

# The PAR Index (Peer Assessment Rating): methods to determine outcome of orthodontic treatment in terms of improvement and standards

S. Richmond, W. C. Shaw, C. T. Roberts, and M. Andrews

Department of Oral Health and Development, University of Manchester, UK

**SUMMARY** In orthodontics it is important to objectively assess whether a worthwhile improvement has been achieved in terms of overall alignment and occlusion for an individual patient or the greater proportion of a practitioner's caseload. An objective measure is described that has been validated against the subjective opinions of 74 dentists. Using the weighted PAR Index it was revealed that at least a 30 per cent reduction in PAR score is required for a case to be considered as 'improved' and a change of 22 PAR points to bring about 'great improvement'. For a practitioner to demonstrate high standards the proportion of an individual's case load falling in the 'worse or no different' category should be negligible and the mean reduction should be as high as possible (e.g. greater than 70 per cent). If the mean percentage reduction in PAR score is high and the proportion of cases that have been 'greatly improved' is also high, this indicates that the practitioner is treating a great proportion of cases with a clear need for treatment to a high standard.

## Introduction

A previous paper has described how the PAR Index score is derived (Richmond *et al.*, 1992); this article will describe how the PAR Index can be used in assessing improvement and the standard of treatment.

It is unrealistic to expect all malocclusions to be treated to an ideal occlusion. The outcome of treatment is often dependent on many factors, e.g. complexity of the case and the expertise of the practitioner. It is, however, important to establish whether a worthwhile improvement has been achieved for an individual case and the proportion of cases that show improvement. There are basically two methods of assessing improvement using the PAR Index:

- (1) reduction in the weighted PAR score;
- (2) percentage reduction in the weighted PAR score.

A greater reduction in weighted PAR score is likelier where there is a high pretreatment score, indicating a greater deviation from normal and perhaps a greater need for treatment. Percentage reduction reflects the change relative to the pretreatment score. Thus, a change in score

from 50 to 10 (a reduction of 40) and a change from 15 to 3 (a reduction of 12) both produce an 80 per cent reduction, but a much greater improvement has been achieved in the first instance. In order to express the degree of improvement resulting from treatment, a method was developed for relating numerical change in the weighted PAR scores to consensus professional judgements.

## Design of the study

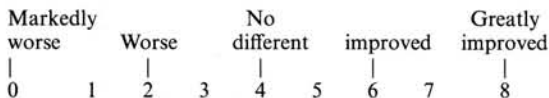
The panel of 74 examiners (the composition of the panel has been described in a previous paper Richmond *et al.*, 1992) was asked to examine 128 pairs of dental casts. Ninety-six pairs consisted of pre- and post-treatment cases from the Dental Practice Board and Manchester/Bristol Dental Hospitals, the remaining 32 pairs were derived from untreated cases from the Cardiff study in 1981 and 1984 (Shaw *et al.*, 1986). The untreated cases were included to 'load' the unimproved/improved area in order to establish a threshold between no change and improvement. A random stratified subsample of the dental casts was arranged in four groups of 32 pairs (A, B, C, and D). The panel of 74 exam-

iners recorded the degree of improvement on a 9-point scale from 0 (markedly worse) to 8 (greatly improved) (Fig. 1). Inter-examiner reliability assessments were derived for the four groups using the chance corrected measure of agreement (O'Connell and Dobson, 1984).

**Results**

*Reliability*

The inter-examiner results are shown in Table 1. The agreement within examiners for the groups ranged from 0.433 to 0.576 (moderate agreement). The low level of reliability suggests some



**Figure 1** Nine-point scale to determine improvement.

**Table 1** Inter-examiner agreement for the panel of 74 examiners: assessment of improvement.

Group (32 pairs)	Number of examiners	Kappa statistic
A	15	0.446
B	18	0.433
C	18	0.450
D	19	0.576

considerable disagreement between the panel of 74 examiners in assessing improvement.

