Systematic Review

Soft-tissue changes in Class II malocclusion patients treated with extractions: a systematic review

Guilherme Janson, Lucas Marzullo Mendes, Cintia Helena Zingaretti Junqueira and Daniela Gamba Garib

Department of Orthodontics, Bauru Dental School, University of São Paulo, Bauru, Brazil

Correspondence to: Guilherme Janson, Department of Orthodontics, Bauru Dental School, University of São Paulo, Alameda Octávio Pinheiro Brisolla 9-75, Bauru-SP 17012-901, Brazil. E-mail: jansong@travelnet.com.br

Summary

Background: Concerns about the effects caused by premolar extractions on the soft-tissue profile have motivated many investigations in different malocclusions.

Objectives: To evaluate the cephalometric facial soft-tissue changes after orthodontic treatment with premolar extractions of Class II division 1 malocclusion subjects.

Search methods: Electronic databases PubMed, Web of Science, Embase, and Scopus were searched.

Selection criteria: Abstracts that appeared to fulfil the initial criteria (premolar extraction; cephalometric soft-tissue analyses/changes) were selected. The full-text original articles were then retrieved. Their references were also hand-searched.

Data collection and analysis: By consensus of two researchers, the articles that fulfilled the selection criteria and quantified facial soft-tissue changes were individually analysed. Some methodological flaws were identified and some articles were excluded. The studies were rated according to the type of study, sample description and homogeneity, malocclusion severity, consideration of confounding factors, validity of the method, and statistical analyses.

Results: Heterogeneous information about malocclusion severity before treatment was found in most articles. Statistically significant soft-tissue changes reported included nasolabial angle (NLA) increasing from 2.4 to 5.40 degrees in 2-premolar extraction protocol and from 1 to 6.84 degrees in 4-premolar extraction protocol. Retrusion of the upper and lower lips were also verified, with less retraction of the lower lip in 2-premolar extraction groups.

Conclusions: When Class II division 1 malocclusion is treated with premolar extractions, the NLA increases and the lips are retracted. However, there is less retraction of the lower lip in the 2-maxillary premolar extraction protocol.

Introduction

Orthodontic treatment can influence patient’s profile and aesthetics, especially when extractions and extensive anterior retraction are involved (1, 2). The effects of extraction and non-extraction therapies have been widely investigated (2–12), but it seems that the debate about the extraction effects is still far from finishing.

Soft-tissue thickness (13–15), pre-treatment labial tension (2, 15), type of malocclusion, crowding (16, 17), and face height (6) are some of the factors that seem to influence the effects of tooth extraction on the soft-tissue profile.

There are many therapeutic approaches to treat Class II malocclusions, such as removable or fixed functional orthopaedic appliances (18–21), extra- or intra-oral distalizing appliances (10, 22, 23),
tooth extractions (9, 11, 24–27), and orthodontic-surgical treatment (28, 29), when there is accentuated skeletal discrepancies.

Evaluation of facial profile and balance is a continuous learning process for orthodontists (7). However, most studies concerned with the effects of orthodontic treatment on facial profile have been based mostly on assumptions than on actual changes in the relationship between the incisors and lips (30–32). Some only suggest favourable changes in the long term (33).

There are concerns that premolar extractions might cause greater lip retrusion and impair the resulting profile more than treatment without extractions (12–14, 27, 34, 35). The speculation is that anterior retraction would result in an undesirable flattened facial appearance. Although many recent studies have refuted this hypothesis (3, 5, 12, 36–40), this issue keeps been studied. However, there are no consistent data regarding the amount of soft-tissue changes in Class II malocclusion treatment with 2- or 4-premolar extractions (13, 34, 41). Therefore, this systematic review aimed to evaluate the soft-tissue changes after orthodontic treatment with premolar extractions in Class II division 1 malocclusion subjects.

