

ED Hannapel  
LE Johnston, Jr.

# Extraction vs. non-extraction: PAR-score reduction as a function of initial susceptibility

## Authors' affiliations:

Eric D. Hannapel, In private practice,  
Caledonia, MI, USA  
Lysle E. Johnston, Jr.,  
Robert W. Browne Professor of Dentistry, and  
Chair, Department of Orthodontics and  
Pediatric Dentistry, University of Michigan,  
Ann Arbor, MI, USA

## Correspondence to:

Prof. Lysle E. Johnston, Jr  
Department of Orthodontics  
and Pediatric Dentistry  
School of Dentistry  
The University of Michigan  
Ann Arbor, MI 48109-1078  
USA  
Tel: +1 734 647-3814  
Fax: +1 734 763-8100  
E-mail: lejlr@umich.edu

**Abstract:** The purpose of this study was to compare the dental effects of extraction and non-extraction edgewise in diagnostic subgroups of known susceptibility. Pretreatment records from a sample of 238 former Class II edgewise patients who could be located and who agreed to participate were subjected to discriminant analysis. As judged by the variables that entered the resulting discriminant function, the extraction decision at the time of treatment (1969–1980) was based largely on dental crowding and protrusion. Based on the continuum of standardized discriminant scores that resulted, four diagnostic subgroups were formed: clear cut extraction and clear cut non-extraction groups at the extremes of the distribution, and borderline extraction and borderline non-extraction groups in the centre (discriminant scores near zero). Ultimately, 125 of the 238 patients agreed to return for follow-up records. On average, the post-treatment interval was about 15 years. The so-called peer assessment rating (PAR, British weighting) was used to evaluate immediate post-treatment and longterm dental improvement. On application to subgroups of graded, empirically appropriate susceptibility, both treatments produced a marked reduction in PAR scores. This improvement was similar to that reported elsewhere for comparable, but less precisely characterized samples. The present extraction patients had somewhat higher initial PAR scores and thus had the least potential to benefit most from treatment. This potential for improvement appears to have been realized by premolar extraction. In contrast, the literature holds out little hope that a non-extraction treatment could have done as well. Accordingly, if the protrusive, crowded patient has the most to gain (and the most to lose in terms of regret), then the speciality has an obligation to view with caution and concern the growing popularity of non-extraction at all costs treatments. For a ponderable segment of our patient pool, it would be a pity to abandon treatments that can do the most in favor of those that can do the least.

**Key words:** extraction; non-extraction; patient's regret; peer assessment rating (PAR)-score improvement; susceptibility

## To cite this article:

*Prog. Orthod.* 3, 2002; 17–21  
Hannapel ED, Johnston LE Jr.:  
Extraction vs. non-extraction: PAR-score reduction as  
a function of initial susceptibility

Copyright © SIDO 2002

ISSN 1399-7513

## Introduction

For any given malocclusion there are a host of treatments that compete for attention and market share. To ask whether and under what circumstances any of them ‘works’ seems almost beside the point. Indeed, if orthodontic history teaches anything, it is that all treatments work well enough to permit survival in clinical practice. Accordingly, the speciality today seems to be looking for a simple, popular, orthodontist-friendly treatment that can be used for all malocclusions. But can a single ‘treatment’ – extraction or non-extraction; early or late; one stage or two – be optimal or even adequate for all patients? Conversely, can the menu of available therapies be so interchangeable that, for a single patient, a treatment can be chosen purely on the basis of convenience?

From the standpoint of the individual, it seems almost a mathematical certainty that all treatments cannot be equally effective. Stated simply, some treatments probably produce better results than others. Thus, for each patient/malocclusion, there may well be one best strategy and  $n-1$  less desirable alternatives. But how are we to decide among them? It is easy to tell when treatments are different; better, however, is quite another matter. In the end, it may boil down to ‘whose ox is to be gored’.

