Evaluation of orthodontists' perception of treatment need and the peer assessment rating (PAR) index

Susan P. McGorray, PhD; Timothy T. Wheeler, DMD, PhD; Stephen D. Keeling, DDS, MS; Lisa Yurkiewicz, DMD, MS; Marie G. Taylor, RDH; Gregory J. King, DMD, DMSc

Abstract: This paper examines the relationship between orthodontists' subjective assessment of treatment need and objective measurements obtained during standardized intra- and extraoral examinations. Logistic regression modeling was used to develop predictive models of treatment need. Data were obtained from 1155 eighth-grade students by four orthodontists who used standardized examination forms to assess demographics, trauma, skeletal relationships, morphologic malocclusion traits, and mandibular function. At the conclusion of the examination, the orthodontist rated the subjective treatment need as none, elective, recommended, soon, or immediate. For some analyses, the categories were collapsed to represent no need and need. The peer assessment rating (PAR) index (American validated version) was computed from the clinical exam findings and scoring of dental models; PAR scores were used to document malocclusion severity and treatment difficulty. Spearman rank correlation coefficients quantified the relationship between PAR scores and need categories. Logistic regression analysis modeled treatment need using components of the PAR index as well as other variables. The components of these models, as well as sensitivity and specificity, were compared with malocclusion severity/treatment difficulty scores obtained from malocclusion assessments using the PAR index. The five subjective treatment need categories and the PAR index scores were significantly correlated (rho=0.62, $p<0.001$). Significant differences were detected between the need and no need groups for all PAR components ($p<0.001$). PAR index scores and predicted probabilities from logistic regression models performed equally well for classification purposes (no need, need). The data suggest that the PAR index is highly correlated with orthodontists' subjective assessment of treatment need when that assessment is made in the absence of financial considerations and patient desires.

Key Words: Treatment need, Malocclusion severity, Peer assessment rating index, Sensitivity, Specificity

Historically, the definition and assessment of malocclusion have been problematic. Dentists' opinions as to the need for orthodontic treatment and the outcome of treatment show wide variation.[1] The nature of orthodontic need assessment is difficult to define and quantify because the dental, functional, and psychosocial benefits derived from orthodontic care are largely unknown.[2] Malocclusion is not an acute condition but a deviation of occlusal and skeletal components from normal, with deviations of individual components often unremarkable in and of themselves;[3] thus, practitioners' and patients' perceptions of treatment need are affected by many different variables.[4] What these variables are and how they are used by individual orthodontists to determine treatment need is unknown. Nevertheless, orthodontists routinely make recommendations to treat or not.

Numerous malocclusion indices have been developed to quantify aspects of morphologic malocclusion. Their number alone suggests the inherent complexity of the problem and the concern with malocclusion as a national health issue.[5,6] Although no one measure or index is likely to be developed that is considered ideal for all purposes,[5] accurate, valid, and reliable assessments of malocclusion are necessary for determination of

Author Address
Dr. Timothy Wheeler
Professor & Chairman
P.O. Box 100444
University of Florida
College of Dentistry
Gainsville, FL 32610-0444

Susan P. McGorray, Department of Statistics, Division of Biostatistics, College of Liberal Arts and Sciences, University of Florida.
Timothy T. Wheeler, Department of Orthodontics, University of Florida, College of Dentistry.
Stephen D. Keeling, Department of Orthodontics, University of Florida, College of Dentistry.
Lisa Yurkiewicz, Department of Orthodontics, University of Florida, College of Dentistry.
Marie G. Taylor, Department of Orthodontics, University of Florida, College of Dentistry.
Gregory J. King, Department of Orthodontics, University of Washington, School of Dentistry.

Submitted: February 1998 **Revised and accepted:** January 1999
Angle Orthod 1999;69(4):325-333.

treatment need priority, allocation of limited resources, and assessment of treatment outcomes.[7]

Shaw et al.[8] proposed that there are five types of indices, each with a distinct purpose. First, there is a diagnostic classification, such as the Angle classification. Second, there are epidemiological indices, such as the occlusal index.[9] A third type of index is that which measures treatment need, such as the handicapping labiolingual deviation index,[10] the orthodontic treatment priority index,[11] the index of orthodontic treatment need (IOTN),[12] and the dental aesthetic index (DAI).[13] A fourth type of index measures treatment outcome, such as the peer assessment rating (PAR) index.[14,15] A last type of index would measure treatment complexity, although this type of index does not currently exist. All these indices share one feature in common: they are derived by scoring various components of malocclusion from study models; only the IOTN and the DAI consider patient's ranking of perceived esthetic impairment. None directly considers the objective measurement of functional, skeletal, or psychosocial parameters. The use of indices provides the opportunity to reduce subjectivity and standardize the criteria upon which judgements are made.[12]