*Categorisation of improvement*

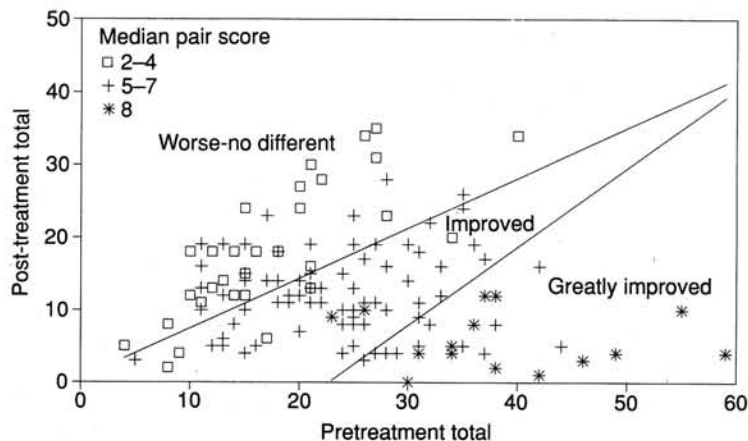
The panel of 74 examiners median scores derived from the 9-point scale were classified into three groups, 'worse or no different', 'improved', and 'greatly improved' (Table 2).

*Discriminant analysis*

Discriminant analysis is a statistical analysis which distinguishes between two or more groups of cases (Nie *et al.*, 1975). Discriminant analysis was used to create a linear separation between the three groups: 'worse or no different', 'improved' and 'greatly improved' related to the panel of 74 examiners subjective assessments of improvement on the pairs of dental casts, and the total weighted PAR scores for the pre- and post-treatment dental casts. As the number of cases was relatively small the discriminant function was undertaken on all 128 cases. The linear discrimination is shown in Fig. 2. The squares represent the panel's median scores 2, 3, and 4,

**Table 2** Classification of improvement derived from the panel of 74 examiners median scores.

Category	Panel's median score
Worse or no different	0, 1, 2, 3, 4
Improved	5, 6, 7
Greatly improved	8



**Figure 2** Categorization of improvement using discriminant analysis: pre- and post-treatment weighted PAR scores versus the panel of 74 examiners median pair scores with predicted class boundaries.

the cross 5, 6, and 7, and the asterisk 8. The nomogram (or graph) is separated into three sections: the upper, 'worse or no different'; the lower, 'greatly improved'; and the middle section, 'improved'.

#### Using the nomogram

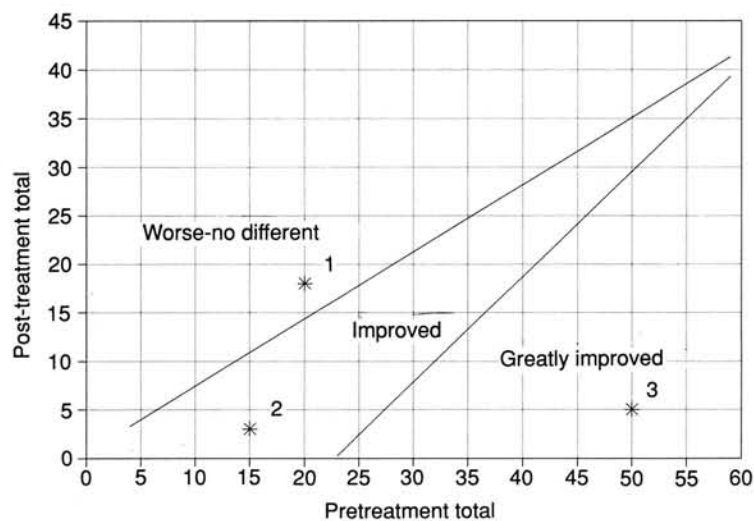
The pre-treatment weighted PAR scores are on the horizontal and the post-treatment weighted PAR scores on the vertical axes (Fig. 3). The scores for the pre- and post-treated cases are read off their respective axes and where the intercept falls indicates the degree of improvement. For example, Case 1 started at 20 and finished at 18, and so this case falls in the 'worse or no different' zone. Case 2 starts at 15 and finishes at 3, indicating that the case has been 'improved'. Case 3 started with a score of 50 and finished at 5; the intercept indicates that the case has been 'greatly improved'.