At the dawn of the 21st century, treatments are supposed to be evidence-based and patient centered. Although the call for ‘evidence based’ treatments seems to have raised a host of red flags (1), it is hard to argue (at least in public) against adopting ‘patient centered’ as a gold standard, bed rock criterion. For many, however, such a focus would represent a major paradigm shift. Rather than ease of application and practice efficiency (‘quicker, cheaper, better – any two will do’), it suggests that we pay more attention to criteria that are of significance to the patient. Nowhere is the potential conflict between the concerns of the orthodontist and the needs of the patient more obvious than in the controversy surrounding ‘extraction’.

Some argue that the extraction of first premolars (‘extraction’) has a drastic negative impact on the profile, whereas others say that the effect is, on average, only 2 mm or so. One school (2) of thought has adopted the mantra that non-extraction treatments produce ‘full complete smiles with nice full dental arches’, while another (3) claims that patients want profiles that are really quite ‘flat’. The way a treatment interacts with the facial profile is controversial, both in terms of what a given treatment does and what actually needs to be done. In the current incarnation of the extraction ‘debate’, dental stability seems to have taken a back seat to a discussion of the impact on the face and smile. Recently, however, the development and increasing

popularity (at least with researchers) of the so-called peer assessment rating (PAR) has served to direct attention once more to the dental effects of treatment – overbite, overjet, midline discrepancy, right and left buccal occlusion, and upper and lower anterior alignment (crowding, spacing, impaction) (4).

The correction of crowding requires space, and extraction is perhaps the surest source. Expansion, either dental or skeletal, has received a great deal of attention as a substitute for premolar extraction. Unfortunately at a rate of 0.7 mm of extra perimeter for each mm of expansion, it would take approximately 12 mm of stable expansion to equal the space created by the extraction of four first premolars followed by careful space-closing mechanics (5). Unfortunately, although expansion is easy, there is little evidence that the changes it produces, especially the hoped for increases in arch perimeter, are stable. The lack of an effective substitute for extraction, rather than some sort of change in the public’s perception of facial beauty, appears to have led to the convenient claim that people prefer the ‘toothy’ effect that contemporary non-extraction, non-compliance, preadjusted treatments tend to produce.

Freud said that ‘biology is destiny’; for patients, the choice of treatments is their destiny. Because each regime produces a restricted range of outcomes (5), it can be argued that a given treatment may be appropriate only to a limited spectrum of patients. Thus, the use of even a reliable, popular, objective index to study the dental effects of, say, ‘early treatment’ (6) or to compare extraction and non-extraction edgewise (7) may produce results that are difficult, if not impossible, to interpret. If a study’s groups are formed purely on the basis of personal treatment decisions, it is impossible to know not only what sort of malocclusions have been studied (8), but also the type of patient to whom the results apply.

The purpose of this simple investigation, therefore, was to examine short- and long-term change in PAR scores in carefully defined diagnostic subgroups of extraction and non-extraction Class II patients. The goal of this approach was to characterize the interaction between malocclusion and treatment strategy. It is hypothesized that PAR score improvement will be a function both of the choice of treatment and the degree of pretreatment crowding and protrusion.

## Patients and methods

The patients studied here and the statistical methods employed to define the various samples have been described in greater detail elsewhere (9–12). Following approval by the Institutional Review Board of Saint Louis University Medical Center, at least five attempts

were made to contact all edgewise Class II, division 1 first-premolar extraction and non-extraction edgewise patients treated in the graduate orthodontic clinic between 1969 and 1980. These treatments were effected by way of standard, 022' non-preadjusted edgewise appliances in conjunction with various combinations of anchorage preparation, Class II elastics, and extra-oral traction, as needed. Because of the passage of 10–20 years since the completion of treatment, most subjects could not be reached; however, in the end, 238 – approximately one in 10 – were located and agreed to participate. These potential subjects were told that they would be contacted for followup records should they meet the empirical admission criteria that would be generated from an analysis of their initial models and lateral cephalograms.

