ARISTOTLE UNIVERSITY OF THESSALONIKI
SCHOOL OF DENTISTRY
DEPARTMENT OF ORTHODONTICS
Chairman: Assoc. Prof. Moschos A. Papadopoulos

CLINICAL EFFECTIVENESS OF CHIN CUP TREATMENT FOR THE MANAGEMENT OF CLASS III MALOCCLUSION: A META-ANALYSIS

MARIA I. CHATZOUDI, D.D.S.

A Thesis Submitted to the Faculty of the Department of Orthodontics in Partial Fulfilment of the Requirements for a Postgraduate Diploma of Specialization in Orthodontics

Thessaloniki 2013

Supervisors

Moschos A. Papadopoulos
Assoc. Professor, Chairman, and Program Director
Department of Orthodontics
School of Dentistry, Aristotle University of Thessaloniki

Ioulia Ioannidou-Marathiotou
Assoc. Professor
Department of Orthodontics
School of Dentistry, Aristotle University of Thessaloniki

To Vasiliki Amanah Belle Chatzoudis
and to all those people stood by my side during this period.

CONTENTS

A. ABSTRACT

Background/Introduction: Chin cup is probably the oldest of the orthopedic approaches to treat growing patients with Class III malocclusion. The clinical effectiveness of this appliance, even though widely used in the orthodontic practice, has not yet been thoroughly investigated under proper evidence based approaches.

Objectives: Aim of this meta-analysis was to qualitatively and quantitatively synthesize the currently existing data in order to investigate the clinical effectiveness of the chin cup treatment utilized for the orthodontic/orthopedic management of growing patients with Class III malocclusion.

Materials and Methods: Detailed electronic and hand searches were performed in order to identify the potentially relevant studies up to September 2011 with no restrictions. Only randomized controlled trials, prospective controlled trials and retrospective controlled trials assessing lateral cephalometric radiographs of patients with Class III malocclusion treated with the chin cup before or during the pubertal growth spurt (6-14 years old) were included. Study selection and data extraction were performed independently by two review authors and the level of their agreement was assessed by means of the Cohen's kappa statistic. Prior to data synthesis, the risk of bias of the included studies was also assessed independently by two review authors and the level of agreement was again assessed using the Cohen's kappa statistic. The risk of bias was assessed with two different approaches; the Antczack et al. (1986) and Jadad et al. (1996) method, and the Downs and Black (1998) scale. Meta-analysis was performed by calculating the Standard Difference in Means and the corresponding 95% Confidence Intervals (CI) for all cephalometric variables under investigation. The random effects model

was used in order to assess the effects of chin cup treatment. Heterogeneity was assessed by means of the Cochrane *Q* statistic and the corresponding *P*-value, as well as by means of the I^2 (within studies) and T^2 (between studies) indexes. The presence of publication bias was initially evaluated through visual inspection of the funnel plots. In case of asymmetry, publication bias was further assessed statistically using the Begg and Mazumdar's test, and the Egger's test. Exploratory analyses were performed when the included studies were less than five. Finally, subgroup analyses were also performed with regard to the modifiers (factors) "force level" and "additional appliances used" and trends were observed when statistical significance was present. The level of significance was set at $P \leq 0.05$ for all tests except for the publication bias tests, where it was set at $P \leq 0.1$.

Results: Initially, 1238 records were retrieved through electronic search, while 5 more studies were added at a later stage through manual searching of the reference lists of the assessed full text articles. After removal of duplicates and following application of specific inclusion and exclusion criteria, 5 studies were eligible for inclusion in this evaluation. The risk of bias assessment revealed that the overall estimate of the quality of the 5 studies included in the analysis was low. The following 13 cephalometric variables were examined in this meta-analysis, divided into 3 categories: (a) skeletal variables in the sagittal plane: SNA, SNB, ANB, Wits appraisal and Co-Gn; (b) skeletal variables in the vertical plane: SN-ML, Gonial angle, N-Me, UFH, LAFH and Co-Go; and (c) dentoalveolar variables: Overjet and Overbite. Publication bias assessment indicated the existence of publication bias for all the examined variables. Meta-analytic synthesis of the data of the primary studies revealed that the cephalometric variables SNB and Gonial angle decreased significantly, ANB, Wits appraisal and SN-ML in-

creased significantly, whereas SNA remained unchanged. Exploratory analysis showed a significant increase of the variables N-Me and Overjet, while for the rest of the variables no statistically significant differences were detected. All these alterations are indicative of a clockwise, backwards and downwards rotation of the mandible with possible concomitant changes in its shape (gonial angle decreases) and in an increase of the overjet.

