

Are the Peer Assessment Rating Index and the Index of Treatment Complexity, Outcome, and Need suitable measures for orthognathic outcomes?

K. M. Templeton*, R. Powell*, M. B. Moore*, A. C. Williams** and J. R. Sandy***

*Royal Devon and Exeter Hospital, Exeter, **King's College Hospital, London, and

***Bristol Dental Hospital, Bristol, UK

SUMMARY The aim of this study was to determine which of two occlusal indices were the most appropriate for use in the assessment of orthognathic outcome. The indices used were the Peer Assessment Rating (PAR) Index and the Index of Treatment Complexity, Outcome, and Need (ICON). These indices were validated against the subjective assessments of treatment outcome and treatment improvement obtained from a panel of experienced orthodontic consultants.

For the subjective assessment, intraexaminer agreement for ranking treatment outcome, from patient study models (30 models), was good. Interexaminer agreement for ranking treatment outcome, in the same way, was good or moderate. Intraexaminer agreement for ranking treatment improvement (30 start and finish pairs of models) was very good or good. Interexaminer agreement for ranking treatment improvement ranged from good to fair.

All the patient study models were scored using PAR and ICON. The level of correlation between PAR and ICON scores of treatment outcome and the subjective ranking of treatment outcome was significant ($P < 0.001$). The level of correlation between PAR and ICON scores of treatment improvement and the subjective ranking of treatment improvement was also significant ($P < 0.001$). It is concluded that both PAR and ICON are suitable indices for assessing the clinical outcome of combined orthodontic treatment and orthognathic surgery.

Introduction

The number of patients presenting for orthodontic and orthognathic treatment is increasing annually and although there are no figures for the UK, there has been an attempt to quantify this in the USA. Bailey *et al.* (2001) suggested that 1.5 million people in the USA have malocclusions severe enough to warrant treatment by a combined approach. If this is extrapolated to the UK then there are potentially 250 000 patients in this category. This is a significant number of potential patients who would require considerable financial support from the funding agencies and/or individuals, depending on health care funding arrangements.

Patient perception of outcomes with combined orthognathic treatment is generally extremely high with satisfaction rates quoted between 50 and 100 per cent (Finlay *et al.*, 1995; Cunningham *et al.*, 1996a,b; Zhou *et al.*, 2001; Siow *et al.*, 2002). Whilst patient satisfaction is an important consideration in the overall judgement of the treatment result, the clinical outcome is also a major factor. As health care moves towards more stringent financial accountability, hard outcome measures of treatment are becoming financially, as well as professionally, important.

If outcomes are to be evaluated then it is in the interest of patients and professionals that any third-party involvement, such as government or health care agencies, should be informed. Orthodontics has been particularly active in the

development of indices to assess treatment need, difficulty, and improvement (Cons *et al.*, 1986; Brook and Shaw, 1989; Richmond *et al.*, 1992; Daniels and Richmond, 2000). If indices are to be used by third-party agencies to assess the clinical outcome of treatment, it is important that clinicians are confident that appropriate measures of outcome are being used to assess their work. Clinical outcome in orthodontics is commonly measured using occlusal indices assessed from study models and lateral cephalometric radiographs.

Occlusal indices

Occlusal indices are tools used to ascribe either numerical or categorical values to malocclusions. Various indices have been developed to enable need for treatment, severity of malocclusion, complexity of malocclusion, or added value during treatment to be defined for an individual or population.

Peer Assessment Rating

One of the most commonly used occlusal indices is the Peer Assessment Rating (PAR) which has been shown to have validity and reproducibility (Richmond *et al.*, 1992; DeGuzman *et al.*, 1995; Firestone *et al.*, 2002a). The index was formulated through a number of meetings by 10

orthodontists who examined over 200 dental study models of subjects at varying stages of development and with varying malocclusions. Pre- and post-treatment casts were discussed until agreement was reached regarding the features that would be assessed in an estimate of malocclusion. PAR uses study casts to score maxillary and mandibular anterior alignment, buccal segment occlusion, overjet, overbite, and centreline discrepancies. Some aspects of occlusion are weighted, with overjet having the highest weighting. If pre- and post-treatment models are rated, the improvement achieved during treatment can be expressed as a percentage.