To characterize the pretreatment differences between the extraction and non-extraction patients, a total of 89 linear and angular measurements were generated and then subjected to discriminant analysis. A linear combination of six of these characteristics – maxillary crowding and protrusion; mandibular crowding, protrusion, and irregularity; profile convexity – provided a highly significant ( $p < 0.01$ ) discrimination between the 132 extraction and the 106 non-extraction patients. Based on standardized discriminant scores, the ex-patients were divided into four groups from which 125 of the original 238 agreed to return for followup records: 33 clear-cut extraction, 33 borderline extraction, 29 borderline non-extraction, and 30 clear-cut non-extraction. The post-treatment interval averaged about 15 years.

The PAR was used to assess the dental effects of treatment on each of the 125 sets of study models. This index was developed to assess malocclusion at any stage of dental development and thereby to measure the success of orthodontic treatment in terms of dentoalveolar change (4). Because of its relatively complete characterization in the literature, the British-weighted version was used here. The senior author was trained in the use of the PAR Index at The Ohio State University. The calibration process was supervised by Dr Kevin D. O'Brien of the University of Manchester, UK. A subsequent reliability exercise on 30 study models produced an intraclass correlation of 0.89. The start (T1), finish (T2), and recall (T3) study models of each of the 125 Class II patients were scored (in random order) according to the British weighting of the PAR Index.

Standard descriptive statistics (means and standard deviations) were calculated for the T1, T2, and T3 study models. According to common usage, a reduction in PAR score of 30% is considered to be 'improved'; 22 points, 'greatly improved'. Based on this categorization,  $\chi^2$  contingency tests were used to test the relationship between treatment and short- and long-term improvement in PAR score.

## Results

Descriptive statistics for pretreatment, post-treatment, and recall PAR scores are summarized in Table 1. It may be noted that the clear-cut extraction patients had the highest PAR scores at the outset; however, after treatment and later on recall, the scores for all four groups were, on average, low and quite similar. The distribution of improvement categories – greatly improved, improved, worse/no different – among the four treatment groups is summarized in Tables 2 & 3. As may be seen, very few of the 125 patients failed to show some sort of improvement. Contingency tests were used to assess the relationship between treatment strategy and categories of PAR-score improvement. In these  $\chi^2$  tests, the 'improved' and 'worse/no different' cells were combined to achieve expected frequencies of at least five. At the end of treatment and perhaps to a degree at recall, the extraction treatments showed disproportionately more subjects in the greatly improved category:  $\chi^2 = 11$  ( $p < 0.01$ ) and 7 ( $0.05 < p < 0.10$ ), respectively.

## Discussion

The various groups studied here were defined by standardized discriminant scores. Ultimately, therefore, they were formed by treatment plans devised some 20–30 years ago. At this time, few would have contemplated using a single treatment for all patients. Some patients were seen as being susceptible to extraction; others, to non-extraction. For the groups of patients studied here, the variables that entered the discriminant function imply that the extraction/non-extraction decision was based largely on crowding and protrusion. Accordingly, although extraction has since fallen out of favor with many orthodontists and their referral sources, the four diagnostic subgroups studied here have a rational basis and thus are at least of heuristic significance, especially when viewed from the standpoint of the patient.

It is clear from the similarity of the recall PAR scores that both the extraction and non-extraction edgewise treatments were effective in producing short- and long-term improvements in the dental characteristics measured by the PAR Index. Indeed, the improvements seen here are similar to those reported for two-phase functional/fixed treatments (6) and for various kinds of extraction and non-extraction treatments in samples containing a mixture of Angle classes (7). These findings, however, do not mean that the various treatment strategies are interchangeable. They mean only that both extraction and non-extraction are effective in patients who are susceptible to the repertoire of changes that each tends to produce.