Conclusion: According to the results of this meta-analysis it can be concluded that the chin cup when utilized for the treatment of growing patients with increasing Class III malocclusion affects significantly and positively towards the treatment goals a number of skeletal and dentoalveolar cephalometric variables, indicating an overall positive effect.

A.1. KEY WORDS: Class III malocclusion; Class III treatment; Chin cup; Meta-analysis; Systematic review

Β. ΠΕΡΙΛΗΨΗ

Εισαγωγή: Η επιγναθίδα αποτελεί ενδεχομένως τον παλαιότερο τρόπο αντιμετώπισης της συγκλεισιακής ανωμαλίας Τάξης ΙΙΙ σε νεαρούς ασθενείς που βρίσκονται σε αύξηση. Ωστόσο, η κλινική της αποτελεσματικότητα, αν και είναι ευρέως αποδεκτή από τον ορθοδοντικό κόσμο, δεν έχει μέχρι στιγμής εξεταστεί διεξοδικά υπό το πρίσμα της ορθοδοντικής βασισμένη στην τεκμηρίωση. **Σκοπός:** Σκοπός της παρούσας μετα-ανάλυσης ήταν η ποιοτική και ποσοτική σύνθεση και στατιστική επεξεργασία των διαθέσιμων δεδομένων ώστε να διερευνηθεί διεξοδικά η κλινική αποτελεσματικότητα της επιγναθίδας, όπως αυτή χρησιμοποιείται για την αντιμετώπιση νεαρών ασθενών με συγκλεισιακή δυσαρμονία Τάξης ΙΙΙ.

Υλικά και Μέθοδοι: Πραγματοποιήθηκε χωρίς περιορισμούς λεπτομερής ηλεκτρονική και χειροκίνητη διερεύνηση μέχρι τον Σεπτέμβριο του 2011 προκειμένου να εντοπισθούν όλες οι πιθανά συσχετιζόμενες εργασίες. Συμπεριελήφθησαν μόνο τυχαιοποιημένες ελεγχόμενες κλινικές μελέτες, προοπτικές κλινικές μελέτες με δείγμα ελέγχου και αναδρομικές κλινικές μελέτες με δείγμα ελέγχου, οι οποίες αξιολογούσαν πλάγιες κεφαλομετρικές ακτινογραφίες νεαρών ασθενών που παρουσίαζαν Τάξη ΙΙΙ και θεραπεύτηκαν με επιγναθίδα πριν ή κατά τη διάρκεια της εφηβικής τους ανάπτυξης (6-14 ετών). Η συλλογή των εργασιών και η εξαγωγή των δεδομένων έγινε από δύο ερευνητές και το επίπεδο συμφωνίας μεταξύ τους εκτιμήθηκε με το στατιστικό δείκτη Cohen’s kappa. Πριν από τη σύνθεση των δεδομένων, ο κίνδυνος μεροληψίας ελέγχθηκε ανεξάρτητα επίσης από δύο ερευνητές και το επίπεδο συμφωνίας τους εκτιμήθηκε ξανά με τη χρήση του στατιστικού δείκτη Cohen’s kappa. Ο κίνδυνος μεροληψίας εκτιμήθηκε με δύο διαφορετικές προσεγγίσεις, τη μέθοδο Antczack et al. (1986) και Jadad et al.

(1996) και την κλίμακα των Downs και Black (1998). Για τη διενέργεια της μετα-ανάλυσης χρησιμοποιήθηκαν για όλες τις υπό εξέταση κεφαλομετρικές μεταβλητές η σταντarισμένη διαφορά των μέσων όρων (Standard Difference in Means, SDM) και τα αντίστοιχα διαστήματα εμπιστοσύνης (Confidence Intervals, CI) στο 95%. Η εκτίμηση των αποτελεσμάτων έγινε με βάση το μοντέλο τυχαίας επίδρασης (random effects model). Η ετερογένεια των συμπεριληφθέντων εργασιών εκτιμήθηκε με το στατιστικό δείκτη Cochrane Q και τα αντίστοιχα επίπεδα στατιστικής σημαντικότητας (P-value) καθώς και με τους δείκτες διακύμανσης I^2 (εντός των εργασιών) και T^2 (μεταξύ των εργασιών). Η ύπαρξη μεροληψίας του τύπου δημοσίευσης εκτιμήθηκε αρχικά με οπτική διερεύνηση των αντίστοιχων χωνοειδών διαγραμμάτων (funnel plots). Εάν παρατηρούνταν ασυμμετρία, τότε χρησιμοποιούνταν επιπλέον και το τεστ των Begg και Mazumdar καθώς και το τεστ του Egger. Στις περιπτώσεις που οι συμπεριληφθείσες εργασίες ήταν λιγότερες από 5, πραγματοποιήθηκαν διερευνητικές αναλύσεις αντί για την κλασσική μετα-ανάλυση. Τέλος, πραγματοποιήθηκε ανάλυση υποομάδων με τη χρήση των τροποποιητών (παραγόντων) «εφαρμοσθείσα δύναμη» και «χρήση επιπλέον συσκευής» και διερευνήθηκαν οι αντίστοιχες «τάσεις» όπου υπήρχε στατιστικά σημαντική διαφορά. Τα επίπεδα εμπιστοσύνης τέθηκαν στο 0,05 ($P \leq 0{,}05$) για όλες τις μεταβλητές εκτός από την ετερογένεια, όπου τέθηκαν στο 0,1 ($P \leq 0{,}1$).

