Assessing Malocclusion—the Time Factor

D. J. TARVIT, M.D.SC.

T. J. FREER, B.D.SC., F.D.S.R.C.S. (ENG.), D.ORTH. R.C.S., PH.D.

Department of Dentistry, University of Queensland Dental School, Brisbane, Australia

Abstract: *Assessment of occlusal status for orthodontic purposes may be undertaken by subjective assessment or by means of an index of occlusion. The assessment method should be reproducible and should achieve the intended outcomes. This study investigated another possible modifying factor— performance over time. Photographs and dental casts of 45 subjects at 12 years of age were compared with follow-up records at 16 years of age. No orthodontic intervention had occurred dur - ing the 4 years. Subjective severity assessments and Aesthetic Component (AC) scores of the Index of Treatment Need (IOTN) for first and second series records were made by four orthodontists. Scores of the Dental Aesthetic Index (DAI) and Dental Health Component (DHC) of the IOTN were made by one of the authors. Implications for hypothetical deci - sions to treat were considered.*

There was a significant reduction in the subjective severity and DAI scores across the whole sample and in the mixed dentition subgroup. There was a non-significant reduction in the AC (IOTN) scores across the whole sample, but there was a significant reduction in the AC (IOTN) scores in the mixed dentition subgroup. The DHC (IOTN) categorisations proved to be more stable over the period of study.

Index words: Assessment Over Time, Dental Aesthetic Index (DAI), Index of Orthodontic Treatment Need (IOTN), Occlusal Indices, Subjective Orthodontic Treatment Need.

Refereed Paper

Introduction

The influence of time upon the accuracy of decisions to provide orthodontic treatment has not been widely reported. Although errors such as over-estimating the need for correction in the 'ugly duckling stage' are well known, subjective clinical decisions in the young patient are not always clear-cut, and there is a need to distinguish between developmental features and persistent traits of malocclusion. The ability to discriminate has not often been tested though Summers (1971) recognized the importance of not biasing the Occlusal Index by unduly weighting the features of normal occlusal development.

Gray and Demirjian (1977) claimed that the most reliable assessments of malocclusion are made in the permanent dentition stage. The designer of an occlusal index may reduce the problems of assessment associated with dental development by restricting its usage to the permanent dentition. In practice many definitive recommendations for treatment are made before completion of the mixed dentition phase.

This study investigated the time related changes in the scores of two indices, the Dental Aesthetic Index (DAI) and the Index of Orthodontic Treatment Need (IOTN), as well as the subjective assessments of four orthodontists using the longitudinal records of an untreated sample of 45 subjects.

Subjects and methods

Dental casts and photographs (facial, profile, and anterior dental) of 105, 12-year-old school children were obtained approximately 4 years prior to this study (Keay *et al.*, 1993). Permission to obtain follow-up records was sought. Subjects who had undergone orthodontic treatment, extractions, or guidance were excluded. Forty-five subjects provided follow-up records (Table 1). Mean age at the time of the initial records was 12 years 3·2 months (± 6·4 months). There was a mean elapsed time of 46.6 months (± 1·8) between the Series 1 and 2 records. The sample included 29 males and 16 females, and 20 subjects were Class I, 21 were Class II division 1, one was Class II division 2, and three were Class III. One subject was of Pacific Island descent and the remainder were Caucasian.

Series 1 and 2 records were combined, and then divided into two groups. The Series 1 and 2 records of the same person were not assigned to the same group. Duplicate sets of records of another 20 persons had been obtained previously and were assigned to the two groups giving a total of 65 in each group. Reliability of measurement was tested using the data from the duplicate records of the extra 20 subjects.

Four orthodontists with previous training and experience in the application of the methods were requested to participate in the study. Each independently examined the two mixed groups, providing a subjective severity assessment of each case. Dental casts and photographs were supplied in order to simulate a typical screening examination. Subjective severity was scored on a 0–10 scale where zero represented classically normal occlusion, 1, 2, and

TABLE 1 *Distribution by sex and dentition type in the original Series 1 subjects*

	Mixed dentition	Permanent dentition	Total
Male	19	10	29
Female	10	6	16
Total	29	16	

0301-228X/98/001000+00$02.00

3 indicated little need for correction, 4, 5, and 6 where correction was considered desirable, but elective, and 7, 8, and 9 where correction was highly desirable. A score of 10 was suggested where treatment was considered essential. Aetiology, treatment cost, and difficulty of treatment were not to be considered. Radiographs were not provided.

On a separate occasion the same four orthodontists provided an aesthetic assessment using the Aesthetic Component (AC) instrument. Dental casts were rated against a monochrome photographic scale as recommended (Richmond *et al.*, 1992). Facial photographs were removed for the assessment of the AC (IOTN).