There have been different interpretations as to how the index can be used. For example, Firestone *et al.* (2002a) felt that the PAR Index was an excellent predictor of treatment need when compared with expert orthodontic opinion. Daniels and Richmond (2000) were clear in their view that the PAR Index was neither designed nor validated as an index of treatment need. Bergström and Halling (1997) found it poor at assessing the outcome of orthodontic treatment when compared with two Swedish indices which both involved clinical examination as opposed to PAR, which is carried out on study models. The latter has the disadvantage that factors such as periodontal health and facial aesthetics are not considered which, by contrast, both Swedish indices use. The PAR Index has also been criticized for being insensitive to some aspects of residual treatment need, such as remaining extraction spaces, rotations, and unfavourable incisor inclinations (Hinman, 1996). Nevertheless, it is popular and widely used.

Index of Orthodontic Treatment Need

The Index of Orthodontic Treatment Need (IOTN; Brook and Shaw, 1989) is based on an index which is used to assess treatment need within the Swedish Health Service. It is divided into two parts, a dental health component (DHC) and an aesthetic component (AC). The two components taken together give an indication of treatment need. This index was designed as a possible means of prioritizing resources to those patients with the most severe malocclusions. Using the aesthetic index on study models is difficult and potentially inaccurate. Comparison between IOTN and the Dental Aesthetic Index (DAI) in order to assess the treatment need of patients has been made (Jenny and Cons, 1996). The DAI was found to be more robust since it combines both the AC and DHC of the malocclusion. IOTN does not do this; it provides categorical data and not a numerical assessment of malocclusion. IOTN uses five categories and these are of most value when assessing the orthodontic needs of a large population. As an index it is insensitive if used to assess treatment changes in individuals, which in fairness it was never designed to do.

Index of Complexity, Outcome, and Need

This index was designed to combine the benefits of the PAR Index with those of IOTN (Daniels and Richmond, 2000).

The Index of Complexity, Outcome, and Need (ICON) has a very high level of validity and in comparison with PAR and IOTN it shows good correlation between IOTN and ICON in terms of treatment need. There is also a good correlation between PAR and ICON in respect of treatment outcome (Firestone *et al.*, 2002b; Fox *et al.*, 2002).

It is currently not known which occlusal index is most valid or reliable in assessing treatment outcome of combined orthodontics and orthognathic surgery. It is recognized that although overjet and reverse overjet are not measured directly in the ICON scoring system, the AC of this index adequately represents the importance of these occlusal traits for the purposes of assessing orthognathic cases (Daniels and Richmond, 2000). In a recent national audit of orthodontic treatment outcome, PAR was used as an outcome measure (McMullan *et al.*, 2003). Orthognathic cases, however, were discounted since PAR was not recognized as a valid measure of outcome in these cases. Bergström and Halling (1997) compared PAR with two Swedish indices of treatment outcome and included a group of patients treated with combined orthodontics and surgery and a group treated with orthodontics alone. The association between PAR and both Swedish indices was low, particularly in the group of patients treated with combined protocols. They concluded that PAR is not a good index for use in evaluating treatment outcome but offered no explanation as to why the discrepancy was greater in patients treated with orthognathic surgery than those treated with orthodontics alone.

The issue of validity of occlusal indices in combined orthognathic cases is often assumed rather than determined. Nurminen *et al.* (1999) used the PAR Index to assess the outcome of combined treatment and felt it appropriate because the weighted features of the index (overjet, overbite, and centreline discrepancies) were features seen in most patients treated by a combined approach. There was no assessment of PAR validity for use in these circumstances. The aim of the present study, therefore, was to determine the validity of PAR and ICON in combined orthognathic cases using study models alone in a retrospective investigation. Since only study models were assessed, it was not felt appropriate to examine the IOTN in this context.

Subjects and methods

Thirty patients were identified who had undergone combined orthodontic and orthognathic treatment within the National Health Service (NHS) in the South West Region of the UK. Patients were included in the study if they fulfilled the following criteria: undergone combined orthognathic/orthodontic treatment, treatment carried out within the NHS in the South West Region, between 1995 and 2001, pre- and post-treatment study models available, and non-syndromic skeletal discrepancy.

