

Treatment Planning from Study Models: An Examiner Variability Study

W. A. B. BROWN, PH.D., D.ORTH., R.C.S.

E. M. HARKNESS, B.D.S., D.ORTH., R.C.S.

A. J. P. COUSINS, B.D.S., D.DO.

K. ISOTUPA, D.ODONT.

The practice of evaluating a malocclusion and of advising an orthodontic treatment plan on the basis of study models and radiographs (which might not necessarily involve a cephalometric film) has declined with the growth of specialisation, but is nevertheless of major importance in the organisation of third party payment for orthodontic treatment.

In the British National Health Service the third party payment agency, the Dental Estimates Board, has operated for nearly thirty years on this basis, but has accepted the need for cephalometric evidence over the latter two years only, and even so within a strictly limited set of circumstances.

The validity and reliability of third party assessment is a matter of current interest to the orthodontic profession; this paper describes an attempt to evaluate one component part only, viz., study models. Interexaminer error in assessment, from study models only, of Angle's classification has been shown to be high by Gravely and Johnson,¹ and it seems probable that the error for less well-defined criteria would be correspondingly higher.

A random sample of nonorthodontic patients (12 year-old school girls from a larger study by Harkness and Brown)² has been studied under a number of headings by four clinicians, but in a nonclinical environment and without any other clinical aids being available to them.

MATERIAL AND METHOD

Fifty sets of study models were ran-

domly selected from the 273 cases which formed the random selection in the Harkness and Brown study.² Each of the sets of study models was identically prepared and trimmed in a similar manner.

The four examiners were orthodontists with comparable experience who combined clinical work with teaching and research; one of the examiners had an orthodontic practice.

The study models were analysed independently by each of the four examiners in three distinct stages; each stage was carried out at weekly intervals.

Stage I was simply to answer the question: Was treatment necessary? This question, as all the other questions, was considered on a five point scale. A rating of 3 implied that the examiner was uncertain or did not know which choice to make. The difference between a score of 1 or 2 (Yes) or 4 and 5 (No) was marginal, and for ease of analysis either 1 or 2 has been treated as of equal status, as has 4 or 5.

Stage II was designed to evaluate why treatment was or was not thought to be necessary. Four factors were considered: (1) aesthetics, (2) function, (3) crowding or spacing, and (4) the need to treat the overjet. As in Stage I a five scale answering system was prepared.

At Stage III an assessment was made of five aspects of treatment procedure:

- (1) Should it be done immediately or delayed?

- (2) Was it thought to be complicated or simple?
- (3) Would the treatment be of long or short duration?
- (4) Would appliances be necessary or not?
- (5) Would the treatment be with or without extraction of teeth?

As for Stages I and II, there was a five scale rating system.

For each of the stages the examiners not only did not know what the other examiners had recorded, but also they did not have access to what they themselves had recorded for the earlier stages.

RESULTS

Stage I

For the 50 cases all examiners agreed that treatment was necessary for 17 and unnecessary for 8 cases. There were 16 sets of study models about which they were uncertain; there was disagreement about whether or not there should be treatment for 9 of the cases.

Stage II

There was total agreement between examiners about the poor aesthetics of 12 cases and the desirable aesthetics of 5 cases, while for 13 all the examiners were uncertain. There was disagreement in about 20 of the cases, that is, no less than 40% of the sample. Only 5 cases were identified with poor function, 17 with good, and again there was disagreement among the examiners about 20 of the cases.

In only 5 cases did the orthodontists disagree about the presence or absence of spacing, but there was a large number (66%) of occasions when all were uncertain.

Comparatively few cases (12%) were thought to be in need of overjet treatment, but there was a comparatively high number (34%) of examiner disagreements.

Stage III

Among the 41 cases thought to be in need of treatment, there were 30 about which the orthodontists disagreed on the time to start treatment. There was a similarly high level of disagreement, 50% and over, about the difficulty of treating the malocclusion, how long the treatment would take, the need for appliances, and even the need for extractions.

