

IN BRIEF

- Yearly consultant appraisal is now a reality, and part of the appraisal folder includes 'results of clinical outcomes as compared to Royal College, Faculty or speciality association recommendations, where available'.
- It is important when setting standards that a focused and robust benchmark is developed, against which the treatment outcomes of individual consultants can be fairly assessed.
- This paper sets out to develop such a standard, and also discuss some of the wider issues surrounding appraisal and personal audit.

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.

R. E. McMullan,¹ B. Doubleday,² J. D. Muir,³ N. W. Harradine⁴ and J. K. Williams⁵

Bristol's much-publicised cardiac surgery problems and subsequent enquiry¹ have drawn attention to the need for audit of treatment outcomes throughout all hospital specialties. Patient anxiety, government policy and the desire of the professions to re-establish public confidence, have further encouraged changes to the system. For medical and dental specialties, such challenges have already been taken up by the Royal Colleges with the establishment of clinical effectiveness committees. Hospitals have modified their procedures and, for consultants, yearly appraisal is already a reality. The Orthodontic Clinical Effectiveness Working Party of the Royal College of Surgeons of England (now the Clinical Effectiveness Committee of the British Orthodontic Society) set up this audit to measure the outcome of fixed appliance treatment and to establish a benchmark for the standard of treatment to be expected from a consultant orthodontist. This paper describes how the audit was carried out, presents the findings and goes on to discuss some of the wider issues involved in audit, clinical governance and appraisal. The Consultant Orthodontists Group of the British Orthodontic Society funded this audit and the results and data set of dental casts remain their property.

Appraisal and Assessment

Appraisal, which is intended to be non-threatening, is a two-way process aimed at addressing the concerns of the consultant as well as the appraiser, and to contribute to the 'personal development plan' for each consultant. This will play a part in the establishment of 'clinical governance',

which is defined as the 'corporate responsibility for treatment outcomes'. The 'personal development plan' – a legally binding document – will, for the first time, acknowledge that treatment outcomes are not only the responsibility of the clinician, but also of the management. Appraisal, however friendly, will result in assessment and will eventually be the basis for re-certification. Audit has become an integral part of annual appraisal and assessment. Part of the appraisal folder includes 'results of clinical outcomes as compared to relevant Royal College, Faculty or speciality association recommendations, where available'.

Index of Orthodontic Treatment Need

Brook *et al.*² described the Index of Orthodontic Treatment Need (IOTN). It classifies malocclusions according to need using a dental health component (DHC) and an aesthetic component (AC). In this audit we only used the DHC, which classifies maloc-

clusions from 1 (mild with little or no need for treatment) to 5 (severe with very great need for treatment). The most severe occlusal trait is used to index the case, and as far as is possible, the subdivisions between the groups are based on scientific evidence as to dental health gain from orthodontic treatment. In an average child population, (11 – 12 year olds), 37% of the population score as IOTN 1 or 2, 27% score IOTN 3, while 36% score IOTN 4 or 5.³

Assessment of tooth irregularity; the PAR index

Orthodontists have routinely assessed one aspect of treatment outcome, ie improvement in tooth irregularity, by the use of pre- and post-treatment dental casts. Such evaluation remained largely subjective until the introduction of the PAR index in 1992.⁴ This quantitative analysis measures tooth irregularity within each arch and the degree of malocclusion between the arches in all

¹Consultant Orthodontist, Department of Orthodontics, Altnagelvin Hospital, Glenshane Rd., Londonderry, Co Londonderry, ²Consultant Orthodontist, Department of Orthodontics, Glasgow Dental Hospital, 378, Sauchiehall Street, Glasgow, Strathclyde, ³Consultant Orthodontist, Orthodontic Department, Central Outpatients Department, North Staffordshire Hospital, Stoke on Trent, Staffordshire, ⁴Consultant Orthodontist, Department of Orthodontics, Bristol Dental Hospital, Lower Maudlin St., Bristol, Avon, ⁵Consultant Orthodontist, Department of Orthodontics, Leeds Dental Institute, Clarendon Way, Leeds, West Yorkshire.