The subjects' study models were obtained from Southmead Hospital, Bristol, and the Royal Devon and Exeter Hospital, Exeter, UK. The patients had been treated by a number of different operators within these hospitals. The pre-treatment models were taken prior to placement of fixed appliances and the post-treatment models at completion of active treatment, when the fixed appliances had been removed. The records selected for this study were representative of a range of malocclusions likely to be treated with a surgical approach. The study models were numbered and the patient names obscured to prevent recognition and therefore potential bias.

Two main types of assessment were carried out:

1. Five experienced consultants working in hospital services in the South West Region were invited to rank the post-treatment study models in order from 'best' to 'worst', which produced a ranking of treatment outcome based on their opinion. The same consultants were then asked to rank the pre- and post-treatment models, which were presented as a pair, and thus produce a ranking in order of treatment improvement. These methods were repeated after 2 weeks to test intraexaminer reliability.
2. The study models were scored using the PAR and ICON indices. The pre-treatment, post-treatment, and treatment improvement scores were recorded. This gave an objective score for both outcome and improvement. Scoring was carried out by one examiner (KT), who had previously been calibrated on a formal course in the use of PAR and ICON.

The data collected were entered into a StatsDirect™ (Stats Direct Limited, Sale, Cheshire, UK) database for analysis.

Statistical analysis

Intra- and interexaminer reliabilities were tested using the weighted kappa (κ) statistic for both treatment outcome and treatment improvement. The rankings from the five consultants were combined to provide composite consultant rankings for treatment outcome and treatment improvement.

The composite results from the consultants' ranking of end of treatment study models were then correlated with the post-treatment PAR and ICON scores using Spearman's correlation coefficient. The composite treatment results from the consultants on the paired pre- and post-treatment study models were also correlated with the changes in PAR and ICON using Spearman's correlation coefficient.

Results

Intra- and interexaminer agreements for treatment outcome and treatment improvement are shown in Tables 1 and 2, respectively. These results show that intraexaminer

Table 1 Intraexaminer agreement using the weighted kappa (κ) statistic after two separate rankings of the treatment outcome models and the treatment improvement models. A weighted κ value of 0.8 represents very good agreement and 0.6 good agreement (Landis and Koch, 1977).

Examiner	κ value (95% CI)	
	Treatment outcome	Treatment improvement
1	0.77 (0.54–0.99)	0.79 (0.56–1.01)
2	0.69 (0.47–0.92)	0.67 (0.44–0.89)
3	0.69 (0.46–0.91)	0.82 (0.59–1.05)
4	0.64 (0.41–0.87)	0.71 (0.48–0.93)
5	0.76 (0.53–0.99)	0.75 (0.52–0.97)

agreement was good or very good in both categories and that the interexaminer agreement was good in most cases and moderate or fair in all others.

The mean subjective rankings for both outcome and improvement were then tested for correlation with the PAR and ICON scores. These results are presented in Figures 1 and 2 in scatter plot format. Spearman's rank correlation was carried out on this data. The results showed a significant correlation between consultant opinion of outcome and PAR [0.72 (0.47–0.85)] and ICON [0.66 (0.39–0.82)] ($P < 0.01$) scores. There was also a significant correlation between consultant opinion of improvement achieved during treatment and PAR [−0.68 (−0.84 to −0.47)] and ICON [−0.65 (−0.82 to −0.37)] ($P < 0.001$) improvement scores.

Discussion

These findings show that the panel of experts exhibited good levels of intraexaminer reliability in judging both outcome of treatment and improvement gained by treatment (average κ values 0.70 and 0.78, respectively). Previous investigations have shown higher levels of intraexaminer reliability (Younis *et al.*, 1997; Firestone *et al.*, 2002a,b). However, these latter studies asked examiners to choose between defined categories, such as 'requires treatment' or 'does not require treatment', or levels of improvement, such as 'good', 'average', or 'poor'. In the present study the examiners were asked to place the study models in ranked order, thus making the process more complex.