**Table 1. Descriptive statistics: means and standard deviations (SD) by stage of treatment**

Group	n	Pretreatment		Post-treatment		Recall	
		Mean	SD	Mean	SD	Mean	SD
Non-extraction							
Clear-cut	29	28.3	8.2	3.9	4.2	10.1	6.3
Borderline	30	31.5	8.3	5.1	4.6	12.5	9.2
Extraction							
Borderline	33	32.7	7.4	5.2	3.9	12.1	6.1
Clear-cut	33	37.7	8.8	5.6	4.1	11.1	7.3

**Table 2. Treatment change in peer assessment rating (PAR) score, by groups**

Group	Greatly improved	Improved	Worse/no different
	n (%)	n (%)	n (%)
Non-extraction			
Clear-cut	17 (59)	12 (41)	0 (0)
Borderline	21 (70)	9 (30)	0 (0)
Extraction			
Borderline	28 (85)	4 (12)	1 (3)
Clear-cut	30 (91)	3 (9)	0 (0)

**Table 3. Overall change in peer assessment rating (PAR) score, by groups**

Group	Greatly improved	Improved	Worse/no different
	n (%)	n (%)	n (%)
Non-extraction			
Clear-cut	11 (38)	16 (55)	2 (7)
Borderline	14 (47)	12 (40)	4 (13)
Extraction			
Borderline	16 (49)	14 (42)	3 (9)
Clear-cut	23 (70)	10 (30)	0 (0)

Patients differ in their susceptibility to a given regimen. Patients with ‘straight’ profiles and no crowding, for example, seem particularly suited to some form of non-extraction treatment. In the present study, however, the two ‘borderline’ groups had discriminant scores that clustered around zero and thus were morphologically similar prior to treatment. For these patients, it seems to have made little difference dentally which treatment was employed. As reported elsewhere (11), however, these same two borderline groups showed modest, but noticeable and statistically significant, differences in profile protrusion that averaged about 2 mm. In this intermediate type of patient therefore the choice of treatment would seem to depend more

on facial aesthetics than on dental characteristics. In this context, it should be noted that patients, themselves, may prefer profiles that are considerably ‘flatter’ than is commonly assumed by those who wish to avoid the rigors of extraction mechanics, both for themselves and for their patients (3).

In contrast to the similarity characteristic of the borderline groups, patients in the two ‘extreme’ groups differed greatly with respect to crowding and protrusion (11, 12). Both the discriminant scores that defined the groups and the present estimates of PAR-score improvement in the extreme extraction group argue that it is the severely crowded and protrusive patient who has the most to gain from orthodontics. From the standpoint of the PAR Index, a patient with a straight profile and little crowding has a limited potential for improvement. Indeed, it takes an initial score of 22 even to be eligible to be ‘greatly improved’ (4). In the present study, it was the extraction patients who had the highest initial PAR scores and who were thus eligible for the greatest improvement. The fact that they actually did undergo the greatest improvement is perhaps significant.

Both short- and long-term, the present results argue against the common, word-of-mouth claim that ‘extraction’ ruins dentition. It suggests instead that it is the clear-cut extraction patient (that is, one who presents with marked protrusion and crowding) for whom orthodontics can do the most. A problem with this inference is that history never reveals its alternatives: perhaps all of the patients would have seen the same reduction in PAR score had they been subjected to the non-extraction alternative. If treatments are interchangeable, why not use the one that is easiest and most popular with patients and referral sources? Unfortunately, the literature provides scant justification for the belief that extraction and non-extraction treatments are interchangeable in application to the extremes.

Peer assessment rating is a dental index. Non-extraction treatments have been shown to have only a limited capacity to correct the crowding and protrusion that contributed to the initial PAR scores of the present extraction patients (3, 13). In contrast, premolar

extraction makes incisor retraction possible and thus can produce a long-term correction of at least some incisor irregularity. Thus, at least from the standpoint of the PAR score, it probably cannot be argued that non-extraction could have produced a comparable improvement in the two extraction groups. If this be so, it is necessary to consider the concept of 'regret', the difference between what a patient gets and what he/she could have had, given the best available treatment.

If there is no effective substitute for premolar extraction, if non-extraction treatments have only a limited capacity to correct crowding and protrusion, and if extraction really is bad, then there may well be a sizeable sub-population of patients for whom orthodontics can do little. The literature, however, contains little support for the common, crude indictments of conventional 'extraction' therapy. Indeed, the present study argues that, compared with outcomes seen in non-extraction patients, premolar extraction can produce dental results that are as good or perhaps even a bit better in a spectrum of patients for whom non-extraction probably can do little. At the very least, our results stand as a reference standard for those who seek to apply non-extraction treatments to the extraction extremes and vice versa. The long-term data should prove interesting.