The IOTN and the PAR index have drawn considerable interest during the past several years as resource allocators and clinical researchers struggle with the difficulties of determining need (e.g., allocating state Medicaid funds) and outcomes of treatment (e.g., clinical trials research). Both indices were developed in England and have been evaluated for reliability and validity; they are being used in the United Kingdom in orthodontic research and administration of public health service orthodontics.[14] The IOTN ranks malocclusion in terms of the significance of various occlusal traits for an individual's dental health and for perceived esthetic impairment; the PAR index assesses treatment difficulty and malocclusion severity.[12] DeGuzman et al.[16] validated the PAR index on a group of U.S. orthodontists to reflect contemporary American orthodontic opinion, by deriving modified weightings for the various components of the index. The percent change in the PAR index has been used as a tool to measure the effectiveness of orthodontic treatment in dento-occlusal changes.[17]

Orthodontists in private practice in the U.S. have not employed standardized indices to determine treatment need; the decision to treat rests on less well-defined criteria. Orthodontic treatment is often performed when elective or no-need situations are present, emphasizing that demand and need are distinct entities.[18] How well does malocclusion severity alone account for orthodontists' assessment of treatment need? This paper examines the relationship between orthodontists' subjective assessment of treatment need and objective measurements obtained during standardized intra- and extraoral clinical examinations. Logistic regression modeling is used to develop predictive models of treatment need. The components of these models, as well as sensitivity and specificity, are compared with malocclusion severity/treatment difficulty scores obtained from malocclusion assessments using the PAR index.

Materials and methods

Subjects

This investigation was conducted at seven public middle schools in Alachua County, Florida, during the months March through May, 1994. These schools were racially and economically diverse, providing variability with respect to the factors of interest in this study. A convenience sample was drawn from the eighth-grade students who had health permission screening forms on file with the school and were present on the day(s) of the screening. A total of 1155 out of 1606 eligible subjects (71.9% of the target sample) were examined. Five students with parental letters of excuse refused participation in the study. Subjects were not selected on the basis of desire for orthodontic treatment, and data describing desire for treatment were not obtained. All subjects were examined, regardless of prior or current orthodontic treatment.

Data collection

This study met the University of Florida criteria for exemption from full review by the Institutional Review Board for Protection of Human Subjects. Permission to screen seven of eight public middle schools was granted by the Superintendent of the Alachua County School Board and the respective school principals.

Depending on school size and schedule conflicts, the number of site visits to each school varied from one to five. Examinations were performed on the school campus in a separate room (typically, a vacant classroom) with the student standing erect in front of the examiner. Each examiner used a millimeter rule, gloves, a tongue depressor for cheek retraction, and a small piece of folded baseplate wax for occlusal impressions and bite registration. Each child was examined by one of the four orthodontists, and all examiners did not visit each site. The mean number of children seen by an examiner was 288, with a range of 118 to 535.

As part of the school screening, data were collected for each student and recorded on standardized forms. Information obtained by a staff assistant included demographic data, history of trauma or previous orthodontic treatment, and the child's perception of treatment need. The examining orthodontist recorded clinical variables and a subjective assessment of the child's orthodontic treatment need.

Demographic information used in this report includes:

1. Age at examination date
2. Sex
3. Race
4. History of previous orthodontic treatment (unilateral space maintainer was classified as no previous treatment)
5. Screening orthodontist
6. School

Clinical information includes:

1. Overjet (mm)
2. Overbite (classified in categories, ranging from 0= none to 4= > 100%)
3. Openbite (classified in categories, ranging from 0= none to 4= greater than or equal to 3 mm)
4. Molar relationship (each side coded in cusp units, ranging from 0 (>FCII) to 9 (>FC III), with 5 representing Class II)
5. Midline deviation

Additional variables required in the PAR index were recorded from models made from the wax bite impressions. These variables include:

1. Upper and lower anterior alignment
2. Right and left crossbite
3. Right and left overbite

This impression method was determined to be valid for the above measures (unpublished).

Table 1
American Validation of the PAR Index, from DeGuzman et al.

Components (and range*)	Raw	Weightings used to measure: Malocclusion severity	Treatment difficulty	Combination
1. Upper anterior alignment (0-25)	1	1	1	1
2. Buccal occlusion (0-14)	1	2	2	2
3. Overjet (0-8)	1	5	4	4.5
4. Overbite (0-7)	1	3	3	3
5. Midline discrepancy (0-2)	1	3	4	3.5

* Range values from Richmond et al.

Need for treatment

At the conclusion of the exam, the examining orthodontist determined the subjective orthodontic treatment need status of the child. Each child was classified into one of the following categories: (1) no treatment need, (2) treatment optional, (3) treatment recommended, (4) treatment need should be attended to soon, and (5) immediate treatment need. For some analyses, classifications were combined defining two subjective orthodontic treatment need (SOTN) groups, those with need (categories 3, 4, and 5) and those with no need (categories 1 and 2).