*Correspondence to: Miss R. E. McMullan
E-mail rmcnullan@alt.n-i.nhs.uk

Refereed Paper

Received 20.08.01; Accepted 11.09.02

© British Dental Journal 2003; 194: 81–84

three planes of space. A set of pre-treatment casts for a very severe malocclusion may achieve a score of 40 or more, while a mild malocclusion will score 10 to 20. Although good treatment will produce a low PAR score, it is very rare for the final models to achieve zero. The success of treatment is commonly described in terms of the percentage improvement in PAR for a given case. The PAR index, although it has limitations, has become a very useful and widely applied measure and one that readily lends itself to the establishment of a standard.

If the pre-treatment PAR score is plotted against the post treatment PAR score, the change in PAR score can be demonstrated as a nomogram. Score changes below 30% are classified as 'worse/no different', while those above 30% are classified as 'improved' with changes over 22 points being classified as 'greatly improved'. Use of the three categories of 'greatly improved', 'improved' and 'worse/no different' as described by Richmond *et al.*⁴ has a good visual impact, but the choice of descriptions has created some discussion and one of the unusual features of the system is that two categories describe percentage-change while one describes an absolute change. The requirement for a minimum absolute change to qualify for the description 'greatly improved' is understandable to avoid giving this accolade to a mildly irregular case that has achieved a final score close to zero. Certain cases however can start with a low PAR score, but have a high treatment need and considerable treatment complexity. Cases involving impacted maxillary canines are an example of such a situation.

Setting a standard

Richmond *et al.*⁴ said: 'For a practitioner to demonstrate high standards, the proportion of an individual's case load falling in the "worse/no different" category should be negligible, and the mean reduction should be as high as possible (viz. greater than 70%)'. In the same paper the authors came to the conclusion: 'If the mean percentage reduction in PAR score is high and the proportion of cases that have "greatly improved" is also high, this indicates a practitioner is treating a great proportion of cases with a clear need for treatment to a high standard.' Both of these statements are considered by the profession to be valid but they are still subjective, being based on the pooled opinion of 74 orthodontists during the validation exercise of the PAR Index, and not upon quantitative analysis of consecutively completed cases. Until now, these are the standards against which consultant orthodontists have measured this aspect of their results.

A large number of published articles have used the PAR Index to describe treat-

ment outcomes.⁵⁻¹³ The paper of greatest relevance to the current audit is that by O'Brien *et al.* in 1993,⁵ in which the authors looked at 17 hospital departments and investigated some 1,392 cases treated with fixed appliances. They examined all cases, including those in which the appliances were removed early, and included all grades of operator. The authors showed that the mean percentage change in PAR ranged from 53% to 78% amongst the departments and that the effectiveness of treatment provision was influenced by the grade of operator, the choice of treatment methods and – interestingly – by 'departmental attitudes and aspirations'. All these factors are important to know, but make it harder to compare like with like.

For this reason, the authors of this paper set out to establish the improvement in PAR for a homogeneous group of prospectively completed upper and lower fixed appliance cases, treated solely by consultant orthodontists in the UK, and from this to establish a focused and robust benchmark against which the treatment outcomes of individual consultants could be sensibly and fairly assessed.

Protocol

All 204 consultant orthodontists on the main list of the Consultant Orthodontist Group of the British Orthodontic Society were invited to participate by submitting models of the first six consecutive patients treated with upper and lower fixed appliances who had been debonded after 1st August 1999. Participants were asked to send only cases that they had treated personally, including any that were debonded early. They were also to include any patients who had had treatment prior to fixed appliance therapy, eg a functional appliance.

