European Journal of Orthodontics 33 (2011) 190–197
doi:10.1093/ejo/cjq052
Advance Access Publication 22 July 2010

Prediction of the outcome of orthodontic treatment of Class III malocclusions—a systematic review

Piotr Fudalej*, Magdalena Dragan** and Barbara Wedrychowska-Szulc***
*Department of Orthodontics, Palacky University, Olomouc, Czech Republic and Department of Pediatric Surgery, Institute of Mother and Child, Warsaw, **Department of Periodontics, Warsaw Medical University and ***Department of Orthodontics, Pomeranian Medical University, Szczecin, Poland

Correspondence to: Piotr Fudalej, Department of Pediatric Surgery, Institute of Mother and Child, Kasprzaka Strasse 17a, 01-211 Warszawa, Poland. E-mail: pfudalej@gmail.com

SUMMARY The purpose of this study was to systematically review the orthodontic literature to assess the effectiveness of a prediction of outcome of orthodontic treatment in subjects with a Class III malocclusion. A structured search of electronic databases, as well as hand searching, retrieved 232 publications concerning the topic. Following application of inclusion and exclusion criteria, 14 studies remained. Among other data, sample ethnicity, treatment method, age at the start and completion of treatment, age at follow-up, outcome measures, and identified predictors were extracted from the relevant studies. A subjective assessment of study quality was performed.

The heterogeneity of the samples and treatment methods prevented carrying out a meta-analysis. Thirty-eight different predictors of treatment outcome were identified: 35 cephalometric and three derived from analysis of study casts. Prediction models comprising three to four predictors were reported in most studies. However, only two shared more than one predictor. Gonial angle was identified most frequently—in five publications. The studies were of low or medium quality.

Due to the large variety of predictors and differences among developed prediction models, the existence of a universal predictor of outcome of treatment of Class III malocclusions is questionable.

Introduction

Treatment of children with a Class III malocclusion represents a challenge in orthodontics because unsuccessful outcome of orthodontic/orthopaedic therapy is relatively frequent (Westwood *et al.*, 2003; Baik, 2007). Despite elimination of the reverse overjet and achievement of an acceptable dental arch relationship during early intervention, relapse is observed irrespective of the treatment modality (Franchi *et al.*, 1997; Tahmina *et al.*, 2000) and at different ages (Battagel, 1993; Franchi *et al.*, 1997). Deterioration of occlusion was found in Class III patients of different ethnicity (Ngan *et al.*, 1997; Westwood *et al.*, 2003) and the incidence of relapse has been reported to be almost 50 per cent (Franchi *et al.*, 1997).

The ability of early classification to either an orthodontic or surgery group would allow efficient triage according to patient treatment need. Subjects, who could be successfully treated with orthodontic/orthopaedic appliances, could receive treatment during childhood or adolescence, while the treatment plan of individuals who eventually would need orthognathic surgery could be modified accordingly. Battagel (1993) was one of the first investigators who recognized a need for a model of prediction of long-term outcome of orthodontic treatment of a Class III malocclusion. She employed discriminant function analysis to identify predictors of relapse in a group of children treated with cervical headgear applied to the mandibular dentition. Battagel (1993) developed a four-variable discriminant model, capable of predicting relapse with high accuracy. Franchi *et al.* (1997) published a study in which a three-variable predictive model was established. However, the predictors identified by Battagel (1993) and Franchi *et al.* (1997) differed substantially. Several other studies (Tahmina *et al.*, 2000; Zentner and Doll, 2001; Zentner *et al.*, 2001) dealing with identification of predictors of the results of Class III treatment have been published. Tahmina *et al.* (2000) examined a sample of Asians treated with a chincup, whereas subjects treated with various methods were evaluated in two, possibly related, studies by Zentner and Doll (2001) and Zentner *et al.* (2001). The predictive variables identified by those authors mostly differed. Although various modalities, treatment timing, or ethnicity might have affected the findings of the above-mentioned studies, the variety of prediction models established raises doubts as to whether identification of reliable predictors is possible. Therefore, the objective of this study was to systematically review the orthodontic literature to assess the possibility of the reliable prediction of orthodontic treatment outcome in subjects with a Class III malocclusion.