Reliability of examiners in scoring malocclusion traits and treatment need

We have previously published the interexaminer reliability of scoring individual malocclusion traits in children during a clinical exam by study examiners during school-based examinations conducted in 1990 and 1991.[19] Median Kappa statistics[20] from our previous work indicate that the reliability of maxillary and mandibular anteroposterior positions, incisor exposure, interlabial gap, and maxillary crowding is poor (K < 0.40). Acceptable reliability exists for mandibular anterior crowding, facial convexity, overbite, overjet, and molar classification (median Kappa statistics ranged from 0.48 to 0.72). Excellent reliability exists for evaluating the presence of a posterior crossbite (K = 0.79).

Data describing the interexaminer reliability of judging treatment need by our group have not previously been reported. Data are available from 16 children examined in the current study. These "reliability" students were randomly selected from among those students screened at each school, and they were examined along with the other students. At each session, the reliability students were examined by each of the orthodontists present at the school that day; the staff assistant directed these students for repeated examinations.

Malocclusion index assessment

Raw PAR index scores were determined for each child by assessing the five PAR components: upper anterior alignment, buccal occlusion, overjet, overbite, and midline deviation; these have been described in detail previously.[12,14,15] Upper anterior alignment was determined by scoring contact displacements. Both right and left buccal occlusion scores considered posterior sagittal interdigitation, posterior crossbite, and posterior openbite. The overjet score considered amount of overjet (positive and negative) and number of anterior teeth in crossbite. Overbite scores represented amount of overbite or openbite. The midline score represented the amount of deviation with respect to the lower incisor. In this study, data on posterior interdigitation, millimeters of overjet, and overbite/openbite were obtained from the clinical exam, performed by one of the four orthodontists. During the clinical exam, to match previously collected data, anterior alignment, posterior and anterior crossbites, and posterior openbite were scored differently from those scored when performing a PAR index evaluation. Thus, the PAR index components of anterior crowding, posterior and anterior crossbites, and posterior openbites were scored from plaster models, made from the wax bite impressions made during each exam. One examiner (LY), who had been trained by a PAR-calibrated examiner, made model scorings. (The PAR-calibrated examiner [SDK] had been trained by one of the developers of the PAR index [K. O'Brien] and tested for reliability against a set of 30 calibration models, used as a standard. The PAR-calibrated examiner's interclass correlation with the standard was 0.87.)

Nineteen models were randomly selected and scored by the PAR-calibrated examiner, with inter-rater reliability assessed.

The individual raw scores for each PAR component were multiplied by established weightings and then summed to establish the total PAR index score. These weightings represent distinct but related entities—malocclusion severity and treatment difficulty—as well as a third composite. Table 1 lists the possible range for each component and describes weighting factors used in this study to represent the current opinions of American orthodontists.[16]

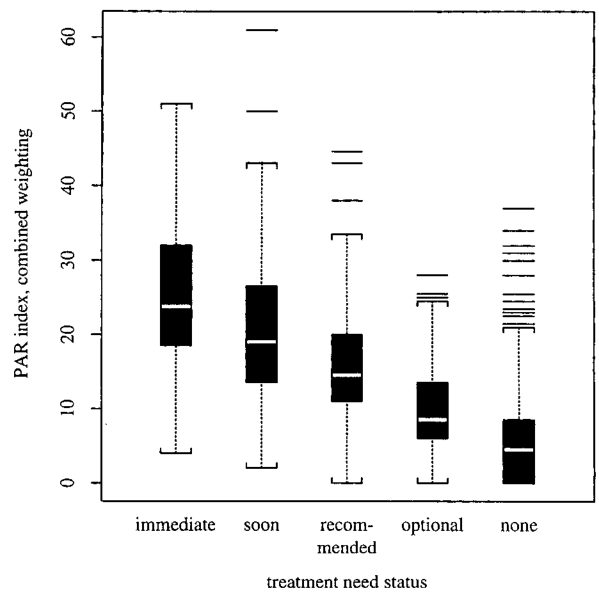

Figure 1
Boxplots show distribution of PAR index scores (combined weighting) for each subjective treatment need classification. Box represents 25th, 50th (white bar), and 75th percentiles, respectively; whiskers show extremes.

Statistical analysis

Interexaminer reliability of treatment need was summarized using percent exact agreement for the five-level subjective orthodontic treatment need variables and also for the dichotomized version. Pairwise comparisons between examiners were made using weighted Kappa statistics and one-sided Z tests.[19] These techniques were also used to evaluate agreement between the model scorer and the PAR-calibrated scorer.