They were asked to exclude patients born with cleft lip and/or palate, orthognathic surgery and severe oligodontia cases. PAR is not recognised as a valid measure of treatment outcome in these patients, as although it will measure improvement in tooth irregularity, this may not be the main treatment objective for these groups of patients. Other indices such as the GOSLON Index^{14,15} are more appropriate to audit the outcome of treatment for cleft lip and palate patients and appropriate outcome indices for orthognathic and severe oligodontia patients are being developed.

It was pointed out to the participants that there was little incentive for anyone to select cases with better outcomes because the results were to be anonymous and an unrealistically high standard of PAR reduction would constitute a future burden as an outcome measure. Each participating consultant was allocated a unique identification number known only to him or her and

the first author. Participants were asked to send pre- and post-treatment casts of the six cases, with only a case identification and consultant audit number marked on the models, to the Bristol Hospital dental laboratory, where orthodontic technicians scored the models.

Every consultant was asked to complete and return, with each set of models, a form giving the IOTN DHC score 1-5 and the appropriate suffix.² This was to permit analysis of the types of cases being accepted for treatment, because there is anecdotal evidence that in some units at least, treatment in the hospital service is rationed to IOTN 4 and 5 unless there are mitigating circumstances. The consultants were also asked to provide any other information relevant to the analysis of the cases, eg the presence of unerupted palatal canines (to facilitate scoring), or the fact that treatment had been discontinued early – together with the reason, eg request by the patient, poor oral hygiene or poor cooperation.

The resultant PAR scores (but not the models) were forwarded to the second author who carried out the statistical analysis.

The collected models are stored in Bristol Dental Hospital and are available for further analysis by permission of the Consultant Orthodontists Group.

RESULTS

Participation

One hundred and fifty-six (77%) of the 204 consultant orthodontists contacted, enrolled in the audit. One hundred and forty (69%) submitted models within the audit period. Some models arrived after the audit period. These were scored and stored with the remainder of the casts but were not included in the audit.

Of the 48 who did not enrol:

- Eleven had retired or were about to retire.
- One was absent from work due to long-term illness.
- One was newly appointed and would not have completed personally treated cases in time. (Because the audit was prospective and ran over a period of one year, most new appointees were able to participate.)
- Sixteen consultants did not enrol because they had the wrong caseload profile. These were either NHS consultants working in tertiary referral centres, or academics within large teaching hospitals. Some academics did, however, participate and the audit illustrates the wide range in case-profiles of personal treatment being carried out by this type of consultant.
- Three consultants disagreed with the audit
- Three consultants expressed a lack of interest in participating in the audit.

Only 13 did not respond at all, despite two communications. This produced an overall response rate of 94%. Given that this is the first audit of its type, the authors felt that this was an excellent response and hope that those who did not participate this time will be encouraged to do so in the future. Of the 140 participants, 134 submitted the full six cases as requested, six sent fewer – for various reasons. This produced a total of 823 cases for analysis (1,646 sets of dental casts), which does provide a unique overview of current consultant orthodontic treatment outcomes.

Minimisation of random error and avoidance of bias.

All technicians scoring the models were calibrated in the PAR index. This means that the error of measurement is to a mean difference of less than two points, RMS less than five and with no systematic bias.¹⁶ Inter-scorer reliability was assessed initially by asking all the technicians involved to score six cases independently. In no case did the accepted PAR scores vary by more than one point between all the scorers. Any set of models that posed a problem during the study was scored by a second technician, and in the event of a difference in the score awarded, the two technicians would discuss and resolve the difference.

It was important to assess as far as is practicable that the records were reliable and that cases had not been 'cherry-picked'. The models were carefully scrutinized by one of the authors as an 'expert eye' and he was reassured that there were no signs that any had been 'doctored' (such as incorrect trimming in order to reduce a final overjet) to improve PAR scores. Furthermore, the general spread of PAR scores before and after treatment suggests that cases had not been specially selected.