#### Assessment of classified cases

The weighted PAR scores (pretreatment, post-treatment, reduction, and percentage reduction) are shown in Table 3 for the three classifications of improvement.

The pretreatment weighted PAR scores for the 'worse or no different' and 'improved' categories have similar arithmetical means and ranges (means: 17 and 23, range: 4–40). The initial weighted PAR scores were much higher in the 'greatly improved' category.

The change in scores was much less in the 'worse or no different' and 'improved' categories. The distinguishing factor between the 'worse or no different' and 'improved' categories was percentage reduction. There was a clear demarcation between the two categories of approximately 3 per cent ('worse-no different', 80 to 25.71 per cent and 'improved', 28.57 to 83.33 per cent). An overlap existed between 'improved' and 'greatly improved' sections in respect of percentage reduction; however, a clear demarcation existed in reduction of the PAR score of 1 point ('improved', 2 to 22 and 'greatly improved', 23 to 55). An analysis of variance indicated a statistically significant difference between the groupings ( $F_{2,128} = 56.02, P < 0.0001$ ). A multiple comparison test was employed (Scheffe  $P < 0.05$ ) and indicated statistically significant differences between the three groups for all the variables recorded (pretreatment, post-treatment, reduction, and percentage reduction in weighted PAR scores). There appeared to be two systems operating in assessing improvement. The line separating the 'worse or no different' from the 'improved' section on the nomogram represents approximately a 30 per cent improvement; that is, a score will need to be reduced by 30 per cent for the case to be considered 'improved'. The line between the 'improved' and 'greatly improved' sections denotes an average change in weighted PAR score of 22. This suggests the case has to



**Figure 3** Categorization of improvement using the nomogram. The pre- and post-treatment PAR scores can be plotted to determine the degree of improvement.

**Table 3** Relationship of pre- and post-treatment, reduction, and percentage reduction in weighted PAR scores for the three classifications of improvement.

	Pretreatment	Post-treatment	Reduction	% Reduction
Worse—no different (45 cases)				
mean	17	19	-1.47	-12.39
standard deviation	7	7	4.54	27.13
minimum	4	5	-9.00	-80.00
maximum	40	35	9.00	25.71
Improved (56 cases)				
mean	23	11	12.29	52.85
standard deviation	7	5	5.21	14.54
minimum	5	2	2	28.57
maximum	37	24	22	83.33
Greatly improved (26 cases)				
mean	37	5	31.65	85.54
standard deviation	8	4	8.25	9.52
minimum	26	0	23	61.90
maximum	59	16	55	100

$F_{2,128} = 56.02, P < 0.0001$ ; Scheffé indicates a significant difference between all groups for all variables.

improve by at least 22 points before it can be classified as 'greatly improved'. Table 4 shows the minimum change in weighted PAR score needed to produce 'improvement' and 'great improvement' in relation to the pretreatment scores.

#### Misclassifications

Seventy per cent of cases were correctly classified, 39 cases fell outside their assigned boundaries. A great proportion of these misclassifications appeared to result from the lack of precision of the panel of 74 examiners in assessing improvement. It would be inappropriate to look at all these misclassified cases, but three cases are illustrated below.

**Table 4** Assessment of the minimum reduction and percentage reduction in weighted PAR scores necessary to bring about 'improvement', and 'great improvement'.

Start PAR	Improvement		Great improvement	
	Reduction%	Reduction	Reduction%	Reduction
50	15	30	21	42
40	12	30	22	55
30	9	30	23	77
20	6	30	—	—
10	3	30	—	—
5	2	40	—	—

#### Case analysis—Fig. 4

Classified as 'improved' using the PAR Index (weighted PAR 23–9); panel of 74 examiners assessment—'greatly improved'.