Αποτελέσματα: Αρχικά μέσω της ηλεκτρονικής αναζήτησης εντοπίσθηκαν 1238 άρθρα, ενώ άλλα 5 προστέθηκαν σε αυτά αργότερα, από τον έλεγχο των αναφορών των εργασιών των οποίων εξετάστηκε το πλήρες κείμενο. Αφού αφαιρέθηκαν οι διπλές εργασίες και κατόπιν εφαρμογής εξειδικευμένων κριτηρίων επιλογής και απόρριψης, 5 εργασίες κρίθηκαν ως κατάλληλες για να συμπεριληφθούν στην παρούσα μελέτη. Η εκτίμηση του κινδύνου μεροληψίας έδειξε ότι η ποιότητα

των 5 αυτών εργασιών ήταν γενικά χαμηλή. Εξετάσθηκαν οι ακόλουθες 13 κεφαλομετρικές μεταβλητές, οι οποίες χωρίσθηκαν σε 3 κατηγορίες: (α) σκελετικές μεταβλητές στο οβελιαίο επίπεδο: SNA, SNB, ANB, Wits appraisal και Co-Gn, (β) σκελετικές μεταβλητές στο κατακόρυφο επίπεδο: SN-ML, Gonial angle, N-Me, UFH, LAFH και Co-Go και τέλος (γ) οδοντοφατνιακές μεταβλητές: Overjet and Overbite. Η μεροληψία τύπου δημοσίευσης ανιχνεύθηκε για όλες τις μεταβλητές. Η μετα-αναλυτική στατιστική επεξεργασία των δεδομένων έδειξε ότι υπήρχε στατιστικά σημαντική μείωση των μεταβλητών SNB και Gonial angle, στατιστικά σημαντική αύξηση τω μεταβλητών, ANB, Wits appraisal και SN-ML, ενώ η μεταβλητή SNA παρέμεινε αμετάβλητη. Η διερευνητική ανάλυση έδειξε σημαντική αύξηση των μεταβλητών N-Me και Overjet ενώ για τις υπόλοιπες μεταβλητές, Co-Gn, Co-Go, UFH, LAFH και Overbite δεν προέκυψαν στατιστικά σημαντικά αποτελέσματα. Οι μεταβολές που βρέθηκαν είναι ενδεικτικές μιας οπίσθιας στροφής της κάτω γνάθου, ενώ ταυτόχρονα φαίνεται να επισυμβαίνει αναδιαμόρφωση του σχήματός της στην περιοχή της γωνίας του Γωνίου, καθώς και αύξηση της οριζόντιας πρόταξης.
Συμπεράσματα: Σύμφωνα με τα αποτελέσματα της παρούσας μετα-ανάλυσης μπορεί να υποστηριχθεί ότι η επιγναθίδα, όταν χρησιμοποιείται για τη θεραπεία νεαρών ασθενών με συγκλεισιακή ανωμαλία Τάξης III που βρίσκονται σε περίοδο αύξησης, επηρεάζει σημαντικά και θετικά ως προς τον θεραπευτικό στόχο αρκετές κεφαλομετρικές μεταβλητές, δίδοντας ένα συνολικά θετικό αποτέλεσμα.

Β.1. ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Συγκλεισιακή ανωμαλία Τάξης III, Θεραπεία Τάξης III, Επιγναθίδα, Μετα-ανάλυση, Συστηματική ανασκόπηση

1. INTRODUCTION

According to Sugawara, most malocclusions and dentofacial deformities do not result from some pathologic process, but rather from a moderate distortion of normal development (Sugawara, 2005). Class III malocclusion appears to develop from a strong interaction between environmental and innate factors (Sugawara, 2005) and has always been a challenge to the orthodontist due to the limited role that orthodontics can play, given a strong and dominant genetic predisposition. Class III malocclusion can be attributed to skeletal, dentoalveolar or functional relationships.