Materials and methods

Search methods

The first phase of this systematic review involved development of a specific protocol and research question based on the Population Intervention Control Outcome Study Design (PICOS) format (Table 1) (42). Based on these specific criteria, the studies should be able to answer the question: ‘Which changes can be expected on the soft-tissue profile of Class II division 1 malocclusion subjects treated with comprehensive orthodontics and premolar extractions?’

A computerized search was conducted using the main terms ‘soft-tissue’, ‘profile’, ‘facial’, ‘Class II’, ‘premolar’, and ‘extraction’ in the following electronic databases: PubMed, Web of Science, Embase, Scopus, and Cochrane (Supplementary Table 1).

Table 1. PICOS format and research question. PICOS, Population Intervention Control Outcome Study Design.

<table>
<thead>
<tr>
<th>Population</th>
<th>Intervention</th>
<th>Comparison</th>
<th>Outcome</th>
<th>Study design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle’s Class II division 1 subjects</td>
<td>Comprehensive orthodontics with premolar extractions</td>
<td>Any other orthodontic treatment and/or a control group</td>
<td>Soft-tissue/profile analyses/changes</td>
<td>Randomized clinical trial, retrospective or prospective cohort studies</td>
</tr>
</tbody>
</table>

Selection criteria

To identify potential articles, the initial search was performed by title and abstract (43). The initial inclusion criteria were: studies published until June 2015; quoting cephalometric soft-tissue changes in samples treated with premolar extractions. The types of study selected were randomized clinical trial, retrospective or prospective cohort studies. Two researchers independently conducted this selection process and, thereafter, evaluated the articles from the selected abstracts. To avoid any biased results, interexaminer conflicts were resolved by discussion on each article to reach a consensus regarding which articles fulfilled the final selection criteria.

The exclusion criteria were: studies that did not have a well-defined Class II group treated with premolar extractions, did not describe the occlusal malocclusion severity, or had inadequate statistical analyses were not included in this review. Case reports, letters to the editors, or experts’ opinions were also rejected. Studies with no systematic and random error analyses or a control group were also excluded. The references of the selected articles were then hand-searched for additional studies.

Data collection and risk of bias analysis

Risk of bias was assessed not only through inclusion but specially through the exclusion criteria. Biased studies that did not describe the extraction protocols used or the Class II severity were systematically excluded. Recognizing that studies with more elaborate methodology and less bias problems may provide more reliable conclusions, the articles ultimately selected were classified through the Cochrane collaboration’s tool for assessing risk of bias (44).

The Cochrane collaboration’s tool for assessing risk of bias (44) helped to evaluate the confidence of the included studies. Several domains were separately verified: sample size, Class II malocclusion severity description, homogeneity regarding extracted teeth, control of confounding factors, validity of method, and statistical analysis. The criteria used to classify each study are described on Table 2.

Each study received for each domain a judgment of low, high, or unclear risk of bias, which means information provided was not enough for a fair evaluation (44). Afterwards, these qualification features were used to classify the articles based on their scientific weight. The studies were finally classified into the following categories:

Table 2. Methodological scoring process.

<table>
<thead>
<tr>
<th>Data</th>
<th>High risk of bias</th>
<th>Unclear risk of bias</th>
<th>Low risk of bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject number</td>
<td>N &lt; 20</td>
<td>20 ≤ N ≤ 29</td>
<td>N ≥ 30</td>
</tr>
<tr>
<td>Description of Class II malocclusion severity</td>
<td>ANB or incisor classification</td>
<td>Overjet</td>
<td>Antero-posterior molar relationship</td>
</tr>
<tr>
<td>Homogeneity regarding extracted teeth</td>
<td>Group had two kinds of treatment protocol</td>
<td>Did not distinguish first premolar from second premolar extractions</td>
<td>Clearly established if two maxillary or four premolars were extracted, and if they were first or second.</td>
</tr>
<tr>
<td>Control of confounding factors</td>
<td>No precautions were taken during record taking</td>
<td>Some precautions were taken during record taking</td>
<td>Precautions were taken to ensure a relaxed position of soft tissues during record taking</td>
</tr>
<tr>
<td>Validity of method</td>
<td>Exam with no standardized technique based on the literature</td>
<td></td>
<td>Exam used for diagnosis had a standardized technique based on the literature</td>
</tr>
<tr>
<td>Statistical analysis</td>
<td>Statistician judged the statistical analyses.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. Low risk of bias (studies with all domains at low risk of bias or one unclear key domain): reasonable bias that would hardly alter the results.
2. Unclear risk of bias (studies with more than one domain at unclear risk of bias): reasonable bias that makes the results suspicious.
3. High risk of bias (studies with one or more domains at high risk of bias): reasonable bias that strongly commits the reliability of the results.