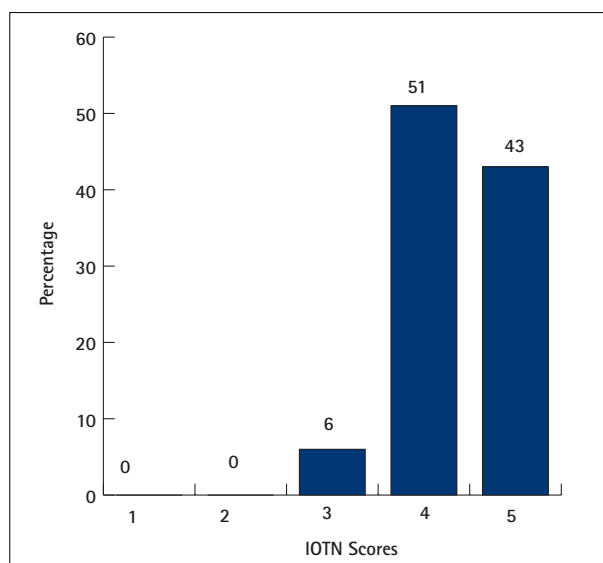
Analysis of the Index of Orthodontic Treatment Need

Sixteen consultants who submitted models either failed to provide information on the IOTN scores or sent incomplete data. The reason for this may warrant further investigation. Of the patients for whom complete scores were available, the majority had malocclusions that were in grades 4 (51%) and 5 (43%) pre-treatment. Only 6% were in grade 3 and none were in grades 1 or 2 (Fig. 1). The IOTN returns seem to support the view that most consultants limit their personal case-loads to cases of higher need.

Analysis of the peer assessment rating (PAR)

The changes in PAR score did not follow a normal distribution (bell curve), due to a few rogue results, especially those in which the PAR score actually increased.

Figure 1 Index of Orthodontic Treatment Need scores at the start of treatment expressed as percentage of patients in each grade.



Because of this fact, the use of parametric analysis would skew the mean to a value lower than which the data actually represent (Table 1). To provide a more accurate picture it was decided to use the median scores together with the interquartile range (Table 1).

The overall outcome of treatment was of a very high standard in relation to previously published results. A median percentage change in PAR score of 84% (interquartile range 71–91%) compares favourably with the findings of O'Brien *et al.*⁵ although it must be remembered that cases in the current audit were treated by consultants only. When plotted on the nomogram, 63% percent of cases fell into the 'greatly improved' category, 34% into the 'improved' category and 3% into the 'worse/no different' category (Fig. 2).

SETTING A NEW STANDARD

There would seem to be two decisions involved when setting standards for treatment outcomes.

- How best to quantify, in terms of PAR score, the complete failure to improve cases significantly – the 'worse/no different' description.
- How best to quantify and describe an acceptable degree of improvement for this case mix as a whole.

A standard for the 'worse /no different' category.

These cases represent 3% of this sample, which is within the previously suggested

standard by Richmond⁴ who suggested that less than 5% of cases should fall into this category. This encouraging figure was achieved even though 10% of cases were reported as being debonded early and despite the fact that some common problems such as ectopic canines are destined to achieve low scores. The only factor that would counsel against adopting 3% as a reasonable standard for this case mix and group of operators is the possibility that some cases which failed to complete had, knowingly or otherwise, been excluded. On balance, the evidence of this audit seems to support the adoption of 3% as a reasonable standard.

A standard for overall improvement.

The interquartile range for percentage change in PAR in this series was 71%–91%. Three quarters of the cases were therefore improved by more than 70% (as a rounded figure), and this seems a sound basis for a standard. Based on the quantitative evidence of this data therefore it is suggested that a standard for PAR score reduction be set for cases of this type when personally treated by consultant orthodontists and should be as follows:

- 75% of cases should exhibit a reduction in PAR score greater than 70%, with 3%, or fewer, cases having a reduction in PAR lower than 30%.

This standard excludes patients with clefts of the lip and palate, orthognathic surgery cases and oligodontia cases.