Spearman rank correlation coefficients quantified the relationships between the PAR index scores and subjective orthodontic treatment need status. For further analysis, the 5-point need scale was dichotomized as need (immediate, soon, or recommended) and no need (optional or no need). Demographic and clinical variables were compared for the two groups. Wilcoxon rank sum tests were used to compare continuous or ordinal variable distributions and Chi-square tests were used to test for differences in proportions for categorical variables. For all statistical tests, a p-value of less than 0.05 was considered statistically significant.

Logistic regression modeling[21] was used to develop a predictive model for the dichotomized need/no need groups. Variables considered for inclusion in the initial model included components and subcomponents of the PAR index (e.g., the PAR component "overjet" has subcomponents overjet and anterior crossbite) plus additional reliable examination variables evaluated as part of this study. A number of coding schemes were considered for some variables. For example, posterior interdigitation was coded as buccal occlusion component score from the PAR index (0 - 7, left and right) and also as total amount of right and left discrepancy from Class I in quarter cusp units (0 - 10). This allows evaluation of different types of relationships between the outcome variable, treatment need, and possible predictor variables. A forward selection procedure was used to assist in determining the initial model. No variables were included a priori. A second interaction model included significant variables from the initial model plus significant interaction terms. To generate this model, all two-way interactions between variables selected in the initial model were considered for inclusion. After significant interaction terms were selected, all variables not in the initial model were again considered for inclusion. As a final step, demographic and screening variables (age, race, sex, screening orthodontist) were considered for inclusion in the two models. Goodness-of-fit of the models was assessed by Hosmer Lemeshow tests.[22]

From a regression model, it is possible to calculate a predicted probability of treatment need for each subject, based on the overall model and the subject's specific clinical values. This predicted probability of treatment need can range from 0.00 to 1.00. Different cutoff points can be specified to define predicted treatment need (classify as predicted treatment need any subject with a predicted probability of treatment need greater than 0.50).

To assess the classification capabilities of the regression models and the PAR index, sensitivity and specificity associated with possible cutoff points were examined. For sensitivity, we looked at the subjects that were actually judged to have treatment need and determined the percent predicted to have treatment need based on our cutoff point criterion. For specificity, we considered those subjects actually judged to have no treatment need and determined the percent that were predicted to have no need. For example, if our cutoff point was set at 0.00, classifying all those with a predicted probability of greater than or equal to 0.00 as having need (i.e., everyone), then our sensitivity was 100% but our specificity 0%. As the cutoff point increases, sensitivity decreases while specificity increases. A cutoff point is often selected that maximizes both sensitivity and specificity. The characteristics of the index or model determine how high this value might be. Sensitivity and specificity were calculated for the PAR index in a similar manner, with cutoff points for the PAR index score used in place of the predicted probability cutoff points. Using sensitivity and specificity, we compared the performance of the PAR index with the performance of the logistic regression model-based predicted need.

Table 2
Comparison of demographic, clinical, and PAR index components between SOTN groups

Variable	*No need* (none, optional) N=751	*Need (recommended,* soon, immediate) N=403	*p-value*
Demographic			
Sex			
% Female	47.0	45.2	0.550*
Race			
% Black	33.1	33.7	0.707*
% White	63.6	62.1	
% Other	3.3	4.2	
Midline			
%0-1/4	84.8	68.2	0.001*
%1/4-1/2	14.0	27.1	
%>1/2	1.2	4.7	
Age			
Mean (s.d.)	13.8 (0.6)	13.9 (0.6)	0.095**
Clinical			
Molar class discrepancy			
Mean (s.d.)	0.66 (1.54)	2.05 (2.44)	0.0001**
Overjet			
Mean (s.d.)	2.96 (1.35)	3.63 (2.23)	0.0001**
Overbite			
Mean (s.d.)	1.37 (0.66)	1.72 (1.09)	0.0001**
PAR			
Upper anterior alignment			
Mean (s.d.)	1.10 (1.57)	2.28 (2.87)	0.0001**
Buccal occlusion			
Mean (s.d.)	0.55 (1.13)	1.74 (1.91)	0.0001**
Overjet			
Mean (s.d.)	0.47 (0.68)	1.21 (1.18)	0.0001**
Overbite			
Mean (s.d.)	0.45 (0.64)	1.13 (1.06)	0.0001**
Midline discrepancy			
Mean (s.d.)	0.17 (0.41)	0.37 (0.57)	0.0001**

* Chi-square test
** Wilcoxon rank sum test

Results

Reliability of the five-category and binary classifications of subjective orthodontic treatment need was assessed on the 16 subjects reviewed by all four orthodontists. Pairwise percent exact agreement using the five-category scale ranged from 25% to 62.5% (median 46.9%). Weighted Kappa statistics ranged from 0.31 to 0.60 (median 0.51). Pairwise agreement using the binary classification of SOTN ranged from 68.8% to 93.8% (median 84.4%), with Kappa statistics ranging from 0.43 to 0.85 (median 0.69).