Descriptive cephalometric analyses have shown that the skeletal Class III pattern is mainly attributed to a jaw discrepancy, with the mandible being disproportionally bigger than the maxilla (Proffit and Fields, 2000). Increased mandibular size, anterior mandibular position and maxillary deficiency are usually observed either each one alone or in combinations (Ngan, 2001; Ritucci and Nanda, 1986). When compared with normal Class I patients, craniofacial growth of these patients usually present a reduction in the size of the cranial base and an approximate +1.5 years retardation in the completion of the growth development (Alexander, 2007; Langlade, 1981; Lu et al., 1993; Ngan, 2001; Ritucci and Nanda, 1986).

In the dentoalveolar type of Class III malocclusion, there is no apparent sagittal skeletal discrepancy (Alexander, 2007; Ngan, 2001). Skeletal variables are within normal limits (Alexander, 2007; Ngan, 2001). The problem is primarily caused by lingual tipping of the maxillary incisors and labial tipping of the mandibular incisors (Alexander, 2007; Ngan, 2001).

In functional Class III malocclusion, also known as pseudo Class III, patients with a forward shift of the mandible on closure may have Class I skeletal pattern and Class I molar relationship in centric occlusion (i.e. the most posterior and upper position of the condyles in the glenoid fossa when patient occludes), which transforms into a Class III skeletal and dental pattern in maximum intercuspation (Ngan, 2001).

1.1. Diagnosis of Class III malocclusion

According to Angle, Class III describes the condition where the mesiobuccal cusp of the upper first molar occludes distally to the mesiobuccal groove of the lower first molar (Angle, 1920). It is also known as prenormal occlusion (Thilander, 1965). This can be the manifestation of an underlying skeletal and/or dentoalveolar background (Altuğ, 1989; Langlade, 1981; Ngan et al., 1996) with (Figure 1):

- a normal maxilla and mandibular prognathism (Figure 1A),
- maxillary retrusion and a normal mandible (Figure 1B),
- normal maxilla and mandible (Figure 1C), and
- maxillary retrusion and mandibular prognathism (Figure 1D).

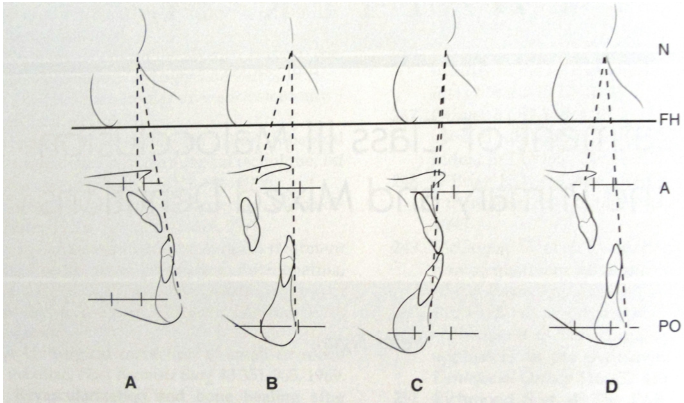

Figure 1. Tracings of four different types of Class III skeletal profiles. (A) Normal maxilla and mandibular prognathism. (B) Maxillary retrusion and normal mandible. (C) Normal maxilla and mandible. (D) Maxillary retrusion and mandibular prognathism. *[Following permission, from: Ngan et al. (1996)]*

The typical skeletal Class III growth pattern is usually presented with (Alexander, 2007; Langlade, 1981; Lu et al., 1993; Proffit and Fields, 2000; Sugawara, 2005):

- reduction of the increase rate of the cranial base,
- increase of the length of the mandibular corpus, and
- growth retardation, which might be completed by the age of 19.5 years in girls and by the age of 21 years in boys.

Langlade (1981), being more specific, described the following cephalometric characteristics (based on the Ricketts analysis) as indicative of the skeletal Class III pattern:

- Increased SN-FH angle (the angle formed between the anterior cranial base and the Frankfurt Horizontal plane).
- Forward position of the ascending ramus of the mandible.
- Forward position of Porion.
- Obtuse facial angle (the angle formed between the Frankfurt Horizontal plane and the Nasion-Pogonion plane).
- Long mandibular corpus.