Table 1 Peer assessment rating (PAR) scores.

	Mean	Range	Median	Interquartile range
Pre treatment PAR score	33	5 to 68	34	26-41
Post treatment PAR score	7	0 to 45	5	3-9
Change in PAR score	27	-10 to 61	27	19-34
Percentage change	78	-73 to 100	84	71-9

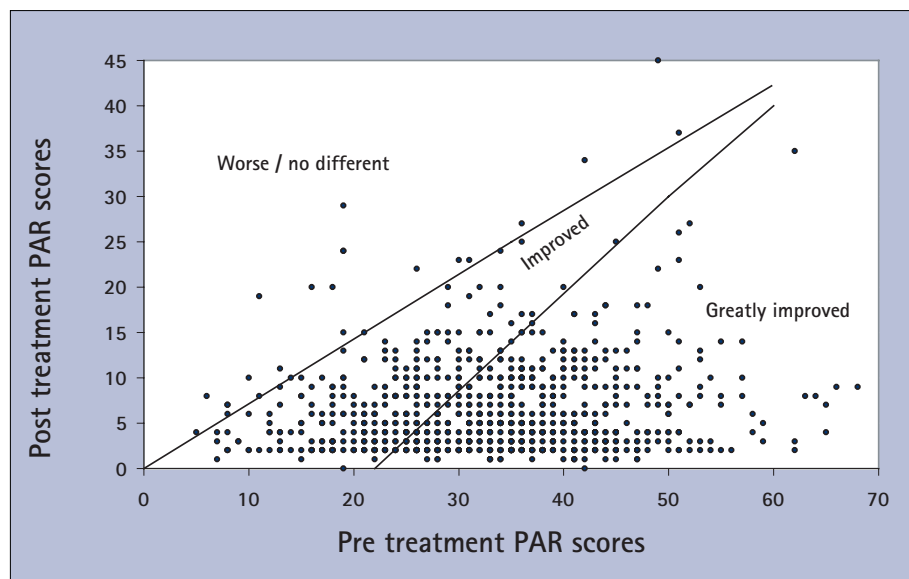


Figure 2 PAR nomogram to show changes in PAR score during treatment.

The non-participants

We deliberately asked for only 6 cases from each consultant as such a small number was intended to make the whole process non-threatening and practicable for participants. The more effort a study demands of the individual, the fewer people are likely to participate. Furthermore, meaningful analysis of an individual's performance cannot be made from such a small sample. This removes the perceived threat of a 'league table' and further encourages participation. In addition to the previous explanations in the protocol sent to all participants, the authors hoped to maximize the numbers enrolling in the audit.

Despite this 13 consultants did not respond and three claimed, 'lack of interest'. A further number enrolled in the audit but did not, eventually, deliver models. Did these consultants not participate because of the audit protocol, or is their default a reflection of their view of audit? The three who 'disagreed' were at least sufficiently constructive to put their thoughts in writing. Why did the others not respond? Can we ask them, and if we did, would they tell us? These are difficult questions that will be encountered by the medical and dental professions as a whole and will have to be faced as clinical appraisal and re-certification are implemented.

Plan to apply findings.

- The results have been presented and discussed at two general meetings of the Consultant Orthodontists Group. These occasions were designed to increase the understanding and consensus view of the results and their appropriate application.
- The Consultant Orthodontists Group committee have been asked to send the proposed new standards to all consultants.
- It is envisaged that these standards will be used as part of individual consultants' appraisals.
- Support and encouragement will continue, through the professional bodies, towards developing new standards of increased validity to audit treatment outcomes in orthognathic and oligodontia cases.

The authors wish to acknowledge the help of the many people, without whom this audit would not have been possible.

The Consultant Orthodontists Group of the British Orthodontic Society, which provided the funding, and for the hard work of the past Hon. Sec, Miss Helen Knight, who compiled the original main consultant orthodontist list.