Material and methods

Search strategy

PubMed, Embase, Cochrane Central Register of Controlled Trials, and Lilacs were searched to the first week of October

2008 using the strategy presented in Table 1. Based on the data from titles and abstracts of the retrieved studies, the following were included: growing patients, subjects treated orthodontically/orthopaedically and articles in English, Polish, Russian, or Spanish. The exclusion criteria were a pseudo-Class III, adults, patients treated surgically, untreated subjects, case or case series reports, review and summary articles, and an observation time shorter than 3 years.

Consensus concerning inclusion/exclusion was undertaken by two authors (PF and MD). The reference lists of these articles were perused and references to related articles were followed-up. Additionally, two orthodontic journals that publish online ahead of print, the *European Journal of Orthodontics* and *Angle Orthodontist*, were hand searched to identify such articles.

Data extraction and quality assessment

The following data were extracted from each study: sample size, gender proportion, ethnicity, treatment method, age at the beginning and completion of treatment, age at follow-up, outcome measures, type of outcome at follow-up, proportion of successful/(uncertain)/relapsed cases, type of statistical analysis used to identify predictors of treatment outcome, and identified predictors (Table 2). The heterogeneity of the samples and treatment methods made it impossible to carry out a meta-analysis.

According to the Centre for Reviews and Dissemination (2009), flaws in the design or conduct of a study can result in bias, and in some cases this can have as much influence on the observed effects as that of treatment. Evaluation of methodological quality gives an indication of the strength of evidence provided by the study. However, no single approach to assessing methodological soundness is appropriate to all systematic reviews. The best approach should be determined by contextual, pragmatic, and methodological considerations (Centre for Reviews and Dissemination, 2009). According to those guidelines, the subjective assessment of quality of investigations included in this systematic review comprised evaluation of description of selection process (including information as to whether the sample consisted of consecutively treated patients), sample size estimation and adequacy of outcome measures, method error estimation, statistical analysis, and validation. The criteria of quality score assignment are presented in Table 3. The quality of the studies was considered as follows: high—total score >9 points, medium/high—total score >7 and ≤9 points, medium—total score >5 and ≤7 points, and low—total score ≤5 points.

Table 1 Search strategy and number of studies found.

Search term	Number of studies			
	PubMed	Embase	Cochrane Central	Lilacs
#1 (Malocclusion Angle Class III) OR ((Class III) AND (Angle OR malocclusion OR bite)) OR prognathism	4706	4031	65	108
#2 Predict* OR Prognosis OR Long-term OR (long term)	1 428 228	851 974	37 384	9882
#3 Surg* OR Osteotomy OR Setback	1 749 195	2 315 960	404	54 210
#1 AND #2 NOT #3	212	125	3	1

Results

The search strategy resulted in the retrieval of 232 publications. Application of inclusion/exclusion criteria allowed identification of 14 relevant studies (Table 2), of which 11 were found both in PubMed and Embase, one exclusively in PubMed, and two were identified through hand searching. No pertinent article was found in the Cochrane Central Register or in the Lilacs database. The main reasons for exclusion were lack of identified predictors of success/relapse, case reports, and review articles.

Despite the fact that nominally 763 subjects were evaluated in the 14 included studies, the actual number of individuals examined was 683, because the sample evaluated by Zentner *et al.* (2001) was most probably part of another sample included in this review (Zentner and Doll, 2001). Gender distribution in three articles was not stated (Zentner and Doll, 2001; Zentner *et al.*, 2001; Wells *et al.*, 2006). Of the 538 subjects examined in the remaining 11 publications, 231 (43%) were males and 307 (57%) were females. In one study, the number of male subjects exceeded the number of females (Battagel, 1993), whereas Yoshida *et al.* (2006) included only females in their sample.

The ethnicity of the examined subjects was described in nine articles (Table 2). Ethnic background can be surmised on the basis of country of origin of five investigations. In two studies (Ngan and Wei, 2004; Ghiz *et al.*, 2005), the samples comprised individuals of different ethnicity (Caucasian and Asian).

Various treatment regimens were employed in the examined samples. In five studies, patients treated with chincups were included: in four the chincup was the sole orthopaedic appliance (Tahmina *et al.*, 2000; Ferro *et al.*, 2003; Ko *et al.*, 2004; Moon *et al.*, 2005), whereas in the investigation by Yoshida *et al.* (2006) a combination of a chincup and facemask was employed. In five investigations, the treatment protocols included the use of facemasks, typically in combination with rapid maxillary expansion (Baccetti *et al.*, 2004; Ngan and Wei, 2004; Ghiz *et al.*, 2005; Wells *et al.*, 2006; Yoshida *et al.*, 2006). Orthopaedic appliances, such as cervical headgear attached to the mandible, or functional appliances were used less frequently (one study, Battagel, 1993). Three studies (Zentner and

Table 2 Characteristics of studies included in this review.

Study	Sample size, gender, and ethnicity	Treatment method, extractions (yes, no, nd)	Age at treatment start, completion, and follow-up	Length of treatment and follow-up	Variable(s) used to evaluate outcome	Type of outcome at follow-up	Success/ (uncertain)/ relapse rate (%)	Statistics	Predictors of success/ relapse
Baccetti *et al.* (2004)	42 (20 M, 22 F), Caucasians	RME + FM; no	8.6, 9.6, 15.0	1.0, 5.4	OJ < 0 mm, Class III molars -> relapse	Success, relapse	71.4/28.6	DFA	Mandibular ramus length (Co–Goi), Ba–T–SBL angle, and MP/SBL angle
Battagel (1993)	34 (21 M, 13 F), Caucasians?	Mn CHG; no	12.4, nd, 20.0	nd, 5.6	OJ, OB > 0 mm -> success; OJ, OB < 0 mm -> relapse; OJ, OB = 0 mm -> uncertain	Success, uncertain, relapse	35.3/26.5/38.2	DFA	U1/PP angle, Labrale superior to soft tissue nasion, Labrale inferior to Sella vertical, number of anterior teeth in crossbite*
Ferro *et al.* (2003)	52 (22 M, 30 F), Caucasians	Splints + elastics + CC (SEC III); nd	9.0, 13.0, 22.0	4.0, 9.0	OJ < or = 0 mm -> relapse	Success, relapse	88.5/11.5	*t*-test	Wits, ANB, SNB, and OB at the end of tx
Franchi *et al.* (1997)	45 (22 M, 23 F), Caucasians?	FA; nd	5.6, nd, 15.8	nd, 6.6	Angle Class, anterior crossbite (concomitant molar and canine Class III and anterior crossbite of at least 1 incisor -> relapse)	Success, relapse	51.1/48.9	DFA	Axis of condylar process to SBL angle (CondAx–SBL), mandibular plane to palatal plane (MP/PP), and mandibular intermolar width
Ghiz *et al.* (2005)	64 (29 M, 35 F), 53% Caucasians, 47% Asians	FM; nd	9.2, nd, nd	nd, nd	OJ > 1 mm and angle Class I -> success	Success, relapse	68.8/31.2	Regression analysis	Position of the condyle with reference to the cranial base (Co–GD), mandibular ramus length (Co–Goi), total mandibular length (Co–Pg), and gonial angle (Ar–Goi–Me)
Ko *et al.* (2004)	40 (18 M, 22 F), Asians?	CC + fixed; Mixed	8.2, 11.1, 17.6	1.9, 6.5	Good facial profile, OJ and OB > 0 mm, Class I molars and canines without severe facial and dental asymmetry -> success	Success, relapse	52.5/47.5	*t*-test	L1/OcclPl, upper gonial angle, AB/OcclPl, AB/MP, ANB, APDI, Wits, Articular angle (S–Ar–Go), gonial angle, facial convexity, AB/facial plane, and L1/A–Pog
Moon *et al.* (2005)	45 (16 M, 29 F), Asians	CC; nd	8.6, 11.7, 17.4	3.1, 5.7	OJ > 2mm, OB > 1.5 mm -> success; OJ, OB < 0 mm -> relapse; intermediate values of OJ and OB -> uncertain	Success, uncertain, relapse	37.8/28.9/33.3	DFA	AB/MP angle; A to N-perp distance
Ngan and Wei (2004)	40 (21 M, 19 F), 38% Caucasians, 62% Asians	RME +FM; no	8.9, 9.8, 13.0	0.9, minimum 3 years	OJ > 1 mm -> success	Success, relapse	50.0/50.0	*t*-test	Growth treatment response vector established based on initial, post-treatment, and follow-up cephalograms

Table 2 (continued).