With regard to the dentoalveolar characteristics in Class III malocclusion, they are often associated with a narrow upper arch and a broad lower arch (Mitchell, 1996; Thilander and Myrberg, 1973). As a result, crowding is present more commonly and to a grater degree in the upper than in the lower arch (Alexander, 2007; Mitchell, 1996; Ngan, 2001; Thilander and Myrberg, 1973). Frequently, the lower arch is characterized by well-aligned or even spaced teeth (Alexander, 2007; Mitchell, 1996; Ngan, 2001; William and Anderson, 1986). Minimal overjet or even anterior crossbites are quite often observed, which are attributed either to the relative discrepancy of the length of the dental arches or to the improper inclination of the maxillary and mandibular incisors or even to occlusal interferences (functional) (Mitchell, 1996; Ngan, 2001). Posterior crossbite is finally another potential characteristic accompanying Class III malocclusion, due to the relative discrepancy in the width of the dental arches (Kerr and Ten Have, 1987; Mitchell, 1996).

According to the deviation of the skeletal characteristics from those describing a normal Class I case, Class III malocclusion can be escalated from mild to moderate and severe (Alexander, 2007; Mitchell, 1996; Langlade, 1981; Ngan et al., 1996). Factors that contribute to this escalation include the length of the mandible and the length of the maxilla relative to each other and to the length of the cranial base, the position of the mandible and the maxilla, the width of the mandible and the maxilla, the amount of crowding observed in the maxilla and spacing or crowding in the mandible (Alexander, 2007; Langlade, 1981; Ngan et al., 1996). Growth pattern in the vertical plane can also affect the severity of Class III patients, which can be horizontal, vertical or average (Alexander, 2007; Langlade, 1981; Ngan et al., 1996; Proffit and Fields, 2000).

More specifically, the longer and wider the corpus of the mandible is, the more anterior its position, the shorter the cranial base, the shorter and narrower the maxilla, the more severe the skeletal Class III pattern is (Alexander, 2007; Langlade, 1981; Ngan et al., 1996; Proffit and Fields, 2000). Additionally, the more vertical the growth pattern is, also the more severe the Class III malocclusion can be regarded (Alexander, 2007; Langlade, 1981; Ngan et al., 1996; Proffit and Fields, 2000). Adversely, in a Class III condition, the smaller and narrower the mandible is, the more posterior its position, the longer the cranial base, the longer and wider the maxilla, the milder this Class III relationship is (Alexander, 2007; Langlade, 1981; Ngan et al., 1996; Proffit and Fields, 2000). Furthermore, the more horizontal the growth pattern is, also the milder the Class III malocclusion can be regarded (Alexander, 2007; Langlade, 1981; Ngan et al., 1996 ; Proffit and Fields, 2000). Horizontal growth patterns respond generally better in treatment modalities than the vertical ones (Alexander, 2007; Langlade, 1981; Ngan et al., 1996; Proffit and Fields, 2000). Moderate skeletal Class III pattern is described by characteristics lying somewhere in between the mild and the severe skeletal Class III conditions (Ngan, 2001).

Dentoalveolar Class III malocclusion, while defined by normal sagittal and vertical skeletal cephalometric variables, it is characterized by dentoalveolar variables approaching the Class III pattern (Ngan, 2001). More analytically, it is mainly attributed to the inclination of the anterior teeth, labial in the maxilla and lingual in the mandible, and the subsequent inclination of the alveolar processes (Alexander, 2007; Mitchell, 1996; Ngan, 2001). Positive overjet or an end-to-end incisal relationship can usually be observed (Alexander, 2007; Mitchell, 1996; Ngan, 2001). Crowding in the maxilla and spacing in the mandible may also co-exist (Alexander, 2007; Mitchell, 1996; Ngan, 2001).

Functional or pseudo-Class III malocclusion is another clinical condition many times met in the orthodontic practice. It is so named the anterior shift of the mandible on closure in order to avoid a premature contact and to achieve maximum intercuspation in a forward position (Proffit and Fields, 2000). The underlying skeletal background is an intermediate form between Class I and Class III malocclusion with the exception of the gonial angle, which is more obtuse in the skeletal Class III pattern and more oblique in the pseudo Class III condition (Ngan, 2001). Therefore, according to Ngan (2001), this measurement is attributed a key diagnostic feature in the differential diagnosis between functional and skeletal Class III relationship. Both prognosis and treatment planning are expected to be much more favorable in the first one (Rakosi, 1997).

1.2. Prevalence of Class III malocclusion

The prevalence of Class III malocclusion varies among different populations. It ranges between 1% and 4% in Caucasians (Ainsworth, 1925; Enrich et al., 1964; Humphreys and Leighton, 1950; Massler and Frankel, 1951; Newman, 1956; Goose et al., 1957; Hill et al., 1959; Thilander and Myrberg, 1973; Magnusson, 1976; Tschill et al., 1997), between 5% and 8% in African Americans (Altemus, 1959; Horowitz and Doyle, 1970; Isiekwe, 1983; Garner and Butt, 1985; Otuyemi and Abidoye, 1993), while in Asian population is much higher, i.e., between 4% and 13% in Japanese (Ishii et al., 1987) and between 4% and 14% in Chinese (Allwright and Burndred, 1964; Lin, 1985).