Agreement was fair to excellent between the PAR index-calibrated orthodontist and the model scorer for all components of the PAR index . Weighted Kappa statistics for upper anterior alignment, right and left posterior crossbite, and anterior crossbite were 0.81, 0.80, 0.85, and 0.42, respectively. In the 19 reliability models, neither scorer detected right or left openbites.

Spearman rank correlation coefficients between orthodontic treatment need status and weightings of the PAR index ranged from 0.61 to 0.62. Correlation was significant in all cases, with the various PAR weightings having limited impact on rank correlation. Subsequently in this paper, only the combination weighting of the PAR index will be used. The PAR index could not be calculated for 41 students due to missing data for one or more of the index components. The boxplot in Figure 1 shows the distributions of these combined weightings of the PAR index for the five categories of orthodontic treatment need status. The median values of the PAR index increase as the need status scale increases.

Comparisons of demographic, clinical, and PAR index components between the two SOTN groups are presented in Table 2. The groups did not differ significantly with respect to age, race, or sex. Significant differences between the two groups were detected for all clinical variables and all PAR index components.

Subjects with complete data for all variables selected in the regression models numbered 1117, with 34.2% having SOTN. Logistic regression models were fit modeling SOTN (need, no need) as a function of the selected covariates. Variable definitions, odds ratio estimates, and 95% confidence intervals for the two models are presented in Table 3. Both models fit the data well, with no significant lack of fit detected (initial model p=0.34, interaction model p=0.51). After identifying important clinical characteristics, the inclusion of additional variables representing sex, race, age, and examining orthodontist did not significantly improve either model.

In the initial model, the odds ratios may be interpreted for each unit of change in a covariate component; the odds of being classified as having SOTN are increased by a factor of the odds ratio estimate.

Table 3
Logistic regression results for predicting treatment need and no need groups. Odds ratio etstimates and 95% confidence intervals are presented

Variable	Initial model	Interaction model
Upper anterior alignment*	1.25 (1.16,1.38)	1.27 (1.16, 1.38)
Openbite*	3.11 (2.14, 4.51)	3.27 (2.24, 4.78)
Overbite*	2.80 (2.21, 3.55)	2.88 (2.26, 3.68)
Overbite for 33 to 66%	0.54 (0.36, 0.79)	0.53 (0.36, 0.78)
Molar class discrepancy	1.96 (1.57, 2.44)	2.18 (1.75, 2.73)
Molar class discrepancy, squared	0.95 (0.92, 0.98)	0.94 (0.91, 0.97)
Lower anterior alignment*	1.26 (1.15, 1.39)	1.28 (1.16, 1.40)
Maximum posterior crossbite	2.91 (2.32, 3.66)	4.01 (3.04, 5.29)
Anterior crossbite*	2.28 (1.68, 3.09)	2.65 (1.90, 3.68)
Overjet*	1.35 (1.06, 1.72)	1.41 (1.10, 1.80)
Molar class discrepancy x maximum posterior crossbite		0.81 (0.74, 0.90)
Anterior crossbite x maximum posterior crossbite		0.65 (0.48, 0.88)

* as coded for the PAR index

Note that all the covariates were significant. In the model with interaction terms, the significance of the interaction terms suggests that the joint effect of two covariates is different than the individual contributions of these two components. For example, the risk of being classified in the treatment need group for subjects with high components of molar class discrepancy and left or right crossbite is down-weighted by the interaction factor of 0.81.

From the two regression models, predicted probability of treatment need was calculated for each subject. The Spearman rank correlation between the PAR index combined score and the predicted probability of SOTN were 0.90 and 0.89 for the initial and interaction models, respectively.

Plots (Figure 2) display the sensitivity and specificity for cutoff points spanning the range of the predicted probabilities for the logistic models and the combined PAR index. The highest values of sensitivity and specificity occur at the cutoff point where the lines cross; these are approximately 77% for the PAR index combination score and 81% for the interaction regression model. In Figure 3, the sensitivity and specificity of all three 3-classification methods are plotted. The similarity of the curves suggests that for classification purposes all three methods perform equivalently.

Discussion

This study examining the subjective treatment need decisions of orthodontists was conducted in an environment without possibility of financial gain for the examiners (three salaried university professors, one orthodontic graduate student) on a large cohort of children who were not identified as wanting treatment. The orthodontists were blinded to any knowledge of the child's perception of need/satisfaction. Treatment need decisions, thus, could be expected to be based on criteria other than financial gain or patient desires. The most general finding of this study is that the unencumbered assessment of need (unencumbered, that is, by financial gain or patient desires) is highly associated with malocclusion severity/treatment difficulty, as measured by the PAR index. These data and the findings of others[15,16] suggest that treatment need, malocclusion severity, and treatment difficulties are closely related concepts.