For all aspects of the approach to treatment there was a comparatively low level of uncertainty, the greatest uncertainty being related to the duration of treatment: There were only four cases about which the examiners were certain that treatment would be long or short. None of the observers expressed any uncertainty about the need for extractions. Each appeared to be clear in his mind whether or not he would extract teeth.

DISCUSSION

It is believed that the results presented here would have been different had the study been carried out on orthodontic patients, but it would have been an impossible task to analyse a random sample of patients in the way these study models have been examined. An important part of this experiment was the use of study models from a random sample of 12-year-old school-girls living in an urban community. These models were not the usual collection of models of patients that had sought advice from an orthodontic department. It was the realisation that none of these study models were of children who were knowingly in need of treatment that assured that the examiners were not immediately biased toward deciding on what the treatment should be. Nevertheless, among the 50 cases analysed, 41 or 82%, were thought by at least one of the examiners to be in need of treatment. Two

of the examiners thought that 40 (80%) were in need of treatment, one thought 45 (90%) and the fourth, 47 (94%) of the cases were in need of treatment. This is a higher percentage than that reported by Haynes³ and Foster and Walpole Day,⁴ who examined 11-12 year-old girls.

The sample, by being restricted to 12 year-old schoolgirls, reduced to a certain extent the conflicting interpretations that might arise from a large dental age range or from differences of sex. Even so, there were nine significant differences between the girls who had reached menarche and those who had not. These differences, as one might expect, related to height, weight and angular measurements of the skull,⁵ which accord with the range of age for different stages of dental and skeletal development reported by Björk.⁶

The sequence of examination

The sequence of examination of the study models was designed to follow the pattern of routine examination. Aesthetic considerations are regarded as being the most important reason for patients to seek treatment, while adequate function of occlusion is thought to be one of the most obvious reasons for treatment. Neither can be effectively appraised from study models and probably, even when evaluated from the patient, are at best subjective appraisals and liable to a large degree of variation. Evaluation of crowding and spacing proved more difficult than had been anticipated; the high level of uncertainty can be attributed to the examining orthodontists being unable to commit themselves on spaces which existed as part of the normal development of occlusion, and were not thought to be variations from normal.

The treatment of the overjet must be one of the most consistent needs of any prolonged treatment and was the

TABLE I
THE TIMING OF TREATMENT

	Number	%
Treat now	7	14
Treat later	2	4
Uncertain if now or later	2	4
Examiners disagree	30	60
No treatment required	9	18

one aspect of this analysis about which there was most unanimity.

The high percentage of disagreement about the time for commencement of treatment, Table I, most likely reflects on the different methods of treatment that could be used; it may, however, represent differing appraisals of the nature of the problem. The same could be true for the analysis of the difficulties of treating the malocclusions, the need for an appliance, and the need for extractions.

Uncertainties and disagreements

The large number of uncertainties among the different aspects of the occlusions that were examined may be particularly high in this type of experiment where the information is restricted to study models. It should be apparent that attempting to derive information from study models does have its limitations. Rank ordering the level of uncertainty, Table II, shows unexpectedly that the orthodontists had the greatest difficulty in deciding about the level of crowding. These varying levels of uncertainty may be a pointer to some of the marginal aspects of orthodontic treatment where not enough information is available such as on the complexities and expected duration of treatment.

Disagreement among the orthodontists in their interpretation of a case and how they would carry out treatment is much easier to understand. In any evaluation of treatment need there is a large degree of subjective interpretation and the vast array of methods for treating a case can readily explain

TABLE II
LEVELS OF UNCERTAINTY:
Number and percentages were based on
n = 50 for each aspect of the study
models that was examined.

Note: The table follows the sequence of
examination.