1.3. Etiology of Class III malocclusion

With regard to the etiology of Class III malocclusion, heredity seems to play a dominant role (Harris et al., 1973), a strong example of which is the Hapsburg family (McGuigan, 1966). Additional etiological factors also involved include the following (Alexander, 2007; Langlade, 1981; Proffit and Fields, 2000; Rakosi, 1997; Rakosi and Schilli, 1981; Sugawara, 2005):

- general factors, such as:
  - hormonal disturbances,
  - syndromes,
  - cleft lip and palate, and
  - nasorespiratory difficulties; and
- regional factors, such as:
  - change in the posture of the mandible,
  - premature loss of 1st molars,
  - disturbance in the normal eruption of incisors,
  - trauma, and
  - habits (such as sucking the upper lip)

The role of the soft tissues in the etiology of Class III malocclusion also needs to be taken into consideration. According to Mitchell (1996), the lips do not comprise a major etiological factor. On the contrary, they tend to tilt both the upper and lower incisors towards each other so that the incisor relationship is less severe than the underlying skeletal pattern (Mitchell, 1996). This is presumably due to an anterior oral seal achieved by upper to lower lip contact, which moulds the upper and lower segment towards each other (Kerr and Ten Have, 1987; Mitchell, 1996; William and Anderson, 1986). However, an exception is observed in patients with increased vertical skeletal proportions where the lips are more likely to be incompetent and an anterior oral seal is often accomplished by tongue to lower lip contact avoiding thus oral breathing (Kerr and Ten Have, 1987; Mitchell, 1996; Proffit and Fields, 2000; William and Anderson, 1986). On the contrary, a flat, anteriorly positioned tongue that lies low in the mouth is considered to be a local epigenetic factor pushing the lower teeth forward, and therefore it should be eliminated during therapy (Rakosi, 1997). Whether tongue posture is a compensatory adaptive phenomenon expressing a result or a primary etiologic factor, is rather controversial (Mitchell, 1996; Rakosi, 1997). According to the authors' opinion, in different patients it could have different roles, approaching either the "adaptive" or the "etiologic" theory. Nasorespiratory difficulties, enlarged tonsils or excessive epipharyngeal lymphoid tissue which are all factors leading to mouth breathing can additionally help to the establishment of a Class III malocclusion (Moyers, 1977; Langlade, 1981; Linder-Aronson, 1970; Rakosi, 1997). Patients with mental retardation presenting compulsive habits like protruding the mandible are also characterized by a high prevalence of mandibular prognathism (Rakosi, 1997).

Other parameters promoting the Class III pattern are occlusal forces created by abnormal eruption and subsequent unfavorable incisal guidance, the premature loss of deciduous molars where the mandible then extends anteriorly in an attempt to establish full occlusal contact during chewing and, finally, overclosure of the mandible caused by loss of posterior occlusal support (Langlade, 1981; Rakosi, 1997).

1.4. Treatment of Class III malocclusion

Treatment of the Class III malocclusion can be achieved with conventional approaches, which includes functional/orthopedic and dentoalveolar modes of treatment or with surgical approaches and is synoptically presented bellow (Alexander, 2007; Chang et al., 2005; Ngan, 2001; Proffit and Fields, 2000; McNamara, 2005; Rakosi, 1997).

Functional/orthopedic treatment attempts maxillary or mandibular growth modification towards the Class I relationship with the aid of extraoral or functional appliances (Alexander, 2007; Chang et al., 2005; McNamara, 2005; Ngan, 2001; Proffit and Fields, 2000; Rakosi, 1997). It promotes changes (a) in the posture of the mandible towards a backward position by remodeling in the temporomandibular joint area involving the glenoid fossa and the condyle, in the ascending ramus as well as in the gonial area or the mandible, (b) in the size of the maxilla by increasing its anteroposterior and sagittal dimensions and (c) in the position of the maxilla which is brought forward and attempted to be forced in a more forward development (Alexander, 2007; Chang et al., 2005; McNamara, 2005; Ngan, 2001; Proffit and Fields, 2000; Rakosi, 1997). Simultaneous enhancement of the downwards development of the maxilla and due to this backwards rotation of the mandible also usually takes place (Alexander, 2007; Chang et al., 2005; McNamara, 2005; Ngan, 2001; Proffit and Fields, 2000; Rakosi, 1997). Prerequisite for the functional/orthopedic treatment is that growth still remains (Alexander, 2007; Chang et al., 2005;

McNamara, 2005; Ngan, 2001; Proffit and Fields, 2000; Rakosi, 1997).