The pretreatment model has an increased overjet of 7 mm. There is mild upper and lower incisor crowding. As a result of treatment the overjet has been reduced by 3 mm to an overjet of 4 mm. Slight spacing has developed anteriorly. The overbite has increased. The final buccal occlusion, although satisfactory is not as good as on the pretreatment models.

#### Case analysis—Fig. 5

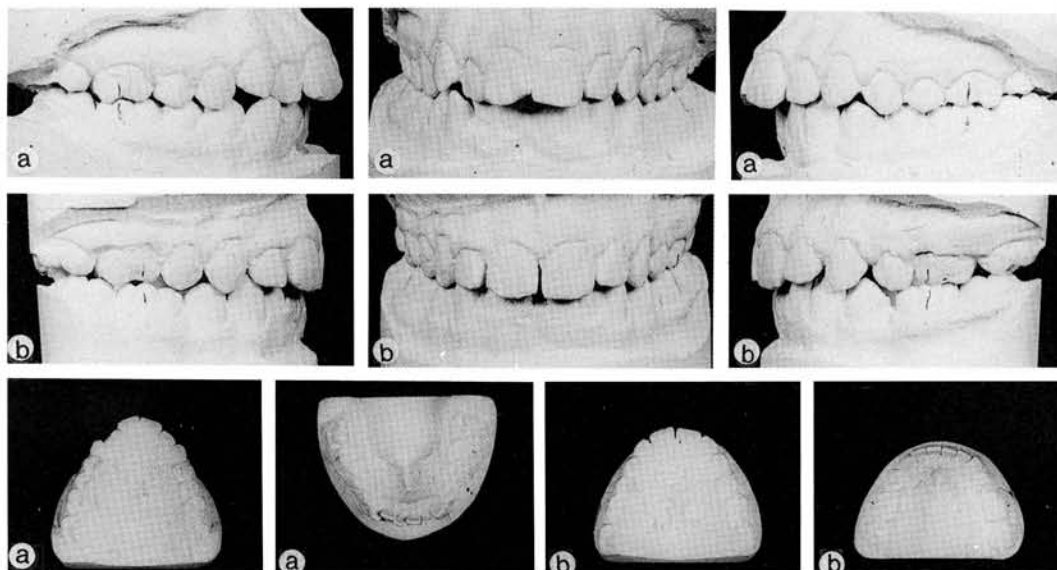
Classified 'greatly improved' using the PAR Index (weighted PAR 44–5); panel of 74 examiners assessment—'improved'.

The pretreatment model has an increased overjet of 12 mm and an increased overbite. As a result of orthodontic treatment the overjet has been reduced from 12 to 3 mm and the overbite has been improved. The buccal occlusion has been improved on the right side with minimal improvement on the left.

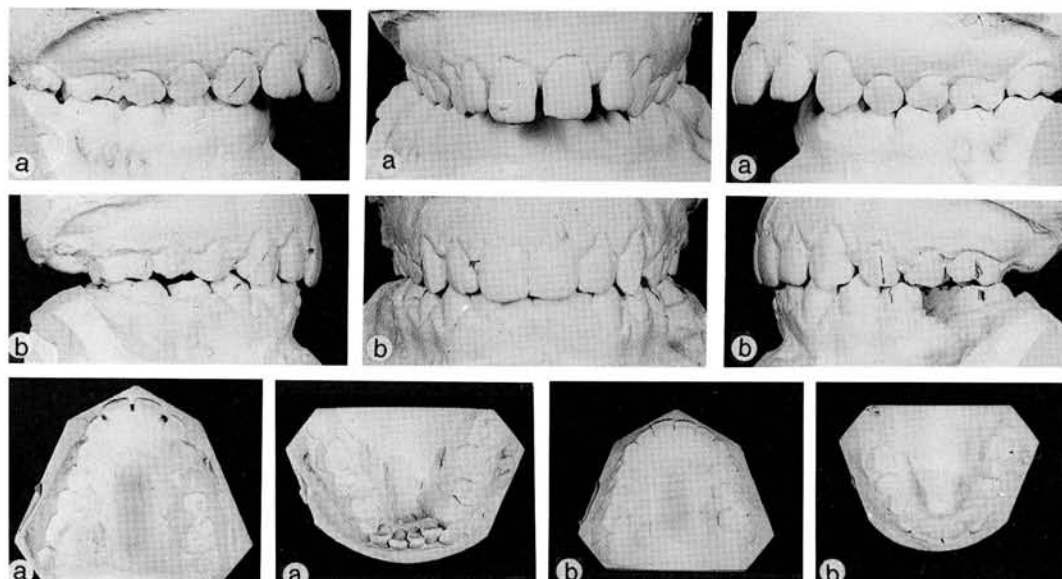
#### Case analysis—Fig. 6

Classified as 'worse or no different' using the PAR Index (weighted PAR 25–23); panel of 74 examiners assessment—'improved'.

The pretreatment model has an increased overjet and the upper lateral incisors are absent. Treatment involved extracting the upper left decidu-



**Figure 4** Case analysis: (a) pretreatment and (b) post-treatment dental casts illustrate the reduction in weighted PAR score from 23 to 9. The panel of 74 examiners assessment 'greatly improved', the nomogram: 'Improved'.

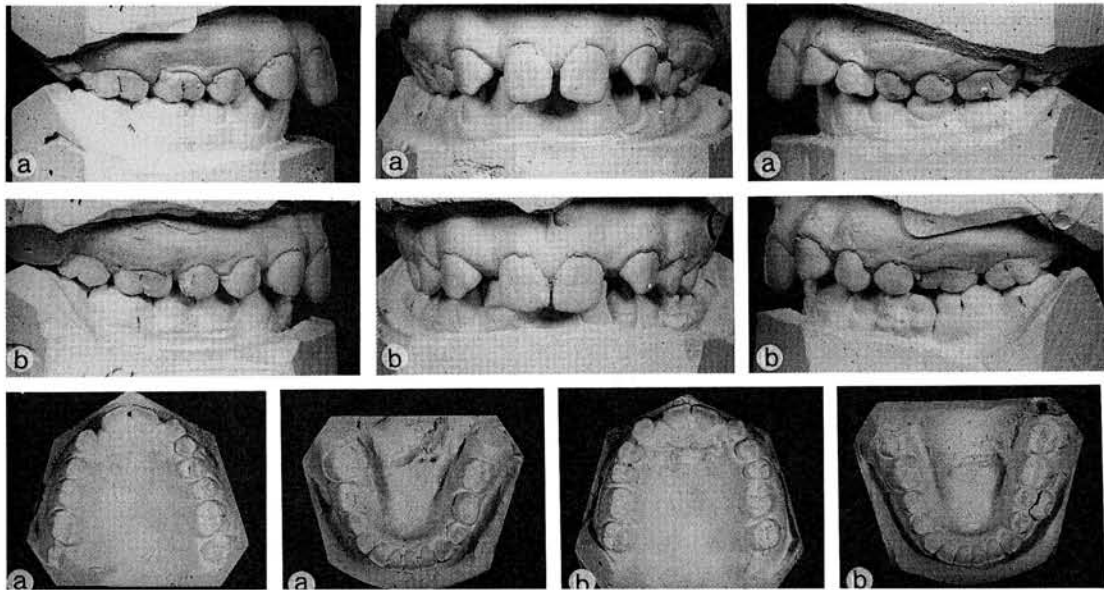


**Figure 5** Case analysis: (a) pretreatment and (b) post-treatment dental casts illustrate the reduction in weighted PAR score from 44 to 5. The panel of 74 examiners assessment 'improved'; the nomogram: 'Greatly improved'.

ous canine and approximating the central incisors and opening the lateral spaces. The post-treatment models indicate that the overjet and overbite are similar to the pretreatment models and the buccal occlusion is not as good as on the pretreatment model.