Since the orthodontists conducting this study were not randomly selected from a large pool of orthodontists, these results may not be generalizable to all practicing orthodontists. However, the results are the only available estimates of what a large-scale study would find. The examinations, although standardized, were conducted in a school setting; the impact that this setting had, when compared with the privacy of the office setting, cannot be determined. Interactions between the orthodontist and the child were minimized, and no contact with the parent(s) occurred. The intent was to remove the influence of child and parent desires on the orthodontist's assessment of need. The data were collected in an unbiased manner, as the relationship between clinical components and subjective assessment of treatment need was identified as an area of interest only after data collection was complete. The intent of the original study was to compare need, demand, and clinical characteristics with similar data collected previously in a younger age group.

Significant differences were detected between the SOTN and no SOTN groups. All subcomponents of the PAR index differed significantly between the two groups. The percent with SOTN did not vary by race or sex, contrary to our previous findings in a group of younger (8- to 10-year-old) children.[18]

A modeling approach was used to determine a parsimonious set of key variables that were predictive of treatment need. While the derivation of the original PAR index was based on rankings, we have used a binary classification of treatment need. The reliability in assessing a 5-point scale of treatment need was less than desirable; better agreement was observed with the dichotomous grouping of need/no need. However, underlying this dichotomy is a continuous probability of treatment need, which may be estimated from the regression models.

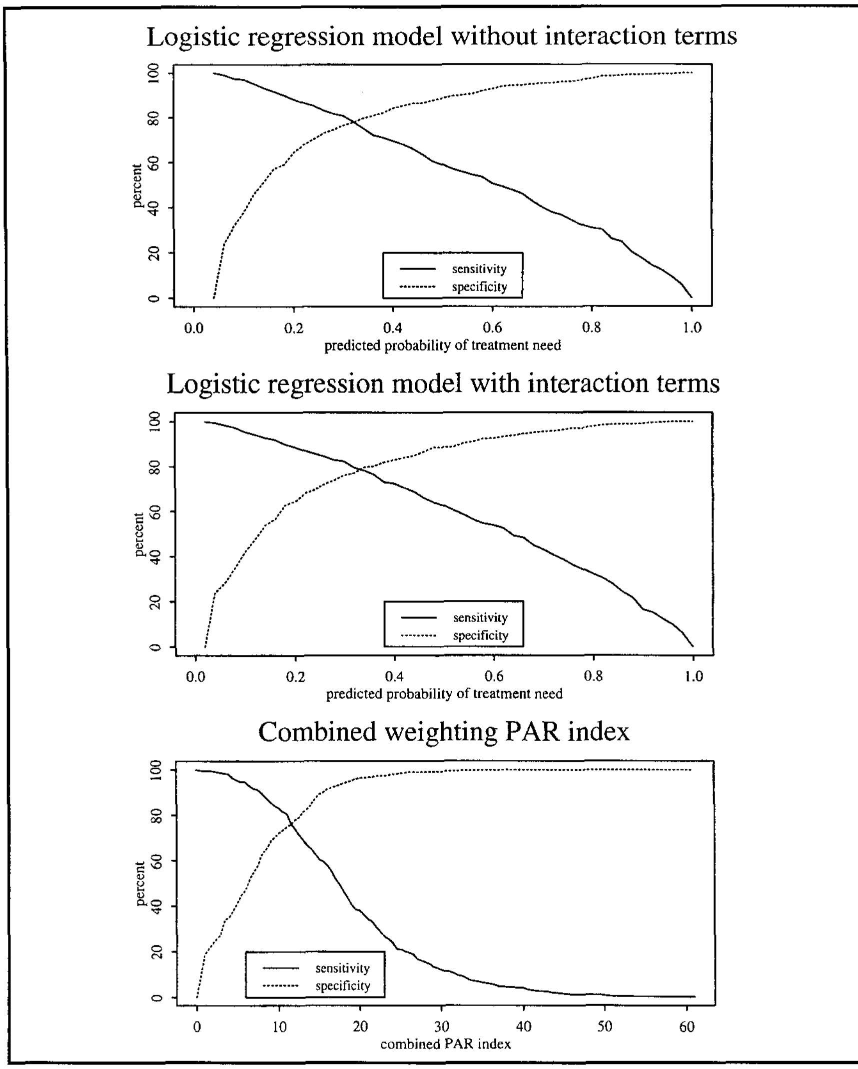

Figure 2
Sensitivity and specificity are displayed for logistic regression models and for the combined weighting PAR index.