Dentoalveolar Class III orthodontic treatment is usually comprised of fixed appliances aimed to change the position and inclination of the teeth. It is applied on the permanent dentition and does not significantly depend on growth (Alexander, 2007; Sugawara, 2005). When not combined with previous orthopedic treatment or later surgical treatment it is also known as camouflage treatment (Mitchell, 1996; Proffit amd Fields, 2000). Camouflage treatment quite often involves extraction of permanent teeth, usually premolars, to relieve crowding and to permit correction of the inclination of the anterior teeth (Proffit and Fields, 2000).

Surgical treatment of the Class III malocclusion is provided to patients having completed their growth and presenting significant degree of skeletal discrepancy towards the Class III skeletal pattern, in which conventional (non surgical) treatment would fail to provide acceptable esthetical and functional results (Alexander, 2007; Mitchell, 1996). It is, in other words, the severe Class III cases where surgical correction applies. The surgical treatment involves a presurgical and postsurgical orthodontic phase of treatment, the first aiming to prepare the teeth for successful jaw surgery and the second to reduce relapse after surgery as much as possible and to provide the patient with acceptable esthetics and occlusion (Proffit and Fields, 2000).

Treatment of the Class III depends both on the age (or growth status) of the patient and the severity of the malocclusion.

1.4.1. Treatment considerations according to the age of the patients

Patients' age is of paramount importance when dealing with Class III malocclusion (Langlade, 1981; Rakosi, 1997; Sugawara, 2005). The younger the patient is, the greater the possibility for a successful conventional (non surgical) orthodontic/orthopedic treatment is, except for patients presenting severe skeletal Class III disharmony (Langlade, 1981; Rakosi, 1997; Sugawara, 2005). Early orthopedic treatment performed either in the deciduous or in the early mixed dentition usually involves the use of extraoral and/or functional appliances (McNamara, 2005; Ngan, 2001; Proffit and Fields, 2000; Rakosi, 1997; Sugawara, 2005). The treatment objectives of this phase according to Sugawara (2005) are:

- The maintenance of good oral hygiene with the help of dental caries risk tests.
- The correction of possible functional deviations of the mandible and the stabilization of the jaw position.
- The improvement of the three-dimensional jaw deformity as far as possible.
- The correction and control of the deviation of the dental midline.
- The accomplishment of desirable anterior occlusion for establishing anterior guidance in the future.
- The establishment of bilateral posterior support.
- The acquisition of the necessary space for the buccal teeth.
- The normalization and enhance of the orofacial functions.

Early orthopedic intervention in growing Class III malocclusion with conventional appliances can be challenging, since growth predictions used to differentiate children with Class III tendency from the normal Class I and identify specific skeletal morphologic factors are reliable to a percent only 70% to 80% (Williams and Anderson, 1986). Indications for early intervention can be good facial esthetics, mild skeletal disharmony, no familial prognathism, antero-posterior functional shift, convergent facial type and symmetric condylar growth (Turpin, 1981). Contraindications are poor facial esthetics, severe skeletal disharmony, hereditary pattern established in the family, no antero-posterior shift, divergent facial type and asymmetric growth. In these cases, patients are left to have their growth completed until a surgical intervention at later age (Turpin, 1981). Factors that may dramatically affect the outcome in orthopedically treated growing Class III children are the proper selection of the specific appliance for the early intervention, patient co-operation (especially concerning the suggested hours of wear), force magnitude and direction (especially for specific appliances like the facemask and the reverse headgear), duration of treatment and patients' age (Thilander, 1965; Turpin, 1981). It seems as if the best results are generally obtained when treatment starts at approximately 4-6 years of age (Turpin, 1981). Rakosi (1997) even suggests the start of the treatment from the 1st year of age as long as the patient can co-operate. The older the patient is, the less the skeletal and the more the dentoalveolar correction is observed (Thilander, 1965). However, according to the clinical practice guidelines for developing Class III malocclusion of Sugawara (2005), orthopedic treatment can be effective even until the beginning of the pubertal growth spurt, approximately 12 years of age in girls and 14 years of age in boys.