Interexaminer reliability was good when judging treatment outcome (average κ value 0.64) but only moderate when judging treatment improvement (average κ value 0.42). There were a larger number of models that needed to be judged in the treatment improvement aspect of the study and the sheer volume made it more difficult to consistently place the models in a ranked order.

If this study were to be repeated it would be helpful if the examiners were asked to band the quality of result as very good, good, average, poor, and very poor. All the examiners found it extremely difficult to consistently rank the cases in the middle of the group. When the data were examined in

Table 2 Interexaminer agreement using the weighted kappa (κ) statistic. A weighted κ value of 0.8 represents very good agreement, 0.6 good agreement, 0.4 moderate agreement and 0.2 fair agreement (Landis and Koch, 1977). Values in brackets are the 95 per cent confidence intervals.

Examiner	1	2	3	4	5
Treatment outcome					
1		0.71 (0.49–0.93)	0.68 (0.45–0.91)	0.63 (0.40–0.86)	0.65 (0.43–0.88)
2	0.71		0.62 (0.40–0.85)	0.54 (0.32–0.77)	0.57 (0.36–0.78)
3	0.68	0.62		0.7 (0.48–0.93)	0.7 (0.48–0.93)
4	0.63	0.54	0.7		0.63 (0.41–0.85)
5	0.65	0.57	0.7	0.63	
Treatment improvement					
1		0.65 (0.43–0.87)	0.55 (0.36–0.75)	0.35 (0.13–0.56)	0.31 (0.10–0.52)
2	0.65		0.65 (0.43–0.87)	0.41 (0.18–0.63)	0.24 (0.02–0.47)
3	0.55	0.65		0.28 (0.07–0.49)	0.22 (0.01–0.44)
4	0.35	0.41	0.28		0.49 (0.26–0.71)

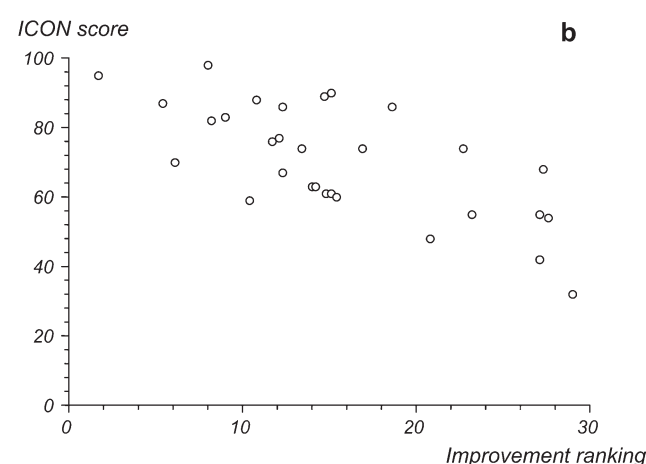
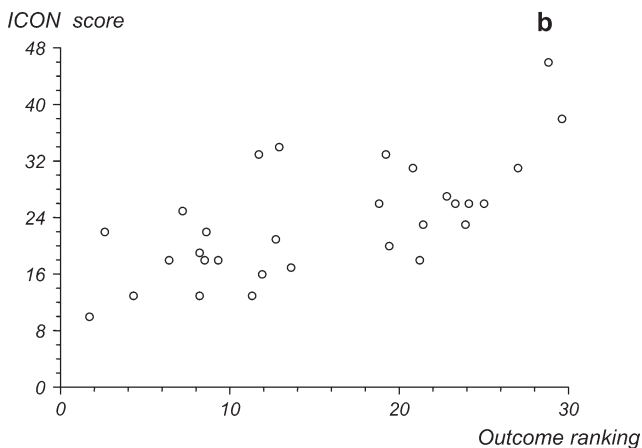
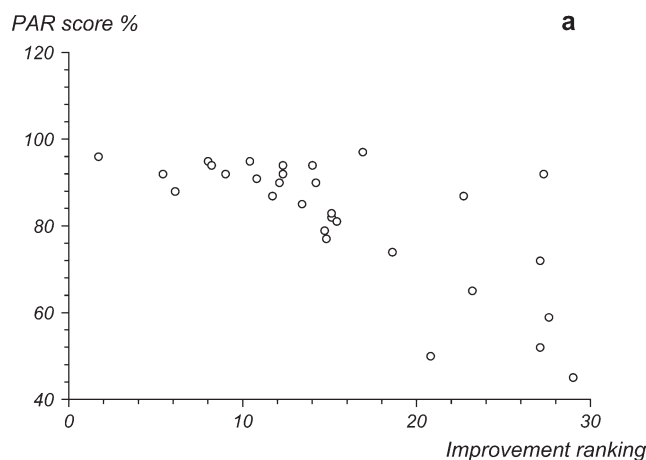
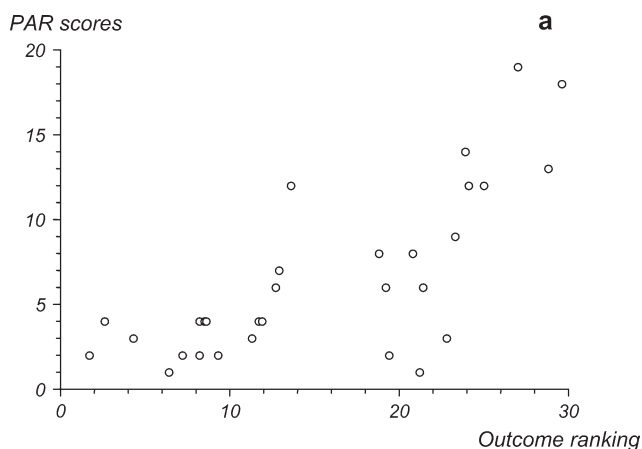