In the meantime, the present communication advances the hypothesis that the orthodontist's greatest opportunity for success may well be the 'extraction' patient who is treated by extraction; the greatest potential for regret, the extraction patient who is not.

## Conclusions

When matched according to empirical measures of patient susceptibility, both extraction and non-extraction treatments produce comparable PAR-score improvement, both short- and long-term. Clear-cut extraction patients, however, seem to have the most to gain from treatment. The present study argues that, for these patients, extraction treatments stand the best chance of meeting this need and hence of fulfilling the specialty's potential to effect meaningful change.

**Acknowledgements:** Portions based on a thesis submitted by Dr Hannapel in partial fulfillment of the requirements of the degree of master of science, Horace H. Rackham School of Graduate Studies, The University of Michigan.

Supported in part by N.I.D.C.R. grant DE08716 and by donations from the Orthodontic Education and Research Foundation (St Louis) and from friends of the University of Michigan Department of Orthodontics and Pediatric Dentistry.

## References

1. Roth RH. Evidence based orthodontics. *Dento-Fac Expressions* 2001;2:2-3.
2. Stoner CC. An interview with AAFO's man of the year 1984 Dr John W Witzig. *Funct Orthod* 1984;1:9-19.
3. Bowman SJ, Johnston LE Jr. The aesthetic impact of extraction and non-extraction treatment on Caucasian patients. *Angle Orthod* 2000;70:3-10.
4. Richmond S, Shaw WC, O'Brien KD, Buchanan IB, Jones R, Stephens CD, Roberts CD, Andrews M. The development of the PAR Index (Peer Assessment Rating): reliability and validity. *Eur J Orthod* 1992;14:125-39.
5. Johnston LE Jr. A comparative analysis of Class II treatments. In: Vig PS, Ribbens KA editors. *Science and Clinical Judgment in Orthodontics*. Ann Arbor, MI, USA: Center for Human Growth and Development; 1986. pp. 103-48.
6. McKnight MM, Daniels CP, Johnston LE Jr. A retrospective study of two-stage treatment outcomes assessed with two modified PAR indices. *Angle Orthod* 1998;68:521-4.
7. Holman JK, Hans MG, Nelson S, Powers MP. An assessment of extraction versus non-extraction orthodontic treatment using the peer assessment rating (PAR) index. *Angle Orthod* 1998;68:527-34.
8. Luke LS, Atchison KA, White SC. Consistency of patient classification in orthodontic diagnosis and treatment planning. *Angle Orthod* 1998;68:513-20.
9. Johnston LE Jr, Paquette DE, Beattie JR, Cassidy DW Jr, McCray JF, Killiany DM. The reduction of susceptibility bias in retrospective comparisons of alternative treatment strategies. In: Vig KD, Vig PS, editors. *Clinical Research as the Basis of Clinical Practice*. Ann Arbor, MI, USA: Center for Human Growth and Development; 1991. pp. 155-77.
10. Johnston LE Jr. A comparative analysis of Class II treatments: a retrospective/prospective alternative. *Clin Orthod Res* 1998;1:142-6.
11. Paquette DE, Beattie JR, Johnston LE Jr. A long-term comparison of non-extraction and bicuspid-extraction edgewise therapy in 'borderline' Class II patients. *Am J Orthod Dentofac Orthop* 1992;102:1-14.
12. Luppannapornlarp S, Johnston LE Jr. The effects of premolar-extraction treatment: a longterm comparison of outcomes in 'clear-cut' extraction and non-extraction Class II patients. *Angle Orthod* 1993;63:257-72.
13. Scott S, Johnston LE Jr. The perceived impact of extraction and non-extraction treatments on matched samples of African American patients. *Am J Orthod Dentofac Orthop* 1999;116:352-8.