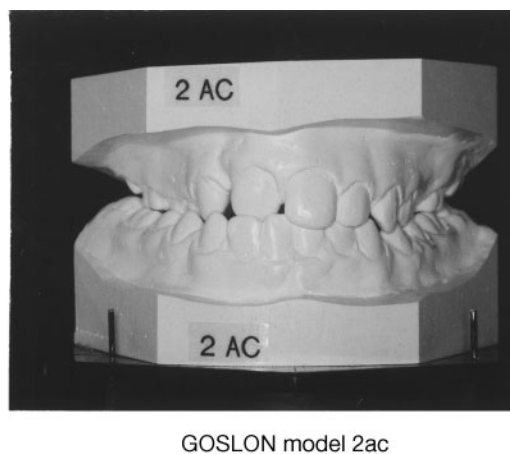
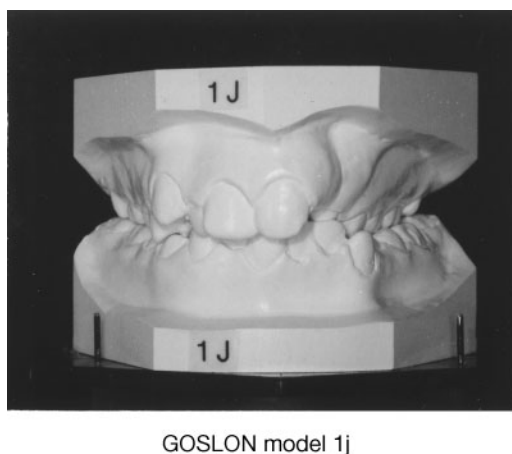


Figure 5 Frontal views: GOSLON reference models 1j and 2ac, which the modified Huddart/Bodenham system places in a more severe category.

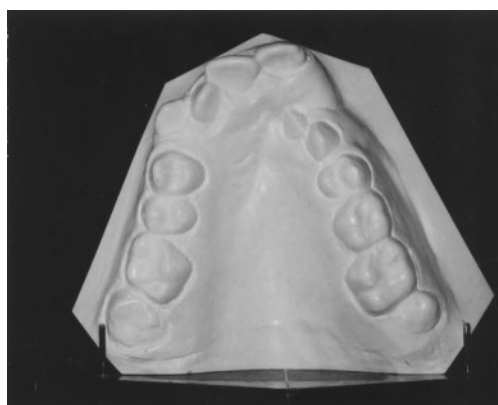
accurately reflect the extent of interarch discrepancy, they do not reflect the fact that both occlusions are within the scope of orthodontic correction—an important consideration when assessing the outcome of surgery.

For measurement of surgical outcome, it is desirable that the score represents the degree of iatrogenic damage rather than skeletal pattern. It could be argued that a disadvantage of the GOSLON score is that it reflects a favourable skeletal pattern rather than a good outcome following surgery. This is demonstrated by model 1k (Figure 6), which is attributed to GOSLON category 1 (excellent). A comparison of the occlusal views of the upper arches of 1j and 1k, both GOSLON category 1 (Figure 7), suggests that with regard to upper arch constriction following surgery, 1j is worse than 1k—a fact that is accurately reflected by the Huddart/Bodenham score.



GOSLON model 1k

Figure 6 Left lateral view: GOSLON reference model 1k, which attracts a good score (1) despite the crossbite tendency of the left buccal segments.



GOSLON model 1j



GOSLON model 1k

Figure 7 Occlusal views of upper arches: GOSLON reference models 1j and 1k. GOSLON places both models in category 1 (excellent), while the modified Huddart/Bodenham system reflects the greater arch constriction in model 1j.

Further development of the Huddart/Bodenham scoring system, using a different weighting factor for incisors and buccal segments, is being undertaken. This will allow it to be used simultaneously as an absolute measure of arch constriction and also a measure that reflects the potential for orthodontic correction.

Disadvantages of the GOSLON and 5-year index scoring systems

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. Mars *et al.* (1987) admitted that 'in some cases precise allocation to a GOSLON category may be ambiguous'. It was also apparent at the calibration courses attended by two of the authors (JC and PM) that when using the GOSLON and 5-year index models there were occasionally two category differences in successive scores on the same models, both in the intra- and inter-examiner reliability tests. Bearing in mind that it is mainly experienced clinicians that are involved, a two category difference in a 5-point scale for the same model may be regarded as an unacceptable level of error. Since the modified Huddart/Bodenham scoring system uses a cumulative score, which is derived from six or eight separate categorical assessments, the effect of random operator error is minimized. A significant error in the overall score would require multiple errors in the same direction.