#### *Comment*

Although a substantial improvement has been achieved in Fig. 4 (a change in PAR score of 14) a greater improvement is shown in Fig. 5 (a change of 39 PAR points). Very little change



**Figure 6** Case analysis: (a) pretreatment and (b) post-treatment dental casts illustrate the reduction in weighted PAR score from 25 to 23. The panel of 74 examiners assessment 'improved'; the nomogram: 'worse or no different'.

has been achieved in Fig. 6. The case illustrated in Fig. 4 has 'improved', but a change of 22 PAR points is required for a case to be judged 'greatly improved'; this has been accomplished in Fig. 5. For a case to be judged 'improved' a 30 per cent reduction in PAR score is required. A reduction of only 8 per cent has been attained in Fig. 6 (worse or no different).

### Discussion

There appeared to be considerable disagreement between the 74 examiners regarding what constituted improvement. Assessing improvement is easy when a case has been treated to an 'ideal' occlusion, but most treated cases still exhibit some occlusal anomalies. For example, it is not uncommon to see an overjet reduced with the development of a cross-bite on one or more of the molars. Removable appliances in particular offer limited control in tooth movement, for example during canine retraction, canines often flare buccally and become rotated. Excessive tipping of teeth is also frequent. Therefore, in the correction of one occlusal anomaly, a less than ideal result may be the consequence of incomplete correction of that element or the introduction of a new occlusal imperfection.

The division in opinions between the panel of 74 examiners in assessing improvement in occlusion highlights the need for an objective and reliable method to record improvement.

The panel of examiners had to decide upon the degree of improvement achieved and weigh up the severity of residual malocclusion subjectively, whereas the nomogram provides the opportunity to assess improvement objectively. The PAR Index has been weighted to give appropriate emphasis to those anomalies which the panel of 74 examiners considered important. Thus, a reduction of a large overjet would be regarded a substantial improvement even if a cross-bite developed on the first molars during treatment.

The study revealed that a 30 per cent reduction in the weighted PAR score is required to produce a change which would be judged as 'improved' and a score change of approximately 22 for 'greatly improved'. This suggests that minor deviations from normal cannot be 'greatly improved', as the case is not severe enough in the first instance. The 30 per cent boundary was the minimal level in which the panel of 74 examiners considered that a case could be 'improved'. However, cases which have been changed by 30 or even 60 per cent, are

both categorized as 'improved' even though one case has been treated twice as well as the other.

The percentage reduction in weighted PAR score could be used to assess or set the standard of orthodontic treatment within a publicly funded health care system. For example, the cases illustrated in Figs 4, 5, 6 have a percentage change in weighted PAR score of 61, 89, and 8, per cent, respectively. The case in Fig. 5 represents a good standard of treatment and has been 'greatly improved'. For a practitioner to demonstrate consistently high standards, the proportion of the case load in the 'worse or no different' category should be negligible and the mean percentage reduction should be as high as possible.

A high standard of treatment may be judged according to the mean percentage reduction in weighted PAR score for an individual practitioner's case load, for example, greater than 70 per cent.

For a practitioner to produce high standards and treat those cases who have perhaps a greater need for treatment, the mean percentage reduction for the case load must not only be high (e.g. greater than 70 per cent), but the percentage of cases having been 'greatly improved' should also be high (e.g. greater than 40 per cent).