Rank Order	Number	%
2 Need for treatment	16	32
3 Evaluation of aesthetics	13	26
6 Evaluation of function	6	12
1 Evaluation of crowding	33	66
7 Need to treat overjet	4	8
8 Timing of treatment	2	4
4 Difficulty of treatment	9	18
4 Duration of treatment	9	18
9 Need for appliance	1	2

disagreements about aspects of treatment. The fact that the four orthodontists had different training backgrounds may exaggerate the disagreements, but may also serve to warn against too much dogmatism about the rights and wrongs of any particular treatment philosophy. Table III rank orders the level of disagreements for each aspect studied and may well serve as a guide to areas where there is less understanding and knowledge than there should be. An important part of an experiment like this is that the examiners should be reasonably matched in experience, but quite apart from trying to match examiners in terms of experience and diagnostic and treatment skills, it is useful to remember that the examiners themselves need to be consistent in their assessments. The degree of their consistency can be assessed from their deci-

TABLE III
LEVELS OF DISAGREEMENT

Rank Order	Number	%
7 Need for treatment	9	18
5 Evaluation of aesthetics	20	40
5 Evaluation of function	20	40
8 Evaluation of crowding	5	10
6 Need to treat overjet	17	34
1 Timing of treatment	30	60
4 Difficulty of treatment	26	52
2 Duration of treatment	28	56
3 Need for appliance	27	54

TABLE IV
INTRAEXAMINER CONSISTENCY
BETWEEN STAGE I AND III

C = consistent I = inconsistent
U = Uncertain to certain

Examiner	C	I	U
1	38	5	7
2	40	4	6
3	42	6	2
4	37	2	11

sions about the need for treatment in Stage I and their views on the nature of treatment in Stage III.

The examiners were consistent in their decisions for Stage I and III in 37 to 42 cases and only inconsistent in from 2 to 6 cases (Table IV). Note was made of the occasions they were uncertain in their initial decision about the need for treatment but definite in their appraisal at Stage III. One examiner by the third stage made a definite decision about 2 cases whilst another made it about 11 cases. These figures suggest that the disagreements which are rank ordered in Table III are more likely to have arisen from the varying opinions about different approaches to treatment than from the personal inconsistencies of the examiners.

The data were appraised for the possibility that the disagreements might be arising from a consistently different view in one of the examiners, but such was not found to be the case.

This study confirms a similar study, Brown and Rønning,⁷ in which 27 examiners with varying experience in orthodontics analysed 30 sets of study models from a class of 11 year-old Finnish schoolboys.

SUMMARY

Four orthodontists of comparable but differing experience examined fifty sets of study models obtained from a random sample of 12 year-old schoolgirls. On the basis of these study models only, assessments were made of need

for treatment, for timing, length and complexity of treatment. The degrees of agreement have been analysed and discussed.

*King's College
Univ. of London
London WC2R 2LS
England*

BIBLIOGRAPHY

1. Gravelly, J. F. and Johnson, D. B.: Angle's classification of malocclusion: An assessment of reliability. *Brit. J. Orthodont.* 1:79-86, 1974.
2. Harkness, E. M. and Brown, W. A. B.: Clinical cephalometric standards: A radiographic study of 12-year-old British girls. *The Orthodontist* 4:23-34, 1972.
3. Haynes, S.: Orthodontic treatment needs in English children aged 11-12 years. *Brit. J. Orthodont.* 1:9-12, 1973.
4. Foster, T. D. and Walpole Day, A. J.: A survey of malocclusion and the need for orthodontic treatment in a Shropshire school population. *Brit. J. Orthodont.* 1:73-78, 1974.
5. Harkness, E. M.: Cephalometric standards: A radiographic study of 12-year-old Cardiff schoolgirls. M.Sc. Thesis. *University of Wales*, 1969.
6. Björk, A.: Timing of interceptive orthodontic measures based on stages of maturation. *Trans. European Orthodont. Soc.* 61-74, 1972.
7. Brown, W. A. B. and Rønning, O.: The evaluation of malocclusion from study models. An examiner variability study. *Proc. Finn. Dent. Soc.* 71:45-52, 1975.