It is instructive to examine outliers in the data set because they often point out the difficulties in applying a fixed set of rules when defining what constitutes orthodontic treatment need. Although need was highly associated with overbite, three of 21 children with greater than 100% overbite were judged to not need treatment; these may represent misclassifications or occlusions with other remarkable features. In the regression models, overbites in the 33% to 66% range, not considered clinically ideal, moderated the risk of needing treatment.

It is also important to note that we evaluated the classification capacities of the two logistic models using the data that generated the models. One would expect these models to perform well on the data upon which they are based. We thought that the inclusion of different or additional variables not used in the PAR index or considering interactions among components would improve the predictability of need; however, only minor improvement resulted. This suggests that the current structure of the American version of the PAR index is sufficient and gives additional credence to the association between treatment need and the PAR index. The reader should note that a different modeling procedure was used in this analysis, compared with the development of the PAR index, yet the results are similar.

It is likely that the associations between treatment need and PAR index scores are limited by inadequacies in the measurement scales and/or their application and by the fact that other factors are involved in the decision to treat. How does one improve on the 80% sensitivity and specificity of the current methods, without lowering one to advance the other? The surest answer is to invest in rigorously designed outcomes research. Only then will it be possible to reduce misclassifications. Whether other methods currently exist that can perform at higher levels of both sensitivity and specificity in identifying 14-year-old children in need of orthodontic treatment has not been shown.

It is not clear what additional criteria orthodontists should be using to decide treatment need, other than patient desires. Obvious considerations would be anomalies in growth and development (missing teeth, exfoliation, caries and periodontal concerns, cleft lip and palate syndromes, occlusal trauma) that are not measured directly by the PAR. However, the PAR index does measure indirectly occlusal features that are influenced by each of these; this appears sufficient.

Several variables and coding schemes used in the PAR index were used in the logistic regression models. Note that lower anterior alignment is included in our regression models but is not used in the U.S. version of the PAR index. Molar class discrepancy is used in the regression models rather than buccal occlusion component scores of the PAR. The molar class discrepancy variable contains information from both sides and has a wider range (0-10) than those of the buccal occlusion PAR subcomponents. The squared molar class discrepancy term in the models, with odds ratio estimate of less than 1.0, suggests that the effect of molar class discrepancy is not linear. Two subcomponents used in the PAR index, left and right posterior openbite, occurred rarely in our data and did not enter the logistic regression models. Midline was also not used in our models. While the weights associated with this component of the PAR index are high, the magnitude of this variable is small (0 - 2), so the overall effect of midline on the PAR index is not large. Variables representing sex, race, and age did not significantly improve the regression models. It is reassuring that, after including clinical features, these factors did not significantly influence the subjective assessment of treatment need. Variables representing screening orthodontists also did not significantly improve either logistic regression model.

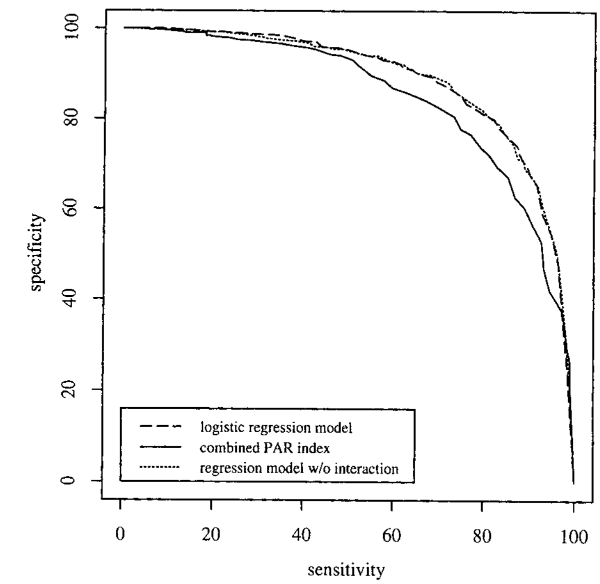

Figure 3
Comparison of sensitivity and specificity for logistic regression models and the combined weighting PAR index.

An esthetic component is not explicitly accounted for in our models. However, esthetic concerns likely played a role in the subjective assessment of orthodontic treatment need. This may be reflected in the odds ratio estimates associated with the objective covariates. This is similar to the approach used by Jenny and Cons.[13]

These data may stimulate some debate in the profession in an era of diminishing government resources; these findings are less likely to affect the practitioner whose patients are well financed and have a high desire for treatment for social concerns. However, these data could be considered by third-party payers in their decision to allocate fixed resources. If one accepts that PAR index scores fairly represent an objective determination of need, as this study suggests, then all one needs to do is balance resources and cutoff values for sensitivity and specificity. Payment plans could fund more treatment by lowering cutoff values, or less by raising cutoffs.