In the adolescent and early post-adolescent period and as long as the Class III malocclusion is characterized by a mild skeletal discrepancy, “camouflage” treatment could be an effective option (Proffit and Fields, 2000; Rakosi, 1997; Sugawara, 2005). The name is attributed to the objective of the treatment, which is to correct the malocclusion while making the underlying skeletal problem less apparent (Proffit and Fields, 2000). The method usually involves extractions of lower teeth to obtain proper molar and incisor relationships and/or the use of Class III elastics (Proffit and Fields, 2000). Camouflage requires that repositioning of the teeth will have a favorable, or at least not detrimental, effect on facial esthetics (Mitchell, 1996; Proffit and Fields, 2000). Therefore, no extractions in the lower arch should be performed in severe Class III cases. If the clinician is in doubt, it is better to wait until the end of growth when surgical procedures can be attempted (Mitchell, 1996; Ngan, 2001; Proffit and Fields, 2000; Rakosi, 1997; Sugawara, 2005).

In adults, where no growth is anticipated any more, moderate and severe Class III cases are to be treated with combined surgical-orthodontic treatment comprised of presurgical orthodontic phase, surgical phase and postsurgical orthodontic phase (Proffit and Fields, 2000; Sugawara, 2005). Presurgical orthodontic treatment is “two jaw independent treatments”, which decompensate and obviously worsens the esthetical problem, allowing, however, for grater skeletal correction at the time of the surgery (Proffit and Fields, 2000). Previous extractions of permanent teeth in the mandible in the premolar or incisor area are detrimental for patients being prepared for surgery since proper decompensation and skeletal correction at the time of the surgery can not be achieved any more (Mitchell, 1996; Ngan, 2001; Proffit and Fields, 2000; Rakosi, 1997; Sugawara, 2005). Adversely, presurgical extractions of premolars in the maxilla are often advantageous (Mitchell, 1996; Ngan, 2001; Proffit and Fields, 2000; Rakosi, 1997; Sugawara, 2005). Postsurgical orthodontic treatment aims at finishing the occlusion while simultaneously prevents significant relapse after surgery (Proffit and Fields, 2000). For mild Class III cases though and especially if the patient is unwilling to undergo the surgical procedure, camouflage treatment could still be an option (Proffit and Fields, 2000). Surgery can be performed either in one jaw only or in both jaws (Proffit and Fields, 2000). For up to 7 mm skeletal discrepancy the one-jaw surgery is usually preferred, while for grater discrepancies and due to the increased postsurgical relapse caused from muscle retraction, the two-jaw surgery is recommended (Proffit and Fields, 2000). The most usual type of one-jaw surgery performed in these patients is bilateral mandibular ramus osteotomy, while when significant deficiency coexists in the mental area an additional genioplasty may also take place (Proffit and Fields, 2000). The incision and the deduction of bone substance occur at the posterior border of the corpus of the mandible (Proffit and Fields, 2000). When two-jaw surgery is deemed appropriate, additional maxillary advancement is also performed (Proffit and Fields, 2000). Le Fort-I, or Le Fort-II surgery are electively chosen depending of the midface deficiency (Proffit and Fields, 2000). Sometimes, even segmental osteotomy is performed, when significant expansion of the maxilla is needed (Proffit and Fields, 2000).

1.4.2. Treatment considerations according to the severity of malocclusion

The severity is the second very important parameter to be taken into consideration when dealing with Class III malocclusion. Many authors consider that severe Class III should be left untreated until the end of growth, at which point a proper surgical procedure can be undertaken, combined with pre- and postsurgical orthodontic treatment (Alexander, 2007; Langlade, 1981; Ngan, 2001; Proffit and Fields, 2000; Sugawara, 2005; Thilander, 1965; Turpin, 1981).

Mild Class III cases though, could be treated conventionally (McNamara, 2005; Proffit and Fields, 2000; Sugawara, 2005). Two phases of conventional treatment can be discerned in these patients: the first, orthopedic one, performed in the deciduous or early mixed dentition with extraoral and/or functional appliances, and the second one, performed during or after pubertal growth spurt, in the permanent dentition, where Class III elastics in combination with fixed appliances and possible extractions in the lower dental arch can provide a competent camouflage treatment (McNamara, 2005; Proffit and Fields, 2000; Sugawara, 2005). From the end of the first phase of treatment to the beginning of the second phase a growth observation period is edged into (McNamara, 2005; Proffit and Fields, 2000; Sugawara, 2005).

While for both severe and mild Class III malocclusions bibliography is quite clear on how they should be treated, protocols of treatment modalities for moderate Class III, especially in growing patients, could be challenging, since the clinician can not always and surely predict the exact severity of the malocclusion in the future (Langlade, 1981; Sugawara, 2005; Williams and Anderson, 1986). Conventional treatment is usually attempted in moderate Class III cases early enough with the optimistic aspiration to avoid surgery (McNamara, 2005; Proffit and Fields, 2000; Rakosi, 1997; Sugawara, 2005; Thilander, 1965; Turpin, 1981). Langlade (1981) described specific cephalometric methods to enable somehow the orthