

## Review

# Class II functional orthopaedic treatment: a systematic review of systematic reviews

V. D'ANTÒ\*<sup>†1</sup>, R. BUCCI\*<sup>1</sup>, L. FRANCHI<sup>‡</sup>, R. RONGO\*, A. MICHELOTTI\* &

R. MARTINA\* *\*Department of Neuroscience, Reproductive Sciences and Oral Sciences, School of Orthodontics and Temporomandibular disorders, University of Naples "Federico II", Naples, †Dentist Unit, Department of Pediatric Surgery, "Bambino Gesù" Children Hospital, Rome, and ‡Department of Surgery and Translational Medicine-Orthodontics, University of Florence, Florence, Italy*

**SUMMARY** This Systematic Review (SR) aims to assess the quality of SRs and Meta-Analyses (MAs) on functional orthopaedic treatment of Class II malocclusion and to summarise and rate the reported effects. Electronic and manual searches were conducted until June 2014. SRs and MAs focusing on the effects of functional orthopaedic treatment of Class II malocclusion in growing patients were included. The methodological quality of the included papers was assessed using the AMSTAR (Assessment of Multiple Systematic Reviews). The design of the primary studies included in each SR was assessed with Level of Research Design scoring. The evidence of the main outcomes was summarised and rated according to a scale of statements. 14 SRs fulfilled the inclusion criteria. The appliances evaluated were as follows: Activator (2 studies), Twin Block (4 studies), headgear (3 studies), Herbst (2 studies), Jasper Jumper (1 study), Bionator (1 study) and Fränkel-2 (1 study). Four studies

reviewed several functional appliances, as a group. The mean AMSTAR score was 6 (ranged 2–10). Six SRs included only controlled clinical trials (CCTs), three SRs included only randomised controlled trials (RCTs), four SRs included both CCTs and RCTs and one SR included also expert opinions. There was some evidence of reduction of the overjet, with different appliances except from headgear; there was some evidence of small maxillary growth restraint with Twin Block and headgear; there was some evidence of elongation of mandibular length, but the clinical relevance of this results is still questionable; there was insufficient evidence to determine an effect on soft tissues.

**KEYWORDS:** malocclusion angle class II/therapy, orthodontic appliances functional, review literature as topic, evidence-based dentistry, adolescent, growth and development

Accepted for publication 28 February 2015

## Background

Class II malocclusion is one of the most frequently encountered orthodontic issue as it occurs in about one-third of the population (1). The efficacy of the functional orthopaedic treatments for such malocclusion is a widely debated topic, with controversial results in orthodontic literature (2).

Systematic Reviews (SRs) and Meta-analyses (MAs) are generally considered appropriate study design for offering a strong level of evidence (3), especially on controversial topics. In addition, SRs are one of the best ways to stay up to date with current medical literature (4) instead of reading an average of 17–20 articles per day (5). A well-conducted SR aims to collect and synthesise all the scientific evidence on a specific topic, according to strict predetermined inclusion and exclusion criteria (6). When possible, SRs might

<sup>1</sup>These two authors contributed equally to this work.

be integrated with MAs to statistically contrast and combine results from different individual studies and to increase the statistical power of the analysis (7).

Approaching the scientific literature using such methodology might reduce the possibility of systematic errors (bias) (8). However, the validity of the results of SRs or MAs might be influenced by different factors; among those, the lack of methodological quality of the individual studies included in the review (9), and the methodological flaws in the development of the SR or MA itself must be taken into consideration.

In 2010, it has been estimated that about 75 trials and 11 SRs of trials were being published every day (10). Moreover, it is likely to find different SRs on the same topic, conducted with different aims and methodologies and leading to conflicting results (11).

In this scenario, the need of overviewing and comparing the results from the existent SRs in a single paper takes place (12). To point out the importance of such 'third level' of evidence, the Cochrane Collaboration has introduced the guidelines for *Overview of Reviews* (6), to summarise multiple Cochrane reviews addressing the effects of two or more potential interventions for a single condition.

To our knowledge, currently no Systematic Review of SRs concerning functional orthopaedic treatment of Class II malocclusion is available. Therefore, the aims of the present study were

- 1 To evaluate the methodological quality of SRs and MAs on functional orthopaedic treatment of Angle Class II malocclusion in growing patients. More specifically, to determine the methodological quality level of the SRs and MAs and to assess the design of the primary studies included in each SR or MA.
- 2 To provide an overview of the reported effects of the treatments and to rate the evidence on which these results are based.

## Materials and methods

The questions to be answered in the present SR are as follows:

- 1 What is the methodological quality level of the SRs and MAs addressing the effects of functional orthopaedic treatment of Class II malocclusion?

- 2 What are the main effects reported in the SRs and MAs about functional orthopaedic treatment of Class II Malocclusion in growing patients and what is the evidence underlying these results?

### Search strategy

For the current study, all the SRs and MAs concerning functional and orthopaedic treatment of Angle Class II malocclusion were analysed. The databases investigated for the systematic literature search were as follows: Medline (Entrez PubMed, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), Latin American and Caribbean Health Sciences (LILACS, <http://lilacs.bvsalud.org>), Scientific Electronic Library Online (SciELO, <http://www.scielo.org>) and the Cochrane Library ([www.cochrane-library.com](http://www.cochrane-library.com)). The survey covered the period from the starting of the databases (1966 for PubMed, 1997 for SciELO, 1982 for LILACS and 1993 for the Cochrane Library) up to September 2013. No language restrictions were set. A further hand-search of orthodontic journals (*European Journal of Orthodontics*, *American Journal of Orthodontics and Dentofacial Orthopedics* and *The Angle Orthodontist*) was performed starting from the first volume available on the digital archives, to include possible overlooked or *in press* papers. Moreover, an exploration of the grey literature (unpublished studies) was performed among the conference abstracts of scientific congresses (European Orthodontic Society and International Association of Dental Research).

The following keywords were used and adapted according to the database rules: 'Functional Orthodontic appliance', 'Angle Class II', Malocclusion, Review, Systematic Review. The search strategies applied for each database are shown in Table 1 (see also: Table S1).

The search was later updated, applying same strategies but customising the publication date range from September 2013 to June 2014.

### Studies selection and data collection

Inclusion criteria:

- 1 To be a Systematic Review or a Meta-Analysis;
- 2 Studies on the effects of functional orthopaedic appliances on Class II skeletal malocclusion;
- 3 Studies on growing patients.

**Table 1.** Search strategy for each database and relative results

Database	Search strategy	Results
PubMed	((‘Activator Appliances’ [Mesh]) OR ‘Orthodontic Appliances, Functional’ [Mesh] OR ‘Orthodontic Appliances, Removable’ [Mesh]) AND (‘Malocclusion, Angle Class II’ [Mesh]) AND (Review* OR Meta-Analysis*)	94
Cochrane Library	Malocclusion Angle Class II; Filter: Review	2
SciELO	Angle Class II Malocclusion AND (Review OR Meta-Analysis)	4
LILACS	(tw:(Angle Class II Malocclusion)) AND (tw:(Review))	23

Exclusion criteria:

- 1 Dual publication;
- 2 Systematic Review of SRs;
- 3 SR updated in a later publication;
- 4 Treatment protocol not involving functional orthopaedics.

Two investigators (V.D. and R.B.) independently read all titles and abstracts. Two of four databases (LILACS and SciELO) were analysed by only one investigator, due to language limitations. Subsequently, full-texts of the references that seemed to fulfil the inclusion criteria were acquired and analysed thoroughly. Finally, only the papers that completely satisfied all the inclusion criteria were selected. Disagreements between the two examiners were discussed and resolved to reach a unanimous consensus. In addition, the reference lists of the included SRs were analysed to identify any further relevant missing papers.