Prudent observers should insist that these findings be confirmed in other studies prior to widespread use in allocating limited resources. Current methods of allocating public resources for orthodontic care have not been sufficiently validated to represent treatment need; some appear to be based on meeting requirements of political constituencies. The specialty should demand that such methods be validated for the sake of patients who depend upon limited government resources.

Conclusions

Data collected to examine the relationship between orthodontists' subjective assessment of treatment need, made in the absence of financial gain considerations and patient desires, and objective measurements on 1155 eighth-grade school children obtained during standardized intra- and extraoral clinical examinations support the following conclusions:

1. Orthodontists' assessments of treatment need were univariately associated with interarch (sagittal, vertical, and transverse) and intraarch (alignment) descriptors of morphologic malocclusion; these descriptors were variously scored in different patterns.

2. The peer assessment rating (PAR) index, which measures treatment difficulty/malocclusion severity, was highly correlated with the orthodontists' subjective assessment of treatment need.

3. Logistic regression modeling, used to predict orthodontists' subjective treatment need, did not yield values of sensitivity and specificity that dramatically improved on those obtained by using only the PAR index combination score (77% for PAR to 81% for the interaction regression model).

4. In eighth-grade school children, orthodontists determined need based on occlusal features available from dental casts. Once occlusal factors were accounted for, sex, race, and age did not impact these decisions.

References

1. Richmond S, O'Brien K, Roberts C, Andrews M. Dentists' variation in the determination of orthodontic treatment need. Br J Orthod 1994;21:65-68.
2. Shaw WC. Factors influencing the desire for orthodontic treatment. Eur J Orthod 1981;3:151-62.
3. Wylie WL. The assessment of anteroposterior dysplasia. Angle Orthod 1947;17:97-109.
4. Shaw WC, O'Brien KD, Richmond S. Quality control in orthodontics: Factors influencing the receipt of orthodontic treatment. Br Dent J 1991;170:66-8.
5. Jago JD. The epidemiology of dental occlusion: A critical appraisal. J Pub Health Dent 1974;34:20-32.
6. Cohen LK, Jago JD. Toward the formulation of sociodental indicators. Int J Health Service 1976;6:681-98.
7. Casko J. Indices for determination of handicapping malocclusions. In: Wei SH, Casko J, eds. Orthodontic care for handicapped persons. Iowa City: University of Iowa, 1987:8-17.
8. Shaw WC, Richmond S, O'Brien KD. The use of occlusal indices: A European perspective. Am J Orthod Dentofac Orthop 1995;107:1-10.
9. Summers CJ. A system for identifying and scoring occlusal disorders: The occlusal index. Am J Orthod 1971;59:552-67.
10. Draker HL. Handicapping labio-lingual deviations: A proposed index for public health purposes. Am J Orthod 1960; 46:295-305.
11. Grainger RM. The orthodontic treatment priority index. Vital and Health Statistics Data Evaluation and Methods Research 1967;2(25).
12. Richmond S, O'Brien K, Buchanan I, Burden D. An introduction to occlusal indices. West Yorkshire, England: ORTHO-CARE (UK) Limited, 1992.
13. Jenny J, Cons NC. Comparing and contrasting two orthodontic indices, the Index of Orthodontic Treatment Need and the Dental Aesthetic Index. Am J Orthod Dentofac Orthop 1996;110:410-16.
14. Richmond S, Shaw WC, O'Brien KD, et al. The development of the PAR index (Peer Assessment Rating): Reliability and validity. Eur J Orthod 1992;14:125-39.
15. Richmond S, Shaw WC, Roberts CT, 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;14:180-97.
16. DeGuzman L, Bahiraei D, Vig KL, Vig PS, Weyant RJ, O'Brien K. The validation of the Peer Assessment Rating index for malocclusion severity and treatment difficulty. Am J Orthod Dentofac Orthop 1995;107:172-76.
17. O'Brien KD, Robbins R, Vig KWL, Vig PS, Shnorhokian H, Weyant R. The effectiveness of Class II Division 1 treatment. Am J Orthod Dentofac Orthop 1995;107:329-34.
18. Wheeler TT, McGorray SP, Yurkiewicz LA, Keeling SD, King GJ. Orthodontic treatment demand and need in 8- to 10-year-old schoolchildren. Am J Orthod Dentofac Orthop 1994;106:22-33.
19. Cohen J. Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psych Bull 1968;70:213-20.
20. Keeling SD, McGorray SP, Wheeler TT, King GJ. Imprecision in orthodontic diagnosis: Reliability of clinical measures of malocclusion. Angle Orthod 1996; 66(5):381-92.
21. Hosmer D, Lemeshow S. Applied logistic regression. New York: John Wiley, 1989.
22. Hosmer D, Lemeshow S. Goodness-of-fit for the multiple logistic regression model. Commun Stat 1980;A9:1043-69.