One of the authors (DT), after training in the methods of the DAI (Cons *et al.*, 1986; Jenny and Cons, 1995) and the IOTN, separately calculated DAI and IOTN Dental Health Component (DHC) scores for both series of dental casts. After 1 month, the indices were recalculated by the same examiner on a random sample of 20 cases to test reproducibility of measurement.

Results

For each of the four orthodontists, weighted kappa statistics (Fleiss and Cohen, 1973) were calculated from the 20 duplicate sets of data for both subjective severity and AC (IOTN) scores and indicated that the levels of reproducibility were acceptable (Table 2). The Wilcoxon matched pairs test was applied to the ordinal data and revealed no significant difference at the 5 per cent level between replicate subjective severity, and AC (IOTN) scores for any of the orthodontists. No significant difference was found between replicate DAI scores of the other examiner (DT). The Spearman rank order correlation was 0·97. DHC (IOTN) scores were found to be 100 per cent reliable.

Agreement between the four orthodontists for subjective severity and AC (IOTN) was calculated on the scores of all 90 sets of records. Friedman analysis of variance by ranks revealed, over both series of records, a tendency for Examiner 1 to assign lower scores for subjective severity, while Examiner 3 assigned higher scores. Examiner 1 also provided lower ratings of the AC (IOTN). The scores of all examiners were used to derive mean subjective assessment and AC (IOTN) scores for the sets of Series 1 and 2 records.

Correlations between the Series 1 and 2 scores of the four methods (Table 3) show that in general the Series 2 scores were more highly correlated although the AC (IOTN) and subjective severity were highly correlated for both series (0.83 and 0.87). DAI and subjective severity were moderately correlated for both series (0·63 and 0·65). DAI and DHC (IOTN) were less well correlated for both series. Correlation between AC (IOTN) and DHC (IOTN) for Series 1 was low (0·48), but improved for Series 2 (0·68).

The orthodontists' mean subjective severity scores revealed highly significant reductions with time ($P < 0{\cdot}01$). Reduction of at least one grade was found in 24 cases (53 per cent). The greatest reduction was four grades and the average was 2·1. The magnitude of the changes in scores over time was not well correlated with the initial severity assessment (Spearman rank order correlation = $-0{\cdot}256$, $P = 0.09$). Interestingly, each orthodontist registered no significant change in severity over time for the permanent dentition subjects, whereas reduction in severity for the mixed dentition subjects was highly significant ($P < 0{\cdot}01$).

The orthodontists' mean assessment scores of AC (IOTN) reduced significantly in the mixed dentition subgroup over the period of the study ($P < 0{\cdot}05$). However, the mean AC (IOTN) scores for Series 1 and 2 showed a non-significant tendency towards reduction over the total sample.

There was no statistically significant difference between the DHC (IOTN) scores of the Series 1 and 2 records. Scores remained unchanged in 35 cases (78 per cent) and showed increased scores in two cases (4 per cent). The remaining eight subjects (18 per cent) demonstrated reductions of one grade and of these, six achieved the lower grade through reduction in overjet. Neither was there any significant change in the mixed dentition subgroup over the period of study.

DAI scores for Series 1 and 2 casts were found to have significantly reduced over time (Wilcoxon matched pairs, two-tailed, $P < 0{\cdot}01$). The greatest reduction was 17 points for one case and the average reduction was 5·1 points. Changes in DAI scores were only weakly correlated with the initial scores (Spearman rank order correlation = 0·31, $P = 0{\cdot}04$). Changes in the scores for each component of the DAI were examined to identify the trait or traits which contributed to reduced scores during time. The overjet score reduced in 58 per cent of the sample (Fig. 1). There was a slight tendency for incisal segment crowding to increase and for spacing to decrease including a decrease in the size of diastemata. Mandibular irregularity tended to increase. In 12 cases the molar relationship score improved by a half-cusp position. Nine of these were first series mixed dentition cases. There was a significant reduction in scores in the mixed dentition subgroup ($P < 0{\cdot}01$).

The effects of changing scores upon hypothetical, initial treatment decisions were investigated. The four sets of assessment scores (subjective, AC, DHC, DAI) for Series 1 and Series 2 were tabulated to show the overall rates of 'no treatment' and 'treatment' (No, Yes; Table 4). The decision boundary used in each case was that recommended by the originators to indicate a definite need for treatment. In addition to the overall numbers of

TABLE 2 *Intra-examiner agreement on subjective severity and AC (IOTN) (Kappa statistic, quadratic weights). Ninety-five per cent confidence limits in brackets*

	Examiner 1	Examiner 2	Examiner 3	Examiner 4
Severity	0·78 (0·56–1·0)	0·86 (0·74–0·97)	0·74 (0·63–0·85)	0·65 (0·43–0·87)
AC (IOTN)	0·57 (0·24–0·90)	0·87 (0·80–0·95)	0·78 (0·56–1·0)	0·90 (0·84–0·96)

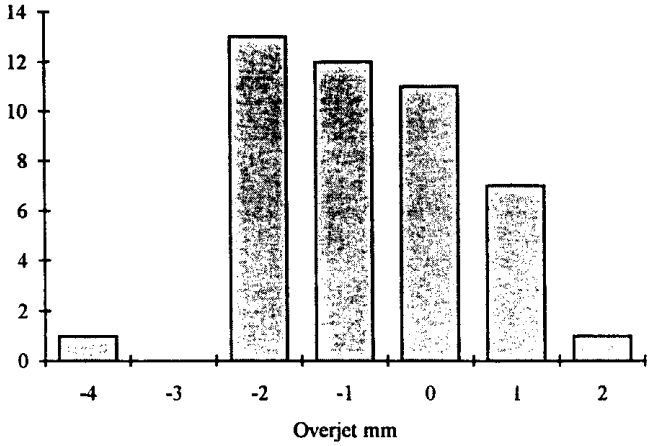

FIG. 1 Distribution of changes in overjet between the Series 1 and 2 records. Negative values indicate reduction.

TABLE 3 *Spearman rank order correlations between methods of assessment for Series 1 and 2 records*

		Subjective Severity	DAI	DHC (IOTN)
DAI	1st	0·63		
	2nd	0·65		
DHC	1st	0·52	0·40	
(IOTN)	2nd	0·77	0·39	
AC	1st	0·83	0·64	0·48
(IOTN)	2nd	0·87	0·71	0·68

treatment reversals between Series 1 and 2, we were interested to determine the number of reversals from No to Yes, as well as from Yes to No. There was a greater number of reversals from Yes to No than from No to Yes for the subjective scores, DHC (IOTN) and DAI. The AC (IOTN) showed an equal number of reversals in both directions.

Discussion

The reduction of scores from three methods of assessment in this study, particularly in the mixed dentition subgroup, is of more than academic interest because it may signify a level of inaccuracy in the original decision to treat which may have clinical implications. It should be noted that small changes in cut-off points and small changes in scores over time may reverse the inclusion or exclusion of individuals whose scores fall near thresholds. Because of this, setting cut-off values in order to include or exclude individual patients should be reviewed frequently. When thresholds are set and applied within a public health programme, inappropriate cases may be placed in the treatment group. Recommendations to treat resulting from failure of the assessment method to consider developmental factors are likely to expose additional patients to unnecessary treatment (Shaw *et al.*, 1991).

The higher proportion of changes from Yes to No based on the orthodontists' subjective assessments suggests a need to consider the time factor in the stability of subjective severity assessments. The ratings of the orthodontists also revealed an important contrast in performance between the mixed dentition and permanent dentition groups. Assessments made in the permanent dentition were more consistent over time whereas there was significant reduction in scores for the mixed dentition group and 90 per cent of the reversed decisions to treat were made in mixed dentition cases. This not only supports the view that better assessments of malocclusion can be made in the permanent dentition, but may hold implications for the validity of decisions to embark on so called 'early treatment'.

AC (IOTN) demonstrated significant reductions in scores in the mixed dentition subgroup indicating a tendency to over-emphasise developmental features in mixed dentition cases. It was highly correlated with the orthodontists' subjective assessments in both series. It should be noted that the AC (IOTN) is a complementary assessment to DHC (IOTN) and is intended to identify those patients for whom the psychological importance of aesthetic impairment may be an overriding consideration. AC (IOTN) demonstrated the lowest number of reversals over the period of study and was the only method not to show a difference between the number of Yes–No and No–Yes reversals.

Assessments using the DHC (IOTN) were relatively stable over time. There was no statistically significant difference between the DHC (IOTN) scores of the Series 1 and 2 records. However, there were seven reversals (16 per cent) in the individual treatment need categories. Resembling a classification rather than an index, the DHC (IOTN) makes no attempt to quantify all features of the dental arrangement and deliberately excludes all but the worst occlusal feature. The DHC (IOTN) does not discriminate between individual cases within grades. This 'broad brush' approach provides a stability that is less easily achieved by a summative index. For example, a reduction in overjet of 2 mm may well be 'absorbed' within IOTN grades of 4 or 5 (treatment indicated), whereas the reduction of 2 mm, weighted by a factor of 2, will reduce

TABLE 4 *Reversals in the Series 1 and 2 treatment recommendations for each method of assessment using the threshold recommended by the originators of the method*

	Series 1		Series 2		Changes		%
	No	Yes	No	Yes	No to Yes	Yes to No	Changed
Subjective score ⩾ 6·5	26	19	36	9	1	11	27
AC (IOTN) ⩾ 7·5	39	6	39	6	3	3	13
DHC (IOTN) ⩾ 4	15	30	18	27	2	5	16
DAI ⩾ 30·5	13	32	22	23	2	11	29

DAI by 4 points. The DAI score may be further reduced if there has been a net improvement in other components.

While it cannot be claimed that the orthodontist's subjective scores represent an unchallengeable benchmark, correlation between the DHC (IOTN) and the subjective opinions of the orthodontic panel was substantially stronger for the 16-year-old sample than for the 12-year-old sample. Based on the premise that the orthodontists' assessments at 16 years, rather than at 12 years, more accurately reflect treatment need, this correlation tends to support the use of the DHC (IOTN) as an assessment tool at an earlier age. However, the relative consistency of the DHC (IOTN) over time may be achieved as a result of over-estimation.

Decreasing DAI scores over time occurred in this longitudinal study and in a cross-sectional study over the same age range (Estioko *et al.*, 1994). It would appear that the DAI, in weighting overjet by a factor of 2, constitutes a weakness if over-sensitivity to developmental symptoms is to be avoided. This study and others have demonstrated reductions in overjet during adolescence (Bjork, 1953), though increasing overjet may be expected in some cases between 6 and 12 years of age (Frölich, 1962). Although overjet may be an important predictor of treatment need (Freer, 1973), and even though weighting for the overjet component of the DAI index was established only after validation against professional and lay opinion, the possibility of a degree of improvement over time appears not to have been considered. By summing a number of components, the DAI is susceptible to a cumulative reduction in scores due to components which show improvement such as molar relationship and spacing. In comparison to the orthodontist's subjective scores, the DAI overestimates the numbers in the treatment group.

Conclusion

The consistency of the reversals in treatment recommendations recorded by three separate methods of assessment suggests that, in a significant percentage of subjects, there is a real change in the occlusal pattern with time and that during adolescence, certain features of malocclusion, including overjet, may improve enough to reduce the recommendation for treatment. In the mixed dentition stage, when many treatment decisions are made, some planned treatment may be unjustified. This study suggests that, during adolescence, certain features of malocclusion, including overjet, can change sufficiently to reduce recommendations for treatment. It is also suggested that both objective and subjective assessment methods may overemphasise the need for treatment at an early age. These interpretations should be tempered by the need to verify these results on similar samples using the same or similar methods.

References

Bjork, A. (1953)
Variability and age changes in overjet and overbite,
American Journal of Orthodontics, **39**, 779–801

Cons, N. C., Jenny, J. and Kohout, F.J. (1986)
DAI: The Dental Aesthetic Index,
Iowa City, College of Dentistry, University of Iowa.

Estioko, L. J., Wright, F. A. C. and Morgan, M. V. (1994)
Orthodontic treatment need of secondary schoolchildren in Heidleberg, Victoria: an epidemiological study using the Dental Aesthetic Index,
Community Dental Health, **11**, 147–151.

Fleiss, J. L. and Cohen, J. (1973)
The equivalence of weighted Kappa and the intra-class correlation coefficient as measures of reliability,
Educational and Psychological Measurement, **33**, 613–619.

Freer, T. J. (1973)
Selection of predictor variables in assessing the severity of malocclusion,
American Journal of Orthodontics, **64**, 155–161.

Frölich, F. J. (1962)
Changes in untreated Class II type malocclusions,
Angle Orthodontist, **32**, 167–179.

Gray, A. S. and Demirjian, A. (1977)
Indexing occlusions for public health programs,
American Journal of Orthodontics, **72**, 191–197.

Jenny, J. and Cons, N. C. (1995)
Establishing malocclusion severity levels on the Dental Aesthetic (DAI) scale,
Australian Dental Journal, **41**, 43–46

Keay, P. A., Freer, T. J. and Basford, K. E. (1993)
Orthodontic treatment need and the Dental Aesthetic Index,
Australian Orthodontic Journal, **13**, 4–7.

Richmond, S., O'Brien, K., Buchanan, I. and Burden, D. J. (1992)
An Introduction to Occlusal Indices,
University of Manchester, Victoria.

Shaw, W. C., O'Brien, K. D., Richmond, S. and Brook, P. (1991)
Quality control in orthodontics: risk/benefit considerations,
British Dental Journal, **170**, 33–37.

Summers, C. J. (1971)
The occlusal index: a system for identifying and scoring occlusal disorders,
American Journal of Orthodontics, **59**, 552–567.