Study	Sample size, gender, and ethnicity	Treatment method, extractions (yes, no, nd)	Age at treatment start, completion, and follow-up	Length of treatment and follow-up	Variable(s) used to evaluate outcome	Type of outcome at follow-up	Success/ (uncertain)/ relapse rate (%)	Statistics	Predictors of success/ relapse
Schuster *et al.* (2003)	88 (39 M, 49 F), Caucasians	Various; nd	9.8, nd, 17.4	nd, nd	Need for surgery at the end of observation period	Success, relapse	73.9/26.1	Regression analysis, DFA	Wits, palatal plane to SN angle (PP/SN), and individualized inclination of the lower incisors
Tahmina *et al.* (2000)	56 (23 M, 33 F), Asians?	CC + FIXED; nd	9.3, 12.1, 18.4	2.8, 6.3	Occlusal status at the end of treatment after pubertal growth	Success, relapse	53.6/46.4	DFA	Gonial angle; N–A–Pog angle; RP/SN angle
Wells *et al.* (2006)	41 (not declared), Caucasians	RME + FM; nd	nd, nd, nd	nd, minimum 5 years	OJ > 0 mm -> success	Success, relapse	75.6/24.4	DFA	Vertical position of PNS and Go, OB, Mn length
Yoshida *et al.* (2006)	32 (0 M, 32 F), Asians	CC + FM + fixed; nd	10.2, 11.6, 17.0	1.4, 5.4	OJ, Angle classification (favourable responses to the orthopaedic appliances during the first phase of treatment, and optimal molar and canine relationships with a normal OJ and OB maintained long term -> success)	Success, relapse	50.0/50.0	DFA; Regression analysis	ANS–Me and gonial angle
Zentner and Doll (2001)	104 (nd), Caucasians	Various; nd	11.7, nd, 16.8	nd, nd	change of PAR Index during treatment	Greatly improved, improved, worse/no different	46.8/40.7/12.5	Correlation analysis	Low value of apical base relationship index (particularly resulting from undersized maxilla)
Zentner *et al.* (2001)	80 (nd), Caucasians?	Various; nd	10.9, nd, 16.0	nd, nd	Change of PAR Index during treatment	Greatly improved, improved, worse/no different	50.0/37.5/12.5	Regression analysis	Two models: (1) apical base relationship and gonial angle; (2) apical base relationship, mid facial length/mandibular length ratio (Co–A to Co–Gn %), net sum of 'maxillary difference' (defined as difference between measured maxillary length and individual standard) and 'mandibular difference' (defined as difference between measured mandibular length and individual standard)

M, males; F, females; nd, not declared; PAR, Peer Assessment Rating; DFA, discriminant function analysis; RME, rapid maxillary expansion; FM, facemask; CHG, cervical headgear; CC, chincup; FA, functional appliance; fixed, fixed appliance; OJ, overjet; OB, overbite; APDI, anteroposterior dysplasia indicator; RP, mandibular ramus plane; and ?, likely.
*Prediction model unsuitable for children who required extractions.

Table 3 Assessment of quality of studies included in the review.

Study	A. Description of selection process	B. Previous estimate of sample size	C. Consecutive cases	D. Choice of outcome measure	E. Analysis of method error	F. Statistical analysis	G. Validation	Judged quality standard	Quality score
	Points assigned							Max points	
	0, 1, or 2	0 or 1	0 or 1	0, 1, or 2	0, 1, or 2	0 or 1	0, 1, or 2	0 to 11	
Baccetti *et al.* (2004)	2	0	0	1	1	1	0	5	Low
Battagel (1993)	1	0	0	2	2	1	1	7	Medium
Ferro *et al.* (2003)	1	0	0	1	2	1	0	5	Low
Franchi *et al.* (1997)	2	0	0	1	2	1	0	6	Medium
Ghiz *et al.* (2005)	1	0	0	1	2	1	0	5	Low
Ko *et al.* (2004)	1	0	0	2	0	1	0	4	Low
Moon *et al.* (2005)	1	0	0	2	0	1	0	4	Low
Ngan and Wei (2004)	1	0	0	1	1	1	0	4	Low
Schuster *et al.* (2003)	2	0	0	1	2	1	0	6	Medium
Tahmina *et al.* (2000)	1	0	0	1	0	1	0	3	Low
Wells *et al.* (2006)	1	0	0	1	1	1	0	4	Low
Yoshida *et al.* (2006)	1	0	0	1	2	1	0	5	Low
Zentner and Doll (2001)	1	0	0	1	0	1	0	3	Low
Zentner *et al.* (2001)	1	0	0	1	0	1	0	3	Low

Description of quality score assignment: A—0, if inadequate description; 1, if some details of sample selection missing; and 2, if in-depth description of sample selection. B—0, if not performed and 1, if performed. C—0, if sample comprised non-consecutive patients or no information regarding this was included and 1, if sample comprised consecutive patients. D—0, if inadequate outcome measure; 1, if partially adequate outcome measure; and 2, if adequate outcome measure. E—0, if method error not evaluated; 1, if partially adequate method error analysis; and 2, if adequate method error. F—0, if inadequate and 1, if adequate. G—0, if not performed; 1, if partially adequate validation; and 2, if adequate validation.

Doll, 2001; Zentner *et al.*, 2001; Schuster *et al.*, 2003) included patients treated by various methods.

The average age at commencement of treatment was 9.4 years and ranged from 5.6 (Franchi *et al.*, 1997) to 12.4 (Battagel, 1993) years. The average age at which the final (end of follow-up) examination was carried out was 17.2 years and ranged from 15.8 (Franchi *et al.*, 1997) to 22 (Ferro *et al.*, 2003) years. The average length of post-treatment follow-up was 6.3 years and ranged from 5.4 (Baccetti *et al.*, 2004; Yoshida *et al.*, 2006) to 9 years (Ferro *et al.*, 2003). It should be emphasized, however, that age at commencement and completion of treatment, age at final records, or length of follow-up were not always reported (Table 2).

Overjet was the most frequently used measure of treatment outcome. It was employed to establish success/(uncertain)/relapse groups by the authors of nine studies (Table 2). However, overjet was the only measure of treatment outcome in three studies (Ferro *et al.*, 2003; Ngan and Wei, 2004; Wells *et al.*, 2006); in other investigations, it was used in conjunction with overbite (Battagel, 1993; Ko *et al.*, 2004; Moon *et al.*, 2005) or with Angle classification (Baccetti *et al.*, 2004; Ko *et al.*, 2004; Ghiz *et al.*, 2005; Yoshida *et al.*, 2006). On the other hand, Zentner and Doll (2001) and Zentner *et al.* (2001) employed the Peer Assessment Rating (PAR) Index to identify successful or relapsed cases. Subjective measures, such as 'need for surgery at the end of observation' (Schuster *et al.*, 2003), 'good facial profile' (Ko *et al.*, 2004), or 'occlusal status' (Tahmina *et al.*, 2000) were also used.

In 10 investigations, treatment outcome was described dichotomously (success versus relapse). The reported success rate ranged from 51.1 (Franchi *et al.*, 1997) to 88.5 (Ferro *et al.*, 2003) per cent. Ngan and Wei (2004) and Yoshida *et al.* (2006) matched success and relapse groups to include the same number of successful and relapsed patients. In those investigations, the success rate could not be calculated. Battagel (1993) and Moon *et al.* (2005) also established an 'uncertain' group, in which treatment outcome was judged as doubtful. If uncertain groups are disregarded, the success rate in the studies by Battagel (1993) and Moon *et al.* (2005) roughly corresponds to 50 per cent. Zentner and Doll (2001) and Zentner *et al.* (2001) employed the PAR Index to establish 'greatly improved', 'improved', and 'worse/no different' groups. Assuming that the results of treatment of patients from the greatly improved and improved groups were successful, the success rate was equal to 87.5 per cent.

In total, 38 different predictors of treatment outcome were identified in 14 studies. Thirty-five were cephalometric variables (20 linear, 13 angular, and two ratios) and three were derived from analysis of study models. Most studies reported a set of three to four predictors (Battagel, 1993; Franchi *et al.*, 1997; Tahmina *et al.*, 2000; Zentner *et al.*, 2001; Ferro *et al.*, 2003; Schuster *et al.*, 2003; Baccetti *et al.*, 2004; Ghiz *et al.*, 2005; Wells *et al.*, 2006). However,

Ko *et al.* (2004) listed 12 variables, which were correlated with successful treatment outcome. On the other hand, Zentner and Doll (2001) and Ngan and Wei (2004) identified one predictor: apical base relationship (Zentner and Doll, 2001) and growth treatment response vector (Ngan and Wei, 2004).

Only two studies shared more than one predictor of treatment outcome. Ferro *et al.* (2003) and Ko *et al.* (2004) listed ANB angle and Wits appraisal as predictors. Gonial angle was the most frequently identified variable by different groups of researchers (Tahmina *et al.*, 2000; Zentner *et al.*, 2001; Ko *et al.*, 2004; Ghiz *et al.*, 2005; Yoshida *et al.*, 2006); however, it was included in the prediction models in only five of 14 investigations (36%). Among other predictors identified in more than one study was the Wits appraisal (three studies), total mandibular length (Co–Pog), mandibular ramus length (Co–Goi), ANB angle, overbite, AB to mandibular plane angle, and apical base relationship (all in two studies).

Discussion

The uncertainty regarding the long-term results of Class III treatment was a stimulus to identify potential predictors of therapeutic success or failure. If an unfavourable outcome of therapy could be anticipated prior to treatment, then the type and timing of orthodontic/orthopaedic treatment could be modified. The authors of all studies included in this review employed a similar strategy of predicting factors – they divided patients long-term out of treatment into two (success versus relapse) or three (success versus uncertain versus relapse, or greatly improved versus improved versus worse/no different) groups demonstrating different therapeutic results. Subsequently, usually pre-treatment, variables correlated with treatment success or relapse, were established. In most studies discriminant function or regression analyses were performed to identify the sets of variables showing the highest prediction capability (i.e. prediction models). Ngan and Wei (2004) took a different approach. They evaluated cephalometric radiographs taken after the first phase of treatment [rapid maxillary expansion and facemask (RME+FM) therapy] and at follow-up. On the basis of post-treatment and follow-up cephalometric records, those authors then developed a 'Growth treatment response vector (GTRV)' – a ratio of horizontal growth changes of the maxilla and mandible determined along the occlusal plane. The GTRV below a certain value was reported to be suggestive of an unsuccessful second phase of treatment with fixed appliances. In such cases comprehensive edgewise treatment should be postponed.

The review of the 14 identified articles demonstrated that there were no studies that shared an identical set of predictors of treatment outcome. On the contrary, there was a substantial diversity of predictors. Of the 14 investigations, only Ferro *et al.* (2003) and Ko *et al.* (2004) enumerated concurrently more than one common variable (ANB and Wits). If the reviewed articles are grouped according to orthodontic/orthopaedic treatment modality (RME+FM or chincup), only little similarity can be observed within each group regarding detected predictors. For example, RME+FM treatment was used in the children examined by Baccetti *et al.* (2004), Ngan and Wei (2004), Ghiz *et al.* (2005), and Wells *et al.* (2006). Despite the extensive cephalometric analyses carried out (except for the investigation by Ngan and Wei, 2004), only two variables having a predictive value were simultaneously established in more than one study: mandibular ramus length (Baccetti *et al.*, 2004; Ghiz *et al.*, 2005) and total mandibular length (Ghiz *et al.*, 2005; Wells *et al.*, 2006). An alternative method of orthopaedic treatment of Class III malocclusion, the chincup, was used in the children followed by Tahmina *et al.* (2000), Ferro *et al.* (2003), Ko *et al.* (2004), and Moon *et al.* (2005). Of the 18 different predicting factors identified in the above mentioned four investigations, 14 were established only in a single study (Table 2) and four (ANB, Wits, gonial angle, and AB to mandibular plane angle) were identified concurrently by the authors of two studies (Table 2). Moreover values of coefficients of correlation between a particular variable and outcome might be low. Zentner *et al.* (2001) reported that the correlation coefficient for a set of two variables, gonial angle and apical base relationship, selected by means of regression analysis was 0.137; Pearson correlation coefficient (r) for gonial angle only was 0.238 ($r^2 = 0.057$), which suggests a rather weak association between the set of selected predictors and treatment outcome.

The variety of the variables assumed to have a predictive value, the dissimilarity of the sets of predictors established by different groups of researchers, and the possible low correlation between a particular predictor and treatment outcome (only a few studies specified values of correlation coefficients) imply that prediction of Class III treatment outcome is questionable. On the other hand many authors reported high classification power for the developed prediction models: Franchi *et al.* (1997) demonstrated 95.6 per cent accuracy of discriminant function, Schuster *et al.* (2003) 93.2 per cent, Tahmina *et al.* (2000) 85.7 per cent, Yoshida *et al.* (2006) 84.4 per cent, and Baccetti *et al.* (2004) 83.3 per cent. A lower classification power (< 80%) was reported by Moon *et al.* (2005) where 77.8 per cent of cases were correctly classified, 73.3 per cent by Battagel (1993) and 73 per cent by Wells *et al.* (2006). In order to reconcile these conflicting data, an innate property (limitation) of prediction models based on discriminant function and regression analyses should be considered. Models for prediction of treatment outcome derived from these statistical procedures prognosticate *post hoc* what has occurred previously. It is not uncommon to obtain very good classification if one uses the same cases from which the classification functions are computed. In order to

determine how well a current classification model 'performs' one must classify *a priori* different cases, that is, cases that were not used to estimate the classification model. Only classification of new cases allows assessment of the predictive validity of the classification model; the classification of old cases only provides a diagnostic tool to identify outliers or areas where classification function seems to be less adequate (Klecka, 1980; StatSoft Inc., 2008).

A priori classification of new cases was carried out only by Battagel (1993). Although correct classification of 7 out of 8 new patients resulted in a discriminative power of 87.5 per cent, the small number of classified cases precludes any firm conclusions. Unfortunately, as other authors did not perform validation procedures, the actual accuracy of the prediction is unknown. It can only be speculated that the low actual predictive power of the developed prediction models might have contributed to the wide disparity between them.

The number of groups created in the reviewed papers was not identical. Two groups, success and relapse, were established in 10 studies, and three groups, success, uncertain and relapse or greatly improved, improved and worse/no different, in four studies. Although similar outcome measures were employed in most investigations, overjet, overbite or Angle classification, they were not used uniformly. For example, Ghiz *et al.* (2005) set overjet at 1 mm as a discriminating value between the success and relapse groups, whereas Wells *et al.* (2006) determined successfully treated patients as having an overjet > 0 mm. Moon *et al.* (2005), in turn, defined their success group as patients with an overjet in excess of 2 mm. Overbite was also used differently in various studies. Battagel (1993) classified patients with overbite > 0 mm to the success group, whereas Moon *et al.* (2005) used an overbite > 1.5 mm for classification to the success group. As a result, success or relapse groups from various investigations were likely not equivalent. Therefore this potential inequivalence might have contributed to disparity among the prediction models.

The success rates, computed on the basis of data from the studies selected for this systematic review, may not be accurate due to inadequate sample selection. Determination of the success rate of a Class III treatment protocol should involve evaluation of consecutive patients. If the sample consists of selected rather than consecutive subjects there is a risk of unduly optimistic rates of success. A child diagnosed with a Class III malocclusion may be treated orthodontically or treatment may be delayed until craniofacial growth is complete. At that time orthodontic/surgical treatment is usually initiated. In, so called, borderline cases making a binary choice, i.e. to start or postpone treatment, is difficult since there are premises both to begin and to delay treatment. Some clinicians may initiate treatment, whereas others may postpone it, in a patient with a malocclusion of the same severity. Therefore, depending on the approach of an orthodontist or orthodontic department, if a similar treatment philosophy is adhered to – more aggressive (meaning orthodontic treatment of children even with severe malocclusions) or more cautious (meaning a delay of orthodontic therapy in subjects with severe Class III) – the success rate may differ despite the use of the same orthodontic protocol or appliance. It is also likely to be higher in a sample in which few borderline subjects are included, and lower in a sample with many borderline cases. Thus, the actual rate of success or relapse can be established provided a sample comprises consecutively diagnosed and treated children. Unfortunately, samples from the reviewed articles comprised selected cases and no information on the criteria regarding the treatment decision: treat versus postpone, was offered.

Conclusions

The following conclusions can be made on the basis of the review of the 14 included studies:

1) The possibility of accurate prediction of outcome of orthodontic/orthopaedic treatment of Class III malocclusion seems questionable.
2) Validation testing of a prediction model on the cases, which were not used to develop it, is mandatory in order to evaluate an actual discriminative power of the prediction model.

References

Baccetti T, Franchi L, McNamara J A 2004 Cephalometric variables predicting the long-term success or failure of combined rapid maxillary expansion and facial mask therapy. American Journal of Orthodontics and Dentofacial Orthopedics 126: 16–22

Baik H S 2007 Limitations in orthopedic and camouflage treatment for Class III malocclusion. Seminars in Orthodontics 13: 158–174

Battagel J 1993 Discriminant analysis: a model for the prediction of relapse in Class III children treated orthodontically by a non-extraction technique. European Journal of Orthodontics 15: 199–209

Centre for Reviews and Dissemination. 2009 Systematic reviews. CRD's Guidance for Undertaking Reviews in Healthcare. University of York NHS Centre for Reviews & Dissemination, York

Ferro A, Nucci L P, Ferro F, Gallo C 2003 Long-term stability of skeletal Class III patients treated with splints, Class III elastics, and chincup. American Journal of Orthodontics and Dentofacial Orthopedics 123: 423–434

Franchi L, Baccetti T, Tollaro I 1997 Predictive variables for the outcome of early functional treatment of Class III malocclusion. American Journal of Orthodontics and Dentofacial Orthopedics 112: 80–86

Ghiz M A, Ngan P, Gunel E 2005 Cephalometric variables to predict future success of early orthopedic Class III treatment. American Journal of Orthodontics and Dentofacial Orthopedics 127: 301–306

Klecka W R 1980 Discriminant analysis. Sage Publications, Newbury Park

Ko Y-I, Baek S-H, Mah J, Zang W-S 2004 Determinant of successful chincup therapy in skeletal Class III malocclusion. American Journal of Orthodontics and Dentofacial Orthopedics 126: 33–41

Moon Y M, Ahn S J, Chang Y I 2005 Cephalometric predictors of long-term stability in the early treatment of Class III malocclusion. Angle Orthodontist 75: 747–753

Ngan P W, Wei S H Y 2004 Early treatment of Class III patients to improve facial esthetics and predict future growth. Hong Kong Dental Journal 1: 28–34

Ngan P W, Hägg U, Yiu C, Wei H 1997 Treatment response and long-term dentofacial adaptations to maxillary expansion and protraction. Seminars in Orthodontics 3: 255–264

Schuster G, Lux C J, Stellzig-Eisenhauer A 2003 Children with Class III malocclusion: development of multivariate statistical models to predict future need for orthognathic surgery. Angle Orthodontist 73: 136–145

StatSoft Inc. 2008 Discriminant function analysis (http://www.statsoft.com/textbook/stdiscan.html) (1 March 2009, date last accessed)

Tahmina K, Tanaka E, Tanne K 2000 Craniofacial morphology in orthodontically treated patients of Class III malocclusion with stable and unstable treatment outcomes. American Journal of Orthodontics and Dentofacial Orthopedics 117: 681–690

Wells A P, Sarver D M, Proffit W R 2006 Long-term efficacy of reverse pull headgear therapy. Angle Orthodontist 76: 915–922

Westwood P V, McNamara Jr J A, Baccetti T, Franchi L, Sarver D M 2003 Long-term effects of Class III treatment with rapid maxillary expansion and facemask therapy followed by fixed appliances. American Journal of Orthodontics and Dentofacial Orthopedics 123: 306–320

Yoshida I, Yamaguchi N, Mizoguchi I 2006 Prediction of post-treatment outcome after combined treatment with maxillary protraction and chincap appliances. European Journal of Orthodontics 28: 89–96

Zentner A, Doll G M 2001 Size discrepancy of apical bases and treatment success in Angle Class III malocclusion. Journal of Orofacial Orthopedics 62: 97–106

Zentner A, Doll G M, Peylo M 2001 Morphological parameters as predictors of successful correction of Class III malocclusion. European Journal of Orthodontics 23: 383–392