Results

After a database search, 154 articles were retrieved from PubMed, 132 from Web of Science, 91 from Embase, 184 from Scopus, 7 from Cochrane, and 16 from a hand-search (Figure 1, Supplementary Table 1). After duplicates were removed, the entire search strategy resulted in 262 abstracts, 61.45 per cent of which were published between 2001 and June 2015. These results demonstrate that studies about the influence of first premolar extraction on the soft-tissue profile have considerably increased in the past years.

After preliminary exclusions, based on titles and abstracts, 102 full-text articles were assessed for eligibility (Figure 1). The main reasons for exclusions were: absence of complete description of malocclusion type, severity of Class II malocclusion, no identification of the removed premolars, and no comparison groups. Finally, 13 articles met the initial inclusion criteria (Figure 1; Supplementary Table 1).

Preliminary description of the 13 articles is shown in Supplementary Table 2. Eight (27, 34, 45–50) were retrospective studies, three (29, 51, 52) prospective, and two unclear (53, 54). All the studies had 14–44 patients in the extraction groups. All of them were growing patients, except in three, where the extraction group was compared to a surgical group (29) and the other two used mini-implants (51, 52).

Malocclusion severity description was diversified (Supplementary Table 2). The references used to measure it were: overjet in three articles (45, 49, 52), molar sagittal relationship in six articles (e.g. half cusp, full cusp) (27, 34, 46–48, 51), ANB in one article (29), ANB and overjet in two articles (53, 54), and according to incisor relationship...
Table 4. Main changes and severity. MLA, mentolabial angle; NLA, nasolabial angle; TPA, transpalatal arch.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Class II severity (inadequate: ANB/incisor; partial: overjet; adequate: at least half cusp)</th>
<th>Extraction subjects</th>
<th>NLA</th>
<th>UL-E</th>
<th>LL-E</th>
<th>UL-S</th>
<th>LL-S</th>
<th>MLA</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL-Sibaie and Hajeer, 2014 (51)</td>
<td>Adequate</td>
<td>28*</td>
<td>9.08</td>
<td>−2.98</td>
<td>−2.50</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Battagel, 1996 (45)</td>
<td>Partial</td>
<td>30 children</td>
<td>5.93</td>
<td>−2.47</td>
<td>−1.42</td>
<td>—</td>
<td>—</td>
<td>−4.21</td>
</tr>
<tr>
<td>de Almeida-Pedrin et al., 2009 (46)</td>
<td>Adequate</td>
<td>30 (boys and girls)</td>
<td>+2.40</td>
<td>−2.60</td>
<td>−1.00</td>
<td>—</td>
<td>—</td>
<td>−0.40</td>
</tr>
<tr>
<td>Janson et al., 2007 (27)</td>
<td>Adequate</td>
<td>22 patients</td>
<td>+5.40</td>
<td>−2.66</td>
<td>−1.15</td>
<td>−2.28</td>
<td>−0.83</td>
<td>−4.9 (m);</td>
</tr>
<tr>
<td>Kinzinger et al., 2009 (29)</td>
<td>Inadequate</td>
<td>20 patients</td>
<td>+2.87</td>
<td>−0.75</td>
<td>+0.75</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Upadhyay et al., 2012 (52)</td>
<td>Inadequate</td>
<td>14 patients</td>
<td>+11.55</td>
<td>−2.41</td>
<td>−2.73</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Weyrich and Lisson, 2009 (34)</td>
<td>Adequate</td>
<td>34</td>
<td>—</td>
<td>−1.97</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Zierhut et al., 2000 (47)</td>
<td>Adequate</td>
<td>23 American Caucasian patients</td>
<td>—</td>
<td>−5.03</td>
<td>−4.19</td>
<td>−4.21</td>
<td>−3.76</td>
<td>—</td>
</tr>
<tr>
<td>Bishara et al., 1994 (48)</td>
<td>Adequate</td>
<td>44 subjects</td>
<td>—</td>
<td>−4.9 (m);</td>
<td>−4.6 (m);</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Finnoy et al., 1987 (49)</td>
<td>Partial</td>
<td>30 patients</td>
<td>+6.50</td>
<td>−3.30</td>
<td>−2.50</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Looi and Mills, 1986 (50)</td>
<td>Inadequate</td>
<td>30 patients</td>
<td>+5.9</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>+5.30</td>
</tr>
<tr>
<td>Verma, 2013 (53, 54)</td>
<td>Partial</td>
<td>50 Indian female patients</td>
<td>+6.84</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>+4.92</td>
</tr>
<tr>
<td>Weyrich and Lisson, 2009 (34)</td>
<td>Adequate</td>
<td>37</td>
<td>+1.00</td>
<td>−3.00</td>
<td>−2.29</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*Mini-implants group: skeletal anchorage.
**TPAs group: conventional anchorage.

(British Standard 4492: 1983) in one article (50). Twelve out of the 13 selected studies described an enlarged overjet (Supplementary Table 2).

The main characteristics of each study are presented in Supplementary Table 3, such as: extracted teeth, treatment appliances, comparison groups, and evaluated changes. Seven of them had 2-premolar extraction groups (27, 29, 45–47, 51, 52), five had 4-premolar extraction groups (48–50, 53, 54), and one had both types of groups (34). All of them were treated with fixed Edgewise appliances.

Among the 13 studies, one compared extractions to pendulum and cervical headgear groups (46); one compared to transpalatal arches (TPAs) conventional anchorage (51); two compared to cervical headgear (27, 47); four compared to mandibular advancement (Frankel, Andresen activator, functional mandibular advance, Forsus Fatigue-Resistant Device-3M) (29, 45, 50, 52); one compared to other extraction protocol (34); three compared with non-extraction approach (49, 53, 54); and one compared with untreated normal group (48). The changes were cephalometrically evaluated.

The most usual cephalometric measurements found were nasolabial angle (NLA) (27, 29, 34, 46, 49–54) and distance from upper and lower lips to Ricketts Aesthetic line (UL-E and LL-E, respectively) (27, 29, 34, 46–49, 51, 52). Less frequently mentioned were the mentolabial angle (MLA) (46, 50, 53, 54), distance from upper and lower lips to Steiner’s S-line (UL-S and LL-S, respectively) (27, 47), and to Subnasal-soft-tissue pogonion line (UL-SnPg’ and LL-SnPg’, respectively) (27). All these variables and their respective values are reported in Supplementary Table 4.

When reported, the NLA increase ranged from 2.4 degrees (46) to 11.55 degrees (52) after 2-maxillary premolar extractions and from 1.0 degrees (34) to 6.84 degrees (53, 54) after 4-premolar extraction protocols.

The upper and lower lips were retracted during treatment regarding Ricketts E-line. The changes in 2-premolar extraction groups varied from −0.75 mm, for both lips (29), to −5.03 mm, for the upper lip (47). These changes in 4-premolar extraction groups varied from −2.29 mm, for the lower lip (34), to −4.9 mm (48), for the lower lip.

Retraction of the upper and lower lips was also shown regarding the distances from upper and lower lips to S-line and to the Subnasal-soft-tissue pogonion line. The changes ranged from −0.83 mm, for the lower lip (27), to −4.21 mm, for the upper lip (47) regarding the S-line and from −2.10 mm, for the upper lip, to −0.47 mm, for the lower lip, regarding the Subnasal-soft-tissue pogonion line, in 2-premolar extraction groups (27). The MLA increased 5.3 degrees in the 4-premolar extraction group (50).

Out of the 13 studies included in this systematic review, 6 were classified as presenting low risk of bias (27, 45, 48, 51, 53, 54), none had unclear risk of bias, and 7 were classified as high risk of bias (29, 34, 46, 47, 49, 50, 52) (Table 3).

Discussion
Inclusion and exclusion criteria
The objective of this systematic review was to disclose the soft-tissue changes reported in the literature following extraction treatment protocols of Class II division 1 malocclusion. The studies should have completely described the following items: malocclusion type and severity, comparison group, time point of evaluation, cephalometric changes, and statistical analyses. However, most studies did not contemplate all these aspects.

Extractions may be indicated by clinicians for distinct reasons, including canine Class II correction and overjet decrease. Almost all of the selected studies described an enlarged initial overjet, which is a common feature in Class II patients (Supplementary Table 2). As overjet decrease is a consequence of Class II treatment, it can be inferred that the main objectives of the extractions were canine relationship and overjet corrections.

The inclusion criteria did not distinguish between studies that analysed extractions of first or second premolars. Despite the thought...
that second mandibular premolar extractions could be more favourable in Class II malocclusion treatment with 4-premolar extractions, it has been shown that anchorage loss is similar between extractions of first or second premolars (55).

Several reasons prevented some articles to be included in this review. When missing data were essential for complete understanding of the soft-tissue profile changes, the article was not included. It was possible to notice the lack of standardization in the literature. Many articles were excluded because they did not have homogeneity in malocclusion severity, extraction protocols, or comparison groups.

Firstly, the type of treated malocclusion was identified. The more homogeneous is the sample, the more reliable are the results of the study. When the malocclusion was not completely specified, or when there was more than one type in the same group, it could not be concluded whether the described effects were applied to one or to the other malocclusion.

Malocclusion severity
When discussing Class II malocclusion treatment, description of the antero-posterior molar relationship severity is also essential. It is directly proportional to the amount of anterior retraction needed. The need in a half cusp Class II case is different from a full Class II, the worst-case scenario. In full Class II malocclusions, the treatment change in soft-tissue profile is more accentuated. When the extraction protocol was not clear, it was not even possible to check if the main question of this review had been contemplated.

Comparison group
A comparison group was required for the articles included in this review. Without it, the changes observed could be confused with growth (56), due to the mean age of some samples. The ideal comparison group would have had no treatment at all. But nowadays, it is difficult to have an untreated full Class II malocclusion sample, impairing measurement of actual treatment effects.

The extraction groups were compared to others submitted to many different treatment approaches. Some were treatment groups, such as headgear (27, 46, 47), functional appliances (45, 52), mini-implants (51); others were control groups (48). Therefore, a meta-analysis could not be performed.

Limitations
Not all the studies described the occlusal antero-posterior severity, which caused their exclusion from the current study. The authors from the included studies considered different parameters as adequate, for severity description. These differences may interfere in the soft-tissue changes.

There was no standardization in the literature even when it comes to group comparison. This lack of standardization suggests that the actual soft-tissue changes that follow premolar extractions in Class II malocclusion may not be very precise.

Soft-tissue changes after premolar extraction treatment protocols
Some variables such as patients’ age, ethnicity, and Class II malocclusion occlusal severity may affect the impact of anterior retraction in the soft-tissue profile. The extraction group ages were quite similar, with almost all studies having observed growing patients (Supplementary Table 2). Only three of them had adults as an inclusion criterion, because the extraction group was compared to a surgical group, which required non-growing patients (29). The others used groups with mini-implants as comparison (51, 52). Although lips retraction due to anterior retraction could be consequent to some antero-posterior mandibular growth, the same could have happened to almost all extraction groups, since all had comparable ages.

Different ethnicities could have played a role in the amount of change observed in the extraction groups as well. Different phenotypes are supposed to have different features such as greater soft-tissue thickness, nose shape, and degree of biprotrusion (57). Therefore, different populations with unique characteristics may respond differently to anterior retraction therapy. However, most of the studies included in this review did not provide details about their sample ethnicity or did not even mention it. Consequently, it was not possible to discuss how this difference may have made the soft-tissue changes vary from one study to another.

Several factors may influence treatment changes on the soft-tissue profile, such as mechanics, anchorage devices used, and phenotypic differences. These various factors may have contributed to the wide variation of results observed in the review. Nevertheless, it would be impossible to match all selected studies regarding all these kind of factors. In this context, it is worthwhile to observe the results sparingly, given the relevant information for each study, arranged in Supplementary Tables 2 and 3.

Several variables were used to evaluate the soft-tissue changes in premolar extraction groups. In order to analyse them, the more frequently used parameters were selected. They were: the NLA, the distance from the upper and lower lips to Ricketts Aesthetic Plane, to Steiner’s S-line, and the MLA (Table 4).

Nasolabial angle
Several authors have reported difficulty in precisely measuring the NLA and their variations, due to its great standard deviation (38). From all the studies that evaluated it, all reported an increase, either in 2- or 4-premolar extraction protocols.

Nevertheless, it is possible to observe that the groups submitted to two extractions had a greater range of variation in NLA changes only because of two outliers (51, 52), compared to patients who had undergone 4-premolar extractions. When there was extraction in both arches, the variation ranged was from 1.0 to 6.5 degrees (34, 48–50). When there were extractions only in the maxillary arch, the variation range increased from 2.4 to 5.40 degrees (2–7) with two outliers showing 9.08 degrees (51) and 11.55 degrees (52) of increase. It is important to observe that both outliers used mini-implants, suggesting the use of this anchorage system with caution when large soft-tissue changes are not desired (Supplementary Table 3).

In the groups subjected to 2-premolar extractions with traditional anchorage, the NLA increase was proportional to the occlusal antero-posterior malocclusion severity. In other words, the groups that had more severe Class II showed greater increases in NLA. This result is understandable, because it suggests greater anterior retraction, with little or no anchorage loss. This is what occurs in complete Class II cases treated with 2-premolar extractions.

In the studies that reported greater NLA increase, mini-implants were used as anchorage in 2-premolar extraction protocol, or four premolars had been removed. This is logical because mini-implants provide greater anchorage compared to other appliances. Besides, when four premolars are extracted, usually, the mandibular incisors will experience some retraction. Consequently, the maxillary incisors will have to be retracted an amount equal to the overjet combined with the amount of mandibular incisors retraction.

Lips projection
The groups that had 2-premolar extractions had greater upper lip retraction in relation to the lower lip. As expected, in the groups in
which four premolars were removed, the amount of retraction of the upper and lower lips had more similarity. At inter-groups comparison, both lips are retracted, with less retraction of the lower lip in the 2-premolar extraction protocol (Table 4).

Conclusions
According to studies with high scientific evidence, when Class II malocclusions are treated with 2-premolar extractions and traditional anchorage, the NLA may increase from 2.4 to 5.40 degrees. If mini-implants are used, it can increase up to 11.55 degrees and when treated with 4-premolar extractions, it increases from 1 to 6.84 degrees.

When extractions are performed, both lips are retracted, with less retraction of the lower lip in the 2-premolar extraction protocol. Therefore, these extraction protocols should be indicated when there is the respective excessive lip protrusions and the mentioned changes are primarily required to improve facial aesthetics.

Supplementary material
Supplementary material is available at European Journal of Orthodontics online.

Funding
São Paulo Research Foundation - FAPESP (2011/04603-6).

Acknowledgements
Authors thank FAPESP for the provided funding.

References