Calibration courses are a prerequisite to using the GOSLON and 5-year indices and even with regular use recalibration is necessary to ensure accuracy. Reference models must always be incorporated into samples that are analysed, adding to the complexity of the exercise. It would be expected that intra- and

inter-rater reliability with the use of the GOSLON index would tend to improve with operator exposure, but ideally an index should not be dependent on increased experience or expertise. In the modified Huddart/Bodenham system, clinical judgement and experience is not required, making it more objective and reliable.

Advantages of the modified Huddart/Bodenham scoring system

Objectivity combined with relative simplicity. No clinical experience is required and therefore non-professional auxiliary staff such as laboratory technicians can accurately score the models. This simplifies the training of assessors and will improve inter-centre collaborative studies. A score can be calculated very quickly, consistently within a minute, and usually in less than 30 seconds.

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. The Eurocleft Steering Group on minimum records for measurement of outcome recommends 5/6 years (Shaw *et al.*, 2000), whilst the International Committee on Cleft Documentation and Measurements (Fong *et al.*, 1998) recommends 5 years. Neither the GOSLON nor the 5-year indices could be used outwith the 10- or 5-year age group, respectively, or for anything but UCLP without revalidation.

Sensitivity. The scale is a continuous scale of severity of arch constriction rather than a categorical scale and therefore provides a greater degree of sensitivity and the ability to differentiate the severity within the categories that would be identified by the GOSLON or 5-year indices. As a continuous numerical scale it is also quantifiable, thus lending itself to statistical analysis. The modified system is, however, an ordinal and not an 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 overjet and buccal crossbite could all be recorded from scanned digital images. A computer programme could be developed that would enable the assimilation of all the data to provide a score. A system that incorporates a subjective element does not allow this computer-based approach.

Conclusions

The modified Huddart/Bodenham scoring system appears to be a valid and reliable indicator of treatment outcome for patients with UCLP. The scoring system can be used for 5- and 10-year-old models, or for any age from 3 years upwards. The correlation and intra- and inter-examiner agreement with all except one 5-year index and two GOSLON cases is excellent. The reasons for the discrepancy in these cases reflect the fact that the Huddart/Bodenham system measures arch constriction (possibly a more accurate reflection of surgical outcome) rather than the potential for orthodontic correction, which is influenced not only by surgical outcome but also by the inherited skeletal pattern. The system is a viable alternative to the 5-year index and GOSLON scoring systems, with major advantages being objectivity, flexibility, and simplicity.

Address for correspondence

Dr Peter Mossey
Unit of Dental & Oral Health
Dundee Dental School
1 Park Place
Dundee DD1 4HR, UK

Acknowledgements

The authors would like to thank Mr Grant McLennan, Glasgow Dental Hospital and School for his advice and support in this study. The tuition received by two of the authors on the GOSLON and 5-year index calibration courses is also acknowledged, as is the co-operation of Dr Mike Mars, Great Ormond Street Hospital, London, for his permission to allow the GOSLON models shown in Figures 5, 6 and 7 to be used as illustrations in this article.

References

- Altman D G 1991 Practical statistics for medical research. Chapman & Hall, London, pp. 403–409
- 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 Standards Advisory Group (CSAG) report: Cleft lip and/or palate 1998 Royal College of Surgeons of England, ISBN 0 11 322103 7
- Fong P H, Ngim R C, Lee S T 1988 Measuring the cleft deformity. *Annals of Academic Medicine Singapore* 17: 343–351
- 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-centre 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
- Shaw W C *et al.* 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 Livingstone, Edinburgh, pp. 1–5
- Shaw W C *et al.* 2000 The EUROCLEFT Project 1996–2000. Standards of care for cleft lip and palate in Europe. In: Shaw W C, Semb G, Nelson P, Brattstrom V, Molsted K, Prah-Andersen B (eds) *Biomedical and Health Research Volume 43*. IOS Press, Amsterdam, p. 22
- Williams A, Shaw W C, Devlin H B 1994 Provision of services for cleft lip and palate in England and Wales. *British Medical Journal* 310: 1406 (letter)
- 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.