The Clinical Effectiveness Committee of the Royal College of Surgeons of England, for their support, invaluable advice and encouragement.

Mr Mark Shields, for the IT support, and the setting up, running and retrieval of information from the audit database.

Mrs Jacinta McDermott, for the administrative support, the data entry to the audit database, and the fielding of endless calls from 'confused' consultant orthodontists.

The Bristol Dental Hospital Orthodontic Laboratory, for all the PAR Index analysis. Their interest and support added to the validity of the whole project.

The technicians involved were:

Peter Gleed, Chief Tech. Orthodontics

Rob Edge, Senior Tech.

Ashley Hunt, Senior Tech.

Michelle George, Senior Tech.

Katie Spicer, Senior Tech.

Mark McGrath, Senior Tech.

Most of all, the authors would like to thank the consultant orthodontists who responded to the audit, and who offered encouragement and at times humour to the project. Their vision of the importance of clinical audit and personal appraisal to the quality of care of their patients, made this audit possible.

- 1 *The report of the public inquiry into children's heart surgery at the Bristol Royal Infirmary 1984-1995: Learning from Bristol. (Cm 5207) The Stationary Office. July 2001*
- 2 Brook PH, Shaw WC. The development of an index for orthodontic treatment priority. *Eur J Orthod* 1989; **11**: 309-332.
- 3 Burden D J. Need for orthodontic treatment in Northern Ireland. *Community Dent Oral Epidemiol* 1995; **23**: 62-3
- 4 Richmond S, Shaw W C, Roberts C T, Andrews M. The PAR Index (Peer Assessment Rating): methods to determine outcome of orthodontic treatment in terms of improvement and standards. *Eur J Orthod* 1992; **14**: 180-187.
- 5 O'Brien K D, Shaw W C, Roberts C T. The use of occlusal indices in assessing the provision of orthodontic treatment by the Hospital Orthodontic Service of England and Wales. *Br J Orthod* 1993; **20**: 25-35.
- 6 Richmond S. Personal audit in orthodontics. *Br J Orthod* 1993; **20**, 131-145.
- 7 Fox N A. The first 100 cases; a personal audit of orthodontic treatment assessed by the PAR (Peer Assessment Rating) Index. *Br Dent J* 1993; **174**: 290-297.
- 8 Taylor P J S, Kerr W J S, McColl J H. Factors associated with the standard and duration of orthodontic treatment. *Br J Orthod* 1996; **23**: 335-341.
- 9 Buchanan I B, Russell J I, Clark J D. Practical application of the PAR Index: an illustrative comparison of the outcome of treating using two fixed appliance techniques. *Br J Orthod* 1996; **23**: 351-357.
- 10 Birkeland K, Furevik J, Boe O E, Wisth P J. Evaluation of treatment and post-treatment changes by the PAR Index. *Eur J Orthod* 1997; **3**: 181-185.
- 11 al Yami E A, Kuijpers-Jagtman A M, van't Hof M A. Occlusal outcome of orthodontic treatment. *Angle Orthod* 1998; **68**: 439-444.
- 12 Burden D J, McGuinness N, McNamara T. Treatment outcome for a sample of patients with Class 11 division 1 malocclusion treated at a regional hospital orthodontic department. *J Ir Dent Assoc* 1998; **44**: 67-69.
- 13 Riedman T, Berg R. Retrospective evaluation of the outcome of orthodontic treatment in adults. *J Orofac Orthoped* 1999; **60**: 108-123.
- 14 Mars M, Plint D A, Houston W J B, 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**: 314-322.
- 15 Atack N E, Hathorn I S, Semb G, Dowell T, Sandy J R. A new index for assessing the surgical outcome in unilateral cleft lip and palate subjects aged 5 - reproducibility and validity. *Cleft Palate-Craniofac J* 1997; **34**: 242-246.
- 16 Richmond S. Personal communication, July 2002.