From the included papers data about Authors, Year of Publication, Study Design, Diagnosis, Number of Patients, Intervention, Control, Outcome, Quality of the included studies, Results, Author’s Conclusions and Author’s Comments on Quality of Studies were independently extracted by two authors (V.D. and R.B.), and the consensus was reached through discussion.

#### *Quality assessment of the included systematic reviews*

For each included SR, the methodological quality was assessed using the AMSTAR (Assessment of Multiple Systematic Reviews) (13). AMSTAR is composed by 11 items, each one can be answered ‘Yes’, when clearly done, ‘No’, when clearly not done, ‘Not Applicable’, when the item is not relevant, such as when a MA was not attempted by the authors, ‘Can’t

**Table 2.** Interpretation of the LRD scores. The scores are based on the type of studies included in the SR

LRD score	Studies included
I	Systematic Review of RCT
II	Randomised clinical trial
III	Study without randomisation, such as a cohort study, case-control study
IV	A non-controlled study, such as cross-sectional study, case series, case reports
V	Narrative review or expert opinion

answer’, when the item is relevant, but not described by the authors. Each ‘Yes’ answer is scored 1 point, while the other answers are scored 0 point. According to the number of criteria met, the quality of the included paper was rated as ‘Low’ (AMSTAR  $\leq 3$ ); ‘Medium’ (AMSTAR 4–7); ‘High’ (AMSTAR  $\geq 8$ ) (14, 15).

Moreover, to assess the design of the primary studies included in each SR the LRD (Level of Research Design scoring) was used (16, 17). The interpretation of such score, which is based on the hierarchy of evidence, is shown in Table 2.

For each included study, both investigators (V.D. and R.B.) independently assessed the methodological quality. There was no blinding for the authors during both quality assessment and data extraction. The inter-examiner reliability for the AMSTAR scores was calculated by means of Cohen’s  $k$  coefficient. Nonetheless, disagreements and discrepancies on the AMSTAR items were discussed and solved to reach a unanimous score.

#### *Synthesis of the results and rating of the evidence*

The main results of the included SRs were summarised according to the appliances examined in the

study. Afterwards, the evidence on which such results are based was rated according to a modified predetermined scale of statements (14, 15). The statements applied took into account: the way the data were pooled (MA or narrative synthesis), the statistical significance of the result and the number of studies/participants on which the result was based. A full explanation of the statements adopted is reported in Table 3. Moreover, a downgrade of the rating was performed (i.e. from sufficient evidence to some evidence) whenever the quality of most of the individual studies addressing a specific outcome was low. The quality of the individual studies was not re-assessed, but reported as assessed by the authors of the reviews.

## Results

### *Papers selection*

The updated electronic search of all databases resulted in 123 references. One article was retrieved from sources other than database, and it was an '*in press*' paper provided by the authors. After duplicates were removed, 115 references were left. Eighty-six references were excluded because the topic was not pertinent or because they were not SRs. The remaining eligible 29 articles were entirely read, and 15 of them were excluded (Fig. 1; Table S2). The most common exclusion criterion was the absence of a systematic search strategy, especially among the oldest papers.

The 14 SRs included and the data extracted from each SR are shown in Table 4 (18–31). One-third of the included SRs (5 of 14) were integrated with MA (18, 20, 28, 30, 31). The number of patients included ranged from 59 to 1763. The diagnosis reported in most of the paper was generally 'Angle Class II malocclusion'; six SRs (19, 23–26, 31) more specifically evaluated Class II Division 1 malocclusion and only in one study (27) vertical facial growth was taken into account as inclusion criterion (Class II hyperdivergent patients). Six SRs (18, 19, 21, 22, 27, 30) included only papers with a comparable Class II untreated group. The appliances studied in the included SRs were as follows: Activator (18, 26); Twin Block (18, 22, 25, 29); headgear (18, 20, 27); Herbst (19, 23); Jasper Jumper (24); Bionator (26); Fränkel-2 (30). Four papers evaluated several functional orthopaedic appliances, as a group (20, 21, 28, 31). The primary outcome of most of the articles (7 SRs) was the effect of treatment on the mandible, measured through different cephalometric methods and reference points.

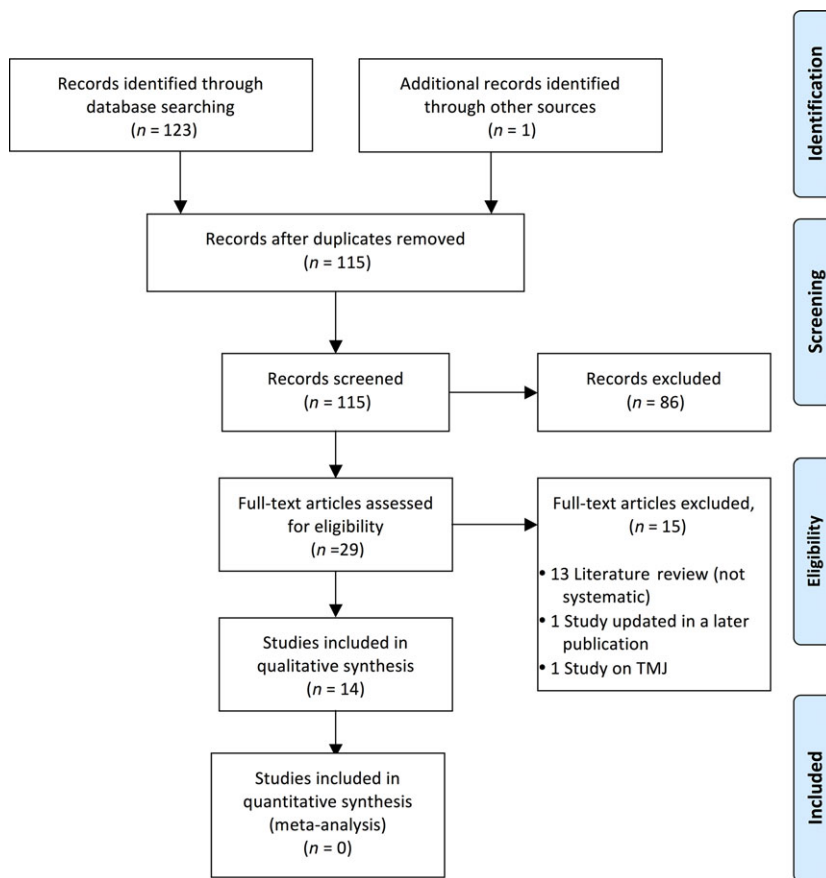
### *Quality of the included systematic reviews*

The Cohen's  $k$  coefficient for the AMSTAR items was 0.91, thus indicating very good interexaminer agreement.

The AMSTAR score ranged from a minimum of 2 to a maximum of 10; the mean score was 6. The single AMSTAR items for each paper and the total AMSTAR scores are shown in Table 5. Three papers

**Table 3.** Scale of Statements adopted to rate the evidence of the outcomes retrieved from each SR

Sufficient evidence	Meta-analysis: statistically significant pooled result that is based on a large number of included studies/participants or Narrative synthesis: large number of studies and/or study participants showing a statistical significance <i>When these conditions are applied to a non-significant result, the interpretation is 'evidence of no effect' (ineffectiveness).</i>
Some evidence	Meta-analysis: statistically significant pooled result that is based on a small number of included studies/participants or Narrative synthesis: small number of studies and/or study participants showing a statistical significance.
Insufficient evidence to support	Underpowering of the included studies to be able to detect an effect of the intervention (small number of studies/participant supporting significant or non-significant results) <i>Not to be interpreted as the first statement. This is about 'no evidence of effect or no evidence of no effect'.</i>
Insufficient evidence to determine	Gap in the evidence (controversial results)



**Fig. 1.** PRISMA Flow Diagram of the included and excluded records.

were rated as 'low quality', 8 papers were rated as 'medium quality', and 3 papers were rated as 'high quality'.

Six papers included only Clinical Controlled Studies (CCTs), three papers included only Randomised Controlled Studies (RCTs), four papers included both CCTs and RCTs, and one paper included also book chapter and expert opinions. The LRD scores are shown in Table 6.

#### *Main outcomes and rating of the evidence*

For this purpose, the papers showing low quality (20, 27, 29) (AMSTAR <4) were excluded.

**Dentoalveolar effects.** Three SRs (19, 22, 23) studied the dentoalveolar effects of functional orthopaedic treatment, while two SRs (18, 31) focused only on OVJ changes.

**Overjet (OVJ)**—There is *some evidence* that functional appliances, considered as a group, significantly

decrease the OVJ [−3.88 mm (19) to −4.17 mm (31)], with higher results for the Twin Block when assessed individually [−6.45 mm (19); −3.3 to −6.9 mm (22)].

There is *insufficient evidence to support* a significant reduction of the OVJ (−4.6 to −5.6 mm) with Splint-Type Herbst appliance (23).

There is *insufficient evidence to determine* an effect of the headgear on the OVJ as controversial results are reported: no significant effect was found by Antonarakis and Kiliaridis (18) while a small significant reduction was reported by Thiruvengatachari *et al.* (31) (−1.07 mm).

**Upper and lower incisors**—There is *some evidence* of proclination of the lower incisors (L1.GoGn: +3.9°) and retroclination of the upper incisor (U1.Mx plane: −9.2°) with Twin Block (22).

There is *insufficient evidence to support* a proclination/ anterior movement of the lower incisors with both Splint-Type (23) and Crown-Banded-Type Herbst appliance (19).

**Table 4.** Data extracted from the 14 systematic reviews and meta-analysis included

Authors, year, reference	Study design, diagnosis No. of patients	Intervention (I) Control groups (C)	Outcome measures	Quality tool and quality of the individual studies	Results	Authors' conclusions (C) Authors' comments on quality of studies (Q)
Antonarakis and Kiliaridis, 2007 (18)	SR and MA of 9 P CTs and RCTs; Class II; 670 subjects	II: Act (HA; Schwarz; Bio); I2: TB; I3: EOT; I4: Combination (EOT/functional); C: Untreated Class II subjects	Maxillary effect (SNA); Mandibular effect (SNB); Intermaxillary relation (ANB); Overjet	Petren <i>et al.</i> : Medium-High* (9/9)	Maxillary effect: Both I2 and I3 control maxillary growth; higher control with I3 (I3:1.03°, I2:1.01°), with lower homogeneity. No significant effect on SNA with I1 and I4 Mandibular effect: I1, I2 and I4 increase mandibular growth; greater effects and high homogeneity with I2(I1:0.66°; I2:1.53°; I4: 1.05°). No significant results on SNB with I3 Intermaxillary relation: All I1, I2, I3 and I4 reduce ANB angle; highest reduction with I2 (I1:0.92°; I2:2.61°; I3:1.38°; I4:1.8°), highest homogeneity with I4. Overjet: I1, I2 and I4 show a significant decreasing of the OJ; highest decrease with I2(I1:3.88 mm; I2:6.45 mm; I4:4.37), highest homogeneity with I4. No significant difference in OJ with I3 I1 determines increase of several mandibular sagittal skeletal variables (2-3 mm); minimal maxillary skeletal effects (few variables were statistically significant different), proclination/anterior movement of the lower incisors, retroclination/posterior movement of the upper incisors, extrusive and anterior direction of movement of the mandibular first molars; distal movement and intrusion of the maxillary first molars (clinically questionable). Overjet and overbite were also reduced	C: All appliances showed an improvement in sagittal intermaxillary relationships (decrease in ANB) when compared to untreated Class II subjects. The use of functional appliances and/or extraoral traction acts mostly in one of the two jaws (mandible for activators and combination appliances and maxilla for extraoral traction) while the twin block group, shows changes on both jaws. Besides the small sagittal skeletal base improvement influencing overjet, the dentoalveolar effect on overjet is brought about by palatal tipping of maxillary and labial tipping of mandibular incisors, respectively Q: Heterogeneity of age, sample size, control groups and appliances
Barnett <i>et al.</i> , 2008 (19)	SR of 3 CCTs; Class II Division I; 102 subjects	II: Crown- or Banded-Type Herbst C: Untreated Class II subjects	Dental and Skeletal cephalometric changes	-	II determines increase of several mandibular sagittal skeletal variables (2-3 mm); minimal maxillary skeletal effects (few variables were statistically significant different), proclination/anterior movement of the lower incisors, retroclination/posterior movement of the upper incisors, extrusive and anterior direction of movement of the mandibular first molars; distal movement and intrusion of the maxillary first molars (clinically questionable). Overjet and overbite were also reduced	C: Dental changes have more impact than skeletal changes Q: No RCT, Poor methodological quality of the studies; frequent use of condylion as reference point for mandibular length measurement, which is well-known to be difficult to determine cephalometrically. Different landmarks/measurements, different group age ranges, different treatment duration

(continued)

Table 4. (continued)

Authors, year, reference	Study design, diagnosis	Intervention (I) Control groups (C)	Outcome measures	Quality tool and quality of the individual studies	Results	Authors' conclusions (C) Authors' comments on quality of studies (Q)
Chen <i>et al.</i> , 2002 (20)	SR and MA of 6 RCTs; Class II; Not Reported	I1: Functional appliances (Bass, Bio, Fr-2, TB); C: No treatment and/or EOT	Mandibular growth (horizontal and vertical dimension)	–	I1 significantly increases only in Ar-Pg and Ar-Gn distances. No effect of the type of appliance	C: There is the need to re-evaluate functional appliance use for mandibular growth enhancement. The clinical effect on mandibular length is little, and probably influenced by reference point. (Ar moves posteriorly and superiorly after functional therapy) The absence of statistically significant difference of angular values (SNB and LIA) was unexpected and might be because all appliances were analysed as a group Q: Methodological limitation; heterogeneity of skeletal age, treatment durations; lack of treatment-control match and patient compliance evaluation
Cozza <i>et al.</i> , 2006 (21)	SR of 4 RCTs and 18 CTs (2 P and 16 R); Class II; 1763 subjects	I1: Functional appliances (Act, Bass, Bio, Fr-2, Herbst, MARA, TB) C: Untreated Class II subjects	Mandibular sagittal position, Total mandibular length, Mandibular ramus height, and Mandibular body length; Efficiency of the appliances	Modified Jadad scale: Low (3/22) Medium (13/22) Medium-High (6/22)	I1 increases mandibular growth in two-thirds of the samples. Changes in mandibular position in relation to the cranial base were not clinically significant The Herbst appliance showed the highest coefficient of efficiency (0.28 mm month) followed by the Twin block (0.23 mm month). Lowest coefficient of efficiency for the Frankel appliance (0.09 per month)	C: The amount of supplementary mandibular growth appears to be significantly larger if the functional treatment is performed at the pubertal peak in skeletal maturation. Q: Quality of the studies from low to medium/high, rare RCTs (neglecting the skeletal maturation)

(continued)

**Table 4.** (continued)

Authors, year, reference	Study design, diagnosis No. of patients	Intervention (I) Control groups (C)	Outcome measures	Quality tool and quality of the individual studies	Results	Authors' conclusions (C) Authors' comments on quality of studies (Q)
Ehsani <i>et al.</i> , 2014 (22)	SR of 10 CTs (6 P and 4 R) and Meta-Analysis of 5 studies; Class II; 664 subjects	II: TB C: Untreated Class II subjects	Skeletal, Dental and Soft tissues effects	Modified risk of bias: High risk (1/10) Medium risk (5/10) Low risk (4/10)	Data from the meta-analysis: II controls maxillary growth (SNA: 0.8°), projected the low jaw slightly forward (SNB: 1.2°), increments the mandibular body length (CoGn: 3 mm) and increases the anterior facial dimensions (ALFH: 2 mm). Moreover, reduces the upper incisor proclination (UI-AnsPns: 9.2°) and increases the lower incisor inclination (LI-GoGn: 3.8°)	C: Dental effects are consistently reported. A clinically significant restraint of maxillary growth was not found. Although the mandibular body length is increased, the facial impact of it is reduced by the simultaneous increment of the face height. Q: Highly heterogeneous and biased studies (various measurements and treatment times, use of historical controls)
Flores-Mir <i>et al.</i> , 2007 (23)	SR of 3 CCTs; Class II div I; Not Reported	II: Splint-Type Herbst; C: No treatment	Dental and Skeletal cephalometric changes	Self-produced checklist: Low (3/3)	Skeletal Effects: II increases anteroposterior length of the mandible (0.7- to 2.9-mm), increases mandibular protrusion (1.2° to 2.9°), decreases intermaxillary discrepancy (-1.5° to -2.1° and -4.2- to -4.9 mm), retrudes maxillary anteroposterior position (<1 mm) and increases posterior (1.4- to 2.5-mm) and anterior (1.2- to 3-mm) facial height Dental effects: II reduces OJ (-4.6- to -5.6-mm) and OB (-2.5 mm), determines mandibular incisor proclination (3.2° to 4.5°), protrusion (1.5- to 4-mm) and extrusion (5.3°) determines mesial movement of lower molars (0.8- to 3.6- mm) (no extrusion), and distal movement of upper molars (2.5- to 5.4- mm) with intrusion (-0.9 mm) and retroclination (5.6°). No significant changes for upper incisors	C: The combination of several small (statistically significant) changes in different skeletal and dental areas produces the overall reported positive change, but they are not likely clinically significant. Q: Secondary level of evidence. Small sample size. Use of different variables and reference points of cephalometric analysis. No homogeneity in treatment and control groups (race, gender, age). Few studies using control group that included Class II patients

(continued)



Table 4. (continued)

Authors, year, reference	Study design, diagnosis No. of patients	Intervention (I) Control groups (C)	Outcome measures	Quality tool and quality of the individual studies	Results	Authors' conclusions (C) Authors' comments on quality of studies (Q)
Flores-Mir <i>et al.</i> , 2006 (24)	SR of 5 CTs (1 P and 4 R); Class II div I; 228 subjects	I1: Jasper Jumper; I2: Herbst; C: No treatment	Soft tissue changes	Self-produced checklist: Low (2/5) Medium (3/5)	I1 increases naso-labial angle, retrudes the position of 'Labrale Superius' relative to the vertical reference plane, and protrudes the position of 'Labrale Inferius' relative to Aesthetic Plane (E-plane) I2 generates a soft menton protrusion, a 'Subnasale' retrusion, contradictory results regarding the anteroposition of the upper lip and no changes in the lower lip	C: There is little of evidence on Jasper Jumper appliance and the results are contradictory. Herbst appliance determines a significant improvement in facial profile. This improvement is not the product of a more forward position of the lower lip but more likely a retrusion of the upper lip. On average, although fixed functional appliances produce some significant statistical changes in the soft tissue profile, the magnitude of the changes may not be perceived as clinically significant. Q: Low level of evidence; reference structures not always reliable
Flores-Mir and Major, 2006a (25)	SR of 2 CCTs; Class II div I; 59 subjects	I1: TB; C: No treatment	Soft tissue changes	Self-produced checklist: Low (1/2) Medium (1/2)	I1 shows no evidence of change of facial convexity, lower lip, nose and soft tissue menton. Controversial changes in the upper lip for the position of <i>labrale superius</i> relative to the aesthetic line, which was in a more retruded position (-1.9 mm)	C: A few studies evaluated the soft tissue profile changes. The twin block appliance seems to not produce a soft tissue profile changes to be perceived as clinically significant. Three-dimensional quantification of soft tissue changes is required Q: Low level of evidence
Flores-Mir and Major, 2006b (26)	SR of 10 CCTs and 1 RCT; Class II div I; 540 subjects	I1: Act; I2: Bio; C: No treatment	Soft tissue changes	Self-produced checklist: Low (10/11) Medium (1/11)	Contradictory results for both I1 and I2	C: Although some statistically significant soft tissue changes were found, for both I1 and I2 the clinical significance is questionable. Q: Methodological weakness; low level of evidence

(continued)

Table 4. (continued)

Authors, year, reference	Study design, diagnosis No. of patients	Intervention (I) Control groups (C)	Outcome measures	Quality tool and quality of the individual studies	Results	Authors' conclusions (C) Authors' comments on quality of studies (Q)
Jacob <i>et al.</i> , 2013 (27)	SR of 4 CCTs; Class II hyperdivergent patients; 221 subjects	II: Extraoral high-pull headgear (maxillary splints/banded molars) C: Untreated Class II hyperdivergents	Skeletal changes (horizontal and vertical); Dental effects (molar eruption)	Modified Antczak <i>et al.</i> : Low – 4 points of 10 (1/4) Medium – 6/7 points of 10 (3/4)	II decreases ANB angle (from 0.9 to 1.5°), decreases overjet (2.6–6.5 mm). Statistically significant posterior displacement of the maxilla (0.1–0.5 mm), distalization of the maxillary molar (0.5–3.3 mm), maxillary molar intrusion (0.4–0.7 mm), retroclination (4.4–11.0°) and intrusion of the maxillary incisors (0.2–2.1 mm) were also reported with II. No effects on the mandible II increases mandibular growth (1.79 mm in the annual mandibular growth) when compared with C, with statistical heterogeneity	C: High-pull headgear treatment improved the AP skeletal relationships, by displacing the maxilla posteriorly but not the vertical skeletal relationships Q: Greater attention to the design and report of studies should be given to improve the quality of such trials
Marsico <i>et al.</i> , 2011 (28)	SR and MA of 4 RCTs; Class II; 338 subjects	II: Functional Appliances (Act, HA, Fr-2, Bio, TB); C: No treatment	Mandibular growth (total length)	Assessment of risk of bias: High risk (1/4) Unclear risk (1/4) Low risk (2/4)	II produces significant reduction of the SNA angle; retroclination of the upper incisors, increase of mandibular length and condyle growth; proclination of the lower incisors; improvement of maxillary mandibular relation	C: The treatment with functional appliances results in change of skeletal pattern (small increases of mandibular length); however, even if statistically significant, appear unlikely to be very clinically significant. The heterogeneity of the results can be attributed to the difference in sample dimension and to the use of different functional appliances). Several benefits must be attributed to the early treatment of Class II malocclusion with functional appliances Q: Heterogeneity regarding cephalometric analyses, variables and reference points C: The alterations were the combination of modifications on the condyle, the mandibular fossa, the basal bone and dentoalveolar alterations. Most of the authors recommend the use of the Twin Block during the pubertal peak Q: nr
Olibone <i>et al.</i> , 2006 (29)	SR of 45 references (articles and book's chapters) Class II; Not reported	II: TB; C: Not reported	Mandibular growth; Maxillary effect; Intermaxillary relation; Upper incisor; Lower incisor	–	II produces significant reduction of the SNA angle; retroclination of the upper incisors, increase of mandibular length and condyle growth; proclination of the lower incisors; improvement of maxillary mandibular relation	C: The alterations were the combination of modifications on the condyle, the mandibular fossa, the basal bone and dentoalveolar alterations. Most of the authors recommend the use of the Twin Block during the pubertal peak Q: nr

(continued)

Table 4. (continued)

Authors, year, reference	Study design, diagnosis No. of patients	Intervention (I) Control groups (C)	Outcome measures	Quality tool and quality of the individual studies	Results	Authors' conclusions (C) Authors' comments on quality of studies (Q)
Perillo <i>et al.</i> , 2010 (30)	SR and MA of 8 CTs (7 R and 1 P) and 1 RCT; Class II; 686 subjects	I1 Fr-2 (FR-2); C: Untreated Class II subjects	Mandibular total length; Mandibular body length; Mandibular ramus height	Modified Jadad scale: Low (7/9) Medium (2/9)	I1 enhances mandibular body length (0.4 mm year), mandibular total length (0.021 mm year) and mandibular ramus height (0.654 mm year)	C: The FR-2 appliance had a statistically significant effect on mandibular growth with a low to moderate clinical impact Q: From low to medium quality of the studies. Heterogeneity in linear measurement, age distribution and treatment duration. Poorly defined initial skeletal diagnosis. Mostly non-randomised and retrospective
Thiruvengatchari <i>et al.</i> , 2013 (31)	SR and MA of 17 RCTs; Prominent Upper Teeth (Class II division 1); 721 subjects	AIM 1: I1: Functional Appliance (TB); Forsus; Andreasen; Fr-2; Bass; Bio; R-Appliance; Dynamax; HA; AIBP; Herbst) I2: EOT C: Adolescent Treatment AIM2: I1: Functional Appliances C: No treatment or different kind of appliance	OVJ; Intermaxillary relation (ANB); Incisal trauma	Assessment of risk of bias: High risk (11/17) Unclear risk (4/17) Low risk (2/17)	I1 and I2 demonstrate significant difference in OVJ and ANB when compared with C, after the first phase of early treatment. At the end of the treatment, no statistically significant difference, except for a significant reduction in the incidence of incisal trauma	C: Early orthodontic treatment for children with prominent upper front teeth is more effective in reducing the incidence of incisal trauma than adolescent orthodontic treatment. There are no other advantages for providing early treatment Q: Overall low quality of the evidence

SR, systematic review; MA, meta-analysis; CCT, controlled clinical trial; P, prospective; R, retrospective; RCT, randomised controlled trial; Act, activator; TB, twin block; EOT, extra oral traction; Bio, bionator; Fr-2, Frankel-2; MARA, mandibular anterior repositioning appliance; HA, harvold activator; AIBP, anterior inclined bite plate; \*stated by the authors. Quality not reported for the individual studies.

**Table 5.** Quality assessment according the AMSTAR items for each SR and total AMSTAR scores. For each Yes answer: 1 point; all the other answers: 0 point

	Antonarakis and Kiliaridis, 2007 (18)	Barnett et al., 2008 (19)	Chen et al., 2002 (20)	Cozza et al., 2006 (21)	Ehsani et al., 2014 (22)	Flores-Mir et al., 2007 (23)	Flores-Mir et al., 2006 (24)	Flores-Mir and Major, 2006a (25)	Flores-Mir and Major, 2006b (26)	Jacob et al., 2013 (27)	Marsico et al., 2011 (28)	Olibone et al., 2006 (29)	Perillo et al., 2010 (30)	Thiruveng- atachari et al., 2013 (31)
Was an 'a priori' design provided?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Was there duplicate study selection and data extraction?	N	Y	N	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y
Was a comprehensive literature search performed?	Y	Y	N	N	Y	Y	Y	Y	Y	N	Y	Y	Y	Y
Was the status of publication (i.e. grey literature) used as an inclusion criterion?	N	N	N	N	Y	N	N	N	N	N	N	N	N	Y
Was a list of studies (included and excluded) provided?	N	Y	Y	N	Y	Y	N	N	Y	N	Y	N	N	Y
Were the characteristics of the included studies provided?	CA	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y
Was the scientific quality of the included studies assessed and documented?	N	CA	N	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y
Were the methods used of the included studies used appropriately in formulating conclusions?	Y	Y	N	NA	Y	Y	NA	NA	NA	NA	Y	NA	Y	Y
Were the methods used to combine the findings of studies appropriate?	N	N	N	NA	N	N	NA	NA	NA	NA	N	NA	N	N
Was the likelihood of publication bias assessed?	N	N	N	N	N	N	N	N	N	N	N	N	N	Y
Was the conflict of interest stated?	4	6	3	5	3	7	7	6	7	9	8	2	7	10
<i>Total AMSTAR Score</i>														

Y, yes; N, no; NA, not applicable; CA, can't answer.

**Table 6.** Study design of the primary studies included in each SR, as assessed according to the LRD scores

Authors, Year, Reference	LRD
Antonarakis and Kiliaridis, 2007 (18)	II–III
Barnett <i>et al.</i> , 2008 (19)	III
Chen <i>et al.</i> , 2002 (20)	II
Cozza <i>et al.</i> , 2006 (21)	II–III
Ehsani <i>et al.</i> , 2014 (22)	III
Flores-Mir <i>et al.</i> , 2007 (23)	III
Flores-Mir <i>et al.</i> , 2006 (24)	III
Flores-Mir and Major, 2006a (25)	III
Flores-Mir and Major, 2006b (26)	II–III
Jacob <i>et al.</i> , 2013 (27)	III
Marsico <i>et al.</i> , 2011 (28)	II
Olibone <i>et al.</i> , 2006 (29)	III–IV–V
Perillo <i>et al.</i> , 2010 (30)	II–III
Thiruvnjatachari <i>et al.</i> , 2013 (31)	II

I: Systematic Review of RCTs; II: Randomised Clinical Trial; III: Study without randomisation; IV: non-controlled study, V: Narrative review/expert opinion.

Upper and lower molars—There is *insufficient evidence to support* a distal movement and intrusion of the upper molars and a mesial movement of the lower molars, reported with Splint-Type (23) and Crown-Banded-Type Herbst appliance (19).

There is *insufficient evidence to determine* a mesio-distal movement of upper and lower molars with Twin Block, due to the controversy of the findings (22).

### Maxillary skeletal effects

Four SRs evaluated the effects of treatment on the upper jaw (18, 19, 22, 23).

There is *some evidence* of a small maxillary growth restraint with Twin Block appliance [SNA:  $-0.7^\circ$ (22) to  $-1.03^\circ$ (18)] and with headgear [SNA:  $-1.01^\circ$ (18)].

There is *some evidence* of a non-significant effect with other activators, considered as a group (Harvold, Bionator, Schwarz) (19).

There is *insufficient evidence to determine* the effect of both Splint-Type (23) and Crown-Banded-Type Herbst Appliance (19) on the upper jaw, which is reported to be very low or even not significant.

### Mandibular skeletal effects

Seven SRs analysed the effects of functional orthopaedic treatment on the lower jaw (18, 19, 21–23, 28, 30).

There is *some evidence* of a significant advancement of mandibular position in relation to cranial base (SNB) with Twin Block appliance [ $1.2^\circ$ (22);  $1.53^\circ$ (18)], while *some evidence* of a very small increase of the same angle was reported with other activators, considered as a group [Harvold, Bionator, Schwarz;  $0.66^\circ$ (18)].

There is *some evidence* of mandibular length increasing after treatment with functional appliances, considered as a group, ranging between 0.8 and 4.7 mm as measured with Co-Gn (or Co-Pg) and between 1.2 and 2.2 mm as measured with Olp-Pg + OLp-Co (21). The same results was reported with an effect size of 0.61 (28).

There is *some evidence* of a significant elongation of Co-Gn with Fränkel-2 (30) appliance and Twin Block (22) appliance individually (1.07 mm/year and 2.9 mm, respectively).

There is *insufficient evidence to support* a significant mandibular length increasing with both Splint-type (23) (0.7–2.7 mm) and Crown-Banded-Type (19) (1.6–2.2 mm) Herbst appliances.

### Soft tissue effects

Four SRs evaluated the effects of functional orthopaedic treatment on soft tissues (22, 24–26).

There is *insufficient evidence to support* an improvement in facial convexity after treatment with fixed appliances (Jasper Jumper (J) and Herbst (H)) (24). In particular, the increase of the naso-labial angle (J) or the retrusion of subnasale point (H) and the protrusion of labrale inferior point (J) or the protrusion of the soft menton (H) are reported.

There is *insufficient evidence to determine* an effect of Twin Block (22, 25) on soft tissues due to the controversy of the reported results: in fact, significant effects were reported in the one SR (22) while non-significant findings were pointed out in another paper (25).

There is *insufficient evidence to determine* an effect on soft tissues with Activator and Bionator as controversial results are reported in one SR (26).

### Discussion

The present Systematic Review aimed to summarise the current evidence from the SRs and MAs on the orthopaedic functional treatment of Class II Malocclusion. In particular, the focus of the present study con-

cerns the quality and the main results of the SRs and MAs addressing this issue.

#### *Quality of the included systematic reviews*

Scientific and rigorous methods are employed in SRs to identify and summarise the literature, to minimise biases that come from narrative reviews. Nonetheless, as with all the other publications, the value of a SR depends on the way it is conducted and on the accuracy of the results (10).

The methodological quality of the included SRs was assessed with the AMSTAR (Assessment of Multiple Systematic Reviews) (13). AMSTAR is a recent valid and reliable quality tool (32), built upon expert opinion and empirical data collected with a previously developed tool (13).

The item 1 of the AMSTAR ('Was an "a priori" design provided?') refers to a registered protocol of the review. The databases for protocol registration, such as PROSPERO (International Prospective Register of Systematic Review) (33), have been recently introduced; therefore, in our study, due to a chronological limitation the presence of the protocol registration was neglected. Affirmative answer to the item 1 was assigned whenever clear predetermined research criteria were provided. Ensuring such approach avoids the review method to be influenced by reviewers' expectations (7).

The AMSTAR scores of the SRs included in the current study showed a wide range of values, between 2 and 10, with an average value of 6. Common factors for the included review to lose point in the AMSTAR score were as follows: not performing a grey literature search (item 4), not assessing the publication bias (item 10) and not providing the conflict of interest of the authors (item 11). However, AMSTAR score have to be carefully interpreted as the single AMSTAR items may have different weights in the overall quality of a SR (34). For instance, reporting the conflict of interest (item 11) has a low impact on the methodology of a SR. On the other hand, the assessment of the scientific quality of the primary study included (item 7) has to be considered a key item, as this evaluation allows the identification of flaws in the primary literature. In 10 of 14 SRs, the quality of the individual studies was documented and reported. Modified Jadad Scale (35) and Assessment of risk of bias (36) were

the most used tools, together with self-produced checklists based on the key of interest, which are also considered valid instruments (9, 24–26). Among the included studies, only the Cochrane review (31) adopted the GRADE approach (37) suggested from the Cochrane collaboration as system for grading the quality of evidence and providing the strength of recommendation.

The paper with the highest AMSTAR score (AMSTAR 10) is a Cochrane Review (31). This result is in accordance with what previously pointed out in several studies (38–40) when comparing the methodology of Cochrane SRs with that of SRs published in paper-based journals; the authors found that the SRs published by the Cochrane Collaboration present less flaws and better methodological quality. These findings suggest that standardised instructions and several peer-review levels improve the methodological soundness of literature.

The AMSTAR score evaluates whether a SR is conducted in appropriate way, but still it neglects information regarding the individual articles included in the SR. To overcome this issue, the AMSTAR score was integrated with the LRD score. The Level of Research Design Scoring has been previously adopted in SR of SRs (34), and it assigns a score to the design of the individual studies according to the hierarchy of evidence (16, 17).

Only one SR (29) included non-controlled studies, book chapters and expert opinions (LRD III-IV-V). This SR showed also the lowest AMSTAR score (AMSTAR 2) and presented a structure closer to a narrative review than to a SR, without providing any definite conclusion. However, it was included in our study because the methodology of the literature search reflects some of the principles of a SR.

Most of the included reviews (6 SRs) included only CCTs. Even if RCTs are considered the best way to investigate the efficacy of dental interventions and to compare different treatment alternatives (41), and MAs of RCTs are considered one of the highest level of evidence (7, 42, 43) only 3 of the included SRs (20, 28, 31) investigated only RCTs. The number of RCTs included in these SRs was variable [6 for Chen *et al.* (20), 4 for Marsico *et al.* (28) and 17 for Thiruvengatachari *et al.* (31)] and only 2 studies overlapped in the 3 searches, because of different inclusion and exclusion criteria.

Interestingly, one of the SR of RCTs (20) was judged of low quality with the AMSTAR score (AMSTAR 3), demonstrating that even the results of a SR of RCTs, which pretends to be the highest level of evidence, have to be carefully interpreted as major methodological flaws can affect the quality of the SR.

#### *Main outcomes and rating of the evidence*

To not provide a simple narrative summary of the results and to assess the quality of body evidence, a predetermined scale of statements was adopted for each of the outcomes analysed. This instrument has been previously adopted in a Cochrane SR of SRs (14, 15), to not re-assess the quality of the studies included within reviews. In the current study, it was not possible to adopt the GRADE approach (37) as suggested by the Cochrane Collaboration, as 'Summary of findings' tables were not reported in any of the included SRs, except for the Cochrane SR (31) and frequently raw data were not available.

The difficulties encountered in our study when synthesising the data extracted from the included SRs and MAs were mainly due to the variability of the inclusion criteria and to the heterogeneity of samples, outcomes, cephalometric landmarks and analysis. Our study pointed out a strong weakness in the initial diagnosis of skeletal Class II malocclusion. All the included SRs set 'Class II malocclusion' as inclusion criterion, but none of them clearly stated how the diagnosis was performed. It was observed that treatment success with functional appliances depends on a great number of confounding variables, including the severity of the baseline conditions. Underestimating this factor does not guarantee generalisation of the conclusion, as the sample might not properly represent the target population (44).

Results from SRs and MAs should be the cornerstone for developing practice guidelines, but due to the limited and biased evidence of the primary studies, the clinical recommendations are always reported to be weak. The most frequently reported flaws of the primary studies were as follows: methodological limitations, absence of a control-matched untreated group, variability of the treatment timing, small sample size and variability of cephalometric analysis and landmarks.

*Dentoalveolar effects.* According to the results provided by the included SRs and MAs, there is a good consensus in literature regarding the effect of reduction of the OVJ after functional orthopaedic treatment. Nevertheless, if the results of the functional appliances in general and of the Twin Block in particular are supported by a good level of evidence, it is not so for the Splint-Type Herbst appliance. Indeed, the SR by Flores-Mir *et al.* (23) which provides results on this outcome is based only on three references judged of low quality by the authors. Regarding the headgear, the evidence supporting the effect on the OVJ was considered insufficient, due to the controversy of results. These controversies are probably related to the different study selection [all studies (18) vs. RCTs (31)] and to the different inclusion criteria of the studies assessing this outcome. In fact, Antonarakis and Kiliaridis (18) chose as diagnostic criterion the Class II malocclusion, while Thiruvengkatachari *et al.* (31) selected the participants as they presented prominent upper front teeth. Therefore, it is likely to observe a greater dental movement when the starting position of the teeth is altered.

Changes in molar position were reported to be small and generally supported by insufficient evidence.

Little information is reported about the long-term effects after functional treatment. In one SR (18), it is reported that skeletal changes seem to be more temporary than dentoalveolar changes, which are more stable.

*Maxillary skeletal effects.* Regarding the evidence provided on maxillary growth restraint, few significant values were reported and most of them were too small to be considered clinically relevant. The best effect of SNA reduction seems to be achieved with headgear (18), while Twin Block shows variable results between significant and non-significant (18, 22). Non-significant values of maxillary growth control were reported with both Splint-Type and Crown-Banded Herbst, but the evidence supporting this result is insufficient due to the small number of primary studies (2 or 3 studies) on which this result is based. In addition, the quality of the individual studies was low in the SR by Flores-Mir *et al.* (23), and even not assessed in the study by Barnett *et al.* (19). Therefore, the current evidence from SRs is not adequate to sug-

gest or discourage the use of Herbst appliance for maxillary skeletal growth control.

*Mandibular skeletal effects.* Enhancement of mandibular length and/or achievement of a more forward position of the mandible, albeit still widely discussed, are frequently desired outcomes as most of the skeletal Class II malocclusion are due to a mandibular retrusion (45).

Addressing all functional appliances as a group Cozza *et al.* (21) reported a wide range of significant and non-significant findings, providing results which are scarcely applicable in the daily practice. The variability of the results in this SR is probably due to the inclusion of retrospective studies, which are susceptible to selection bias, and studies with historical samples, which suffer from the secular growth trends, occurred within the craniofacial region over the past century (46). Moreover, data from treatment with removable and fixed appliances were pooled in this review: this choice can influence the results as the two techniques differ for working hours, length of treatment time, optimal treatment timing and mode of bite-jumping (47). Considering the primary studies included in this SR, in which the pubertal peak was included in the treatment timing, clinical significance of supplementary mandibular elongation (>2 mm) was reported in all studies except one. According to this finding, the authors of this SR support the hypothesis that the short-term supplementary mandibular growth appears to be significantly larger when the functional treatment is performed at the adolescent growth spurt.

Even though all the SRs and MAs included in our study set the treatment of growing subjects as inclusion criterion, none of them put efforts in assessing the skeletal age. Only in one MAs (18), the studies were included only if the age of the participants was reported.

Barnett *et al.* (19) and Flores-Mir *et al.* (23) reported a significant elongation of the mandible with Crown-Banded and Splint-Type Herbst Appliance, respectively, but the literature supporting these outcomes was judged to be insufficient due to the small number and low quality of the primary studies. Comparing the effect of Acrylic-Splint Herbst with Crown or Banded Herbst Appliance, the differences seem to be small and not relevant, but more research is needed on this issue.

In the MA by Perillo *et al.* (30) on Fränkel-2 appliance, a significant but small increase of mandibular total length was found. However, the sensitivity analysis pointed out a negative correlation between the quality of the included studies and the retrieved results, making questionable the clinical relevance of the findings.

The most recent MAs (28) points out an effect size of the treatment of 0.61 when comparing Class II subjects treated with different functional appliances with untreated control groups. This finding is the result of the standardisation of different cephalometric measures of mandibular length, which accounts differently for jaw divergence (Co-Pg, Co-Gn and Olp-Pg+OLp-Co). In addition, the amount of mandibular length reported as the result of the conversion of the effect size (1.79 mm) is higher than that reported in the individual studies included in the SR. This controversy pointed out that major flaws could affect also a MA of RCTs rated of high quality with the AMSTAR score.

*Soft tissues effects.* Regarding soft tissues, better results seem to be obtained with fixed functional appliances than with removable, especially when Herbst appliance is used (24–26). The authors report the improvement of the profile to be mainly due to the retrusion of the upper lip, rather than to the protrusion of the lower lip. However, all the SRs assessing this outcome reported controversial results based on the low-quality primary studies; hence, this evidence has to be considered insufficient.

In addition, none of the primary studies included in the three SRs assessed the changes in facial profile by means of three-dimensional scanning, which is considered a reliable, non-invasive and free of radiation technique for assessing facial form (48). Due to the superimposition of the hard tissues, conventional cephalometric analyses are considered not adequately capable to detect the soft tissue structure, so the results regarding the soft tissues effects might have been underestimated.

## Future research

According to our findings, the registration of the protocol and the implementation of the use of PRISMA guidelines (10) might improve the methodological quality of future SRs. In addition, the use of the GRADE as tool to assess the quality of the primary



studies and to provide the strength of recommendation can give a substantial contribution to the clinical conclusions and give more values to the future evidence from SRs of SRs.

Moreover, it seems more useful for future SRs to analyse more homogeneous group of patients (selected according initial diagnosis, skeletal maturation and vertical growth pattern) and appliances, as reporting an aggregate pooled effect might be misleading if there are important reasons to explain variable treatment effects across different types of patients (7). Finally, the evidence from the included SRs and MAs demonstrates that more research is needed on long-term effects of functional orthopaedic treatment.

## Conclusions

- 1 The SRs on functional orthopaedic treatment of Class II malocclusion present a heterogeneous methodological quality. Only two SRs were judged of high quality.
- 2 Three of the 14 papers analysed, include *only* RCTs and numerous SRs report a low quality of the individual studies.
- 3 Clinicians should be aware of the existent tool to assess strength and weakness of the SRs and MAs, to adequately recognise whenever limited information can be obtained from such studies.
- 4 In general, there is still no sufficient evidence to suggest or to discourage the orthopaedic functional treatment in Class II patients. The lack of definite evidence is mainly due to the small number of primary studies for each outcome and the low quality of most of the individual studies.
- 5 There is some evidence of reduction of OVJ with several functional appliances, except from Herbst appliance, due to the poor quality of literature, and headgear, due to the controversial results reported with this appliance.
- 6 There is some evidence of a small maxillary growth control with headgear and Twin Block.
- 7 In the short term, there is some evidence of mandibular length increasing after treatment with several functional appliances, but not with Herbst appliance, which presents poor quality of literature. However, the clinical relevance of the reported

results is still questionable and long-term data are not available.

- 8 There is insufficient evidence to support the effect of functional orthopaedic treatment on soft tissue.

## Disclosure/Acknowledgments

The authors of this manuscript declare that they have no conflict of interest. This study did not receive any funding.

## References

1. McLain JB, Proffit WR. Oral health status in the United States: prevalence of malocclusion. *J Dent Educ.* 1985;49:386–396.
2. Aelbers CMF, Dermaut LR. Orthopedics in orthodontics: part I, fiction or reality? A review of literature. *Am J Orthod Dentofac Orthop.* 2002;121:560–562.
3. Papadopoulos MA, Gkiauouris I. A critical evaluation of meta-analyses in orthodontics. *Am J Orthod Dentofac Orthop.* 2007;131:589–599.
4. Davidoff F, Haynes B, Sackett D, Smith R. Evidence-based medicine: a new journal to help doctors identify the information they need. *Br Med J.* 1995;31:1085–1086.
5. Lau J, Ioannidis JPA, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet.* 1998;351:123–127.
6. Becker LA, Oxman AD. Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. In: Higgins JPT, Green S, eds. *The cochrane collaboration.* 2011. Available at: [www.cochrane-handbook.org](http://www.cochrane-handbook.org), accessed 15 March 2015.
7. Garg AX, Hackam D, Tonelli M. Systematic review and meta-analysis: when one study is just not enough. *Clin J Am Soc Nephrol.* 2008;3:253–260.
8. Mulrow CD. Rationale for systematic reviews. *Br Med J.* 1994;3:597–599.
9. Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *Br Med J.* 2001;323:42–46.
10. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *Br Med J.* 2009;339:b2700.
11. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med.* 2010;7:e1000326.
12. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Med Res Methodol.* 2011;11:15.
13. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C *et al.* Development of AMSTAR: a measurement

- tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol.* 2007;15:10.
14. Ryan R, Santesso N, Hill S, Lowe D, Kaufman C, Grimshaw J. Consumer-oriented interventions for evidence-based prescribing and medicines use: an overview of systematic reviews. *Cochrane Database Syst Rev.* 2011;(11):CD007768.
  15. Ryan R, Santesso N, Lowe D, Hill S, Grimshaw J, Prictor M *et al.* Interventions to improve safe and effective medicines use by consumers: an overview of systematic reviews. *Cochrane Database Syst Rev.* 2014;(29):CD007768.
  16. Antes G. Evidence-based medicine. *Internist (Berl).* 1998;39:899–908.
  17. Cooke IE. Finding the evidence. *Baillière's Clin Obstet Gynaecol.* 1996;10:561–567.
  18. Antonarakis GS, Kiliaridis S. Short-term anteroposterior treatment effects of functional appliances and extraoral traction on class II malocclusion. A meta-analysis. *Angle Orthod.* 2007;77:907–914.
  19. Barnett GA, Higgins DW, Major PW, Flores-Mir C. Immediate skeletal and dentoalveolar effects of the crown- or banded type Herbst appliance on Class II division 1 malocclusion. *Angle Orthod.* 2008;78:361–369.
  20. Chen JY, Will LA, Niederman R. Analysis of efficacy of functional appliances on mandibular growth. *Am J Orthod Dentofac Orthop.* 2002;122:470–476.
  21. Cozza P, Baccetti T, Franchi L, De Toffol L, McNamara JA Jr. Mandibular changes produced by functional appliances in Class II malocclusion: a systematic review. *Am J Orthod Dentofac Orthop.* 2006;129:e1–e12.
  22. Ehsani S, Nebbe B, Normando D, Lagravere MO, Flores-Mir C. Short-term treatment effects produced by the Twin-block appliance: a systematic review and meta-analysis. *Eur J Orthod.* 2014;37:170–176.
  23. Flores-Mir C, Aye H, Goswami A, Charkhandeh S. Skeletal and dental changes in Class II division 1 malocclusions treated with splint-type Herbst appliances. A systematic review. *Angle Orthod.* 2007;77:376–381.
  24. Flores-Mir C, Major MP, Major PW. Soft tissue changes with fixed functional appliances in Class II division 1. *Angle Orthod.* 2006;76:712–720.
  25. Flores-Mir C, Major PW. Cephalometric facial soft tissue changes with the twin block appliance in Class II division 1 malocclusion patients. A systematic review. *Angle Orthod.* 2006;76:876–881.
  26. Flores-Mir C, Major PW. A systematic review of cephalometric facial soft tissue changes with the Activator and Bio-nator appliances in Class II division 1 subjects. *Eur J Orthod.* 2006;28:586–593.
  27. Jacob HB, Buschang PH, dos Santos-Pinto A. Class II malocclusion treatment using high-pull headgear with a splint: a systematic review. *Dental Press J Orthod.* 2013;18:e1–e7.
  28. Marsico E, Gatto E, Burrascano M, Matarese G, Cordasco G. Effectiveness of orthodontic treatment with functional appliances on mandibular growth in the short term. *Am J Orthod Dentofac Orthop.* 2011;139:24–36.
  29. Olibone V, Lys L, Guimarãe AS, Atta JY. Influência do aparelho propulsor Twin Block no crescimento mandibular: revisão sistemática da literatura/Influence of the Twin Block propulsor appliance in the mandibular growth: systematic review of literature. *Dental Press J Orthod.* 2006;11:19–27.
  30. Perillo L, Cannavale R, Ferro F, Franchi L, Masucci C, Chiodini P *et al.* Meta-analysis of skeletal mandibular changes during Frankel appliance treatment. *Eur J Orthod.* 2011;33:84–92.
  31. Thiruvengkatachari B, Harrison JE, Worthington HV, O'Brien KD. Orthodontic treatment for prominent upper front teeth (Class II malocclusion) in children. *Cochrane Database Syst Rev.* 2013;(11):CD003452.
  32. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J *et al.* AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol.* 2009;62:1013–1020.
  33. Booth A, Clarke M, Gherzi D, Moher D, Petticrew M, Stewart L. An international registry of systematic review protocols. *Lancet.* 2011;377:108–109.
  34. List T, Axelsson S. Management of TMD: evidence from systematic reviews and meta-analyses. *J Oral Rehabil.* 2010;37:430–451.
  35. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ *et al.* Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials.* 1996;17:1–12.
  36. Higgins JP, Altman DG, Gozscze PC, Jüni P, Moher D, Oxman AD *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomized trials. *BMC Med Res Methodol.* 2011;343:d5928.
  37. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S *et al.* Grading quality of evidence and strength of recommendations. *BMC Med Res Methodol.* 2004;19:1490.
  38. Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M *et al.* Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. *BMC Med Res Methodol.* 2000;26:537–540.
  39. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M *et al.* Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA.* 1998;280:278–280.
  40. Fleming PS, Seehra J, Polychronopoulou A, Fedorowicz Z, Pandis N. A PRISMA assessment of the reporting quality of systematic reviews in orthodontics. *Angle Orthod.* 2013;83:158–163.
  41. Pocock SJ. *Clinical trials – a practical approach.* Chichester: John Wiley & Sons Ltd.; 1996.
  42. Hadorn DC, Baker D, Hodges JS, Hicks N. Rating the quality of evidence for clinical practice guidelines. *J Clin Epidemiol.* 1996;49:749–754.
  43. Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor CD *et al.* Users' guides to the medical literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. Evidence-Based Medicine Working Group. *JAMA.* 2000;284:1290–1296.
  44. Khorsan R, Crawford C. How to assess the external validity and model validity of therapeutic trials: a conceptual approach to systematic review methodology. *Evid Based Complement Alternat Med.* 2014;2014:694804.

45. McNamara JA Jr. Components of class II malocclusion in children 8–10 years of age. *Angle Orthod.* 1981;51:177–202.
46. Antoun JS, Cameron C, Sew Hoy W, Herbison P, Farella M. Evidence of secular trends in a collection of historical craniofacial growth studies. *Eur J Orthod.* 2015;37:60–66.
47. Shen G, Hägg U, Darendeliler M. Skeletal effects of bite jumping therapy on the mandible – removable vs. fixed functional appliances. *Orthod Craniofac Res.* 2005;8: 2–10.
48. Rongo R, Antoun JS, Lim YX, Dias G, Valletta R, Farella M. Three-dimensional evaluation of the relationship between jaw divergence and facial soft tissue dimensions. *Angle Orthod.* 2014;84:788–794.

Correspondence: Vincenzo D'Antò, Department of Neuroscience, Reproductive Sciences and Oral Sciences, School of Orthodontics and Temporomandibular disorders, University of Naples "Federico II", Via Pansini, 5 – 80131 Naples, Italy.

E-mail: vincenzo.danto@unina.it

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Table S1.** Medline (via PubMed) search strategy.

**Table S2.** References excluded after the full-text reading and reason for the exclusion.