Figure 1 (a) Peer Assessment Rating (PAR) and (b) Index of Treatment Complexity, Outcome, and Need (ICON) scores of post-treatment study models compared with treatment outcome ranking by consultants.

Figure 2 (a) Peer Assessment Rating (PAR) and (b) Index of Treatment Complexity, Outcome, and Need (ICON) improvement scores of pre- and post-treatment study models compared with improvement ranking by consultants.

detail, the intra- and interexaminer reliabilities were high within the top five and bottom five ranking positions. The most significant anomalies seemed to arise when the examiners ranked the treatment improvement highly but the occlusal index score was low. The study casts for these specific cases were re-examined for commonality of occlusal

traits. There was no obvious similarity in these cases; they exhibited a variety of malocclusions and had been treated in different units.

The results also showed that both PAR and ICON exhibited a good level of agreement with the opinion of the

examiners. This was true for both treatment outcome and treatment improvement. PAR and ICON have previously been shown to have a strong correlation when assessing the outcome in patients treated orthodontically (Fox *et al.*, 2002). The latter study also included some patients being treated with combined orthodontics and orthognathic surgery. It is well-known that PAR is weighted heavily on positive and negative overjets and it is not therefore surprising that PAR correlates well with the subjective opinion of orthodontists in relation to treatment improvement in orthognathic cases. The same close relationship is clear for the ICON, which is quick and easy to use and gives an indication of treatment complexity and need.

Conclusions

1. The five consultants showed good inter- and intraexaminer reliabilities when assessing the outcome of patients treated with a combination of orthodontics and orthognathic surgery.
2. The five consultants showed moderate interexaminer reliability when assessing treatment improvement of patients treated with combined orthodontics and orthognathic surgery.
3. Both PAR and ICON showed a significant correlation with subjective opinion for treatment outcome and treatment improvement.
4. PAR and ICON are both suitable occlusal indices for assessing outcome and improvement in patients treated with a combined orthodontic and orthognathic approach.

Address for correspondence

Professor J. R. Sandy
Division of Child Dental Health
Bristol Dental Hospital
Lower Maudlin Street
Bristol BS1 2LY
UK
E-mail: jonathan.sandy@bris.ac.uk

Acknowledgements

We are very grateful to all the consultants who willingly gave their time. Our thanks also go to the members of staff at both Southmead Hospital and The Royal Devon and Exeter Hospital for their assistance in locating the study models used. This study was supported by a research grant from the British Orthodontic Society Foundation (RJ3844).