*Investigation of orthodontic treatment standards in the General Dental Services in England and Wales*

In a random sample of 1010 cases treated with removable and/or fixed appliances in the General Dental Services of England and Wales (Richmond, 1990), one out of every five patients would be categorized as 'worse or no different' (Table 5). The overall mean percentage reduction in weighted PAR score was 55 per cent. When the standard of treatment was assessed

related to the appliance used (Table 6), it was found that the use of upper and lower fixed appliances produced the best standard of treatment (71.4 per cent reduction in weighted PAR score) and the patient was less likely to be worse after treatment (implied by the lower 95 per cent confidence limits). Single arch removable and fixed appliances showed a mean percentage reduction of around 50 per cent and the 95 per cent confidence limits show a large negative range indicating a substantial proportion of patients did not benefit from orthodontic treatment.

*A pilot study to assess the standard of orthodontic treatment in Norway*

A sample of 220 cases was collected from Norwegian specialist orthodontists. Twenty consecutively finished cases were requested from six practitioners. The remaining 100 cases were chosen randomly from the model stores of five practices visited by one of the authors. The weighted PAR scores were applied to the pre- and post-treated cases. The mean percentage reduction in weighted PAR score for all cases was 78 per cent (a good standard of treatment). Only 4 per cent of patients were categorized as 'worse or no different' (Table 5).

**Table 6** Percentage reduction in weighted PAR score related to appliance type.

	No of cases	% Mean reduction	95% Confidence limits
Upper removable	511	49.3	36.2-81
Upper and lower removable	48	50.4	-27.3-80.7
One arch fixed	149	54.6	-72.5-88.0
Upper and lower fixed	196	71.4	-1.6-92
Combination	106	55.3	-47.4-86.5

**Table 5** Outcome of treatment by assessing improvement using the nomogram (percentages).

Outcome	General Dental Services of England and Wales	Norwegian orthodontic service
Worse or no different	20.7	4
Improved	56.9	61
Greatly improved	22.4	35
Number of cases	1010	220

Although in this article it is not appropriate to make direct comparisons between British and Norwegian standards, it does highlight the need for further investigation and perhaps extend the comparison to other countries.

### Conclusions

A measure has been developed to assess improvement objectively. Using this index it was revealed that at least 30 per cent reduction was needed for a case to be judged 'improved' and a change in score usually of 22 to bring about a change judged to be 'greatly improved'.

For a practitioner to demonstrate high standards, we submit that the proportion of an individual's case load lying in the 'worse or no different' category should be negligible and the mean percentage reduction should be as high as possible (e.g. greater than 70 per cent). The greater the mean percentage reduction in weighted PAR score the higher the standard of orthodontics achieved. If the mean percentage reduction is high and the proportion of cases that have been 'greatly improved' is also high this indicates that the practitioner is treating a great proportion of cases with a clear need for treatment, to a high standard.

### Address for correspondence

Dr Stephen Richmond  
Department of Oral Health and Development  
Coupland III Building  
University of Manchester  
Higher Cambridge Street  
Manchester M13 6FH

### Acknowledgements

The authors would like to thank Kevin O'Brien, Iain Buchanan, Rosemary Jones, and Chris Stephens for their help during this study. We are grateful for the help received from the Norwegian orthodontists. We also would like to acknowledge the financial support of the Dental Practice Board.

### References

- Nie N H, Hull C H, Jenkins J G, Steinbrenner K, Bent D H 1975 Statistical package for the social sciences. McGraw-Hill, New York
- O'Connell D L, Dobson A J 1984 General observer agreement measures on individual subjects and groups of subjects. *Biometrics* 40: 973-983
- Richmond S 1990 A critical evaluation of orthodontic treatment in the General Dental Services of England and Wales. PhD thesis, University of Manchester
- Richmond S, Shaw W C, O'Brien K D, Buchanan I, Jones R, Stephens C D, Roberts C T, Andrews M 1992 The development of the PAR Index (Peer Assessment Rating): reliability and validity. *European Journal of Orthodontics* 14: 125-139
- Shaw W C, Addy M, Dummer P M H, Ray C, Frude N 1986 Dental and social effects of malocclusion and effectiveness of orthodontic treatment: a strategy for investigation. *Community Dentistry and Oral Epidemiology* 14: 60-64