References

Bailey L-J, Haltiwanger G H, Blakey G H, Proffit W R 2001 Who seeks surgical-orthodontic treatment: a current review. *International Journal of Adult Orthodontics and Orthognathic Surgery* 16: 280–292

- Bergström K, Halling A 1997 Comparison of three indices in evaluation of orthodontic treatment outcome. *Acta Odontologica Scandinavica* 55: 36–43
- Brook P H, Shaw W C 1989 The development of an index of orthodontic treatment priority. *European Journal of Orthodontics* 11: 309–320
- Cons N C, Jenny J, Kohout F J 1986 DAI: the Dental Aesthetic Index. College of Dentistry, University of Iowa
- Cunningham S, Crean S J, Hunt N P, Harris M 1996a Preparation, perceptions and problems: a long-term follow-up study of orthognathic surgery. *International Journal of Adult Orthodontics and Orthognathic Surgery* 11: 41–47
- Cunningham S, Hunt N P, Feinmann C 1996b Perceptions of outcome following orthognathic surgery. *British Journal of Oral and Maxillofacial Surgery* 34: 210–213
- Daniels C, Richmond S 2000 The development of the index of complexity, outcome and need (ICON). *Journal of Orthodontics* 27: 149–62
- DeGuzman L, Bahiraei D, Vig K W L, Vig P S, Weyant R J, O'Brien K D 1995 The validation of the peer assessment rating for malocclusion severity and treatment difficulty. *American Journal of Orthodontics and Dentofacial Orthopedics* 107: 172–176
- Finlay P M, Atkinson J M, Moos K F 1995 Orthognathic surgery: patient expectations; psychological profile and satisfaction with outcome. *British Journal of Oral and Maxillofacial Surgery* 33: 9–14
- Firestone A R, Beck F M, Beglin F M, Vig K W L 2002a Evaluation of the Peer Assessment Rating index as an index of orthodontic treatment need. *American Journal of Orthodontics and Dentofacial Orthopedics* 122: 463–469
- Firestone A R, Beck F M, Beglin F M, Vig K W L 2002b Validity of the Index of Complexity, Outcome and Need (ICON) in determining orthodontic treatment need. *Angle Orthodontist* 72: 15–20
- Fox N A, Daniels C, Gilgrass T 2002 A comparison of the Index of Complexity, Outcome and Need (ICON) with the Peer Assessment Rating (PAR) and the Index of Orthodontic Treatment Need (IOTN). *British Dental Journal* 24: 193 225–230
- Hinman C 1996 The Dental Practice Board. Orthodontics—the current status. *British Journal of Orthodontics* 22: 287–290
- Jenny J, Cons N 1996 Comparing and contrasting two orthodontic indices, the index of orthodontic treatment need and the Dental Aesthetic Index. *American Journal of Orthodontics and Dentofacial Orthopedics* 110: 410–416
- Landis J R, Koch G G 1977 The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174
- McMullan R E, Doubleday B, Muir J D, Harradine N W, Williams J K 2003 Development of a treatment outcome standard as a result of a clinical audit of the outcome of fixed appliance therapy undertaken by hospital-based consultant orthodontists in the UK. *British Dental Journal* 194: 81–84
- Nurminen L, Pietilä T, Vinkka-Puhakka H 1999 Motivation for and satisfaction with orthodontic-surgical treatment: a retrospective study of 28 patients. *European Journal of Orthodontics* 21: 79–87
- Richmond S *et al.* 1992 The development of the PAR index (Peer Assessment Rating): reliability and validity. *European Journal of Orthodontics* 14: 125–139
- Siow K K, Ong S T, Lian C B, Ngeow W C 2002 Satisfaction of orthognathic surgical patients in a Malaysian population. *Journal of Oral Science* 44: 165–171
- Younis J W, Vig K W L, Rinchuse D J, Weyant R J 1997 A validation study of three indexes of orthodontic treatment need in the United States. *Community Dentistry and Oral Epidemiology* 25: 358–362
- Zhou Y H, Hägg U, Rabie A B M 2001 Patient satisfaction following orthognathic surgical correction of skeletal Class III malocclusion. *International Journal of Adult Orthodontics and Orthognathic Surgery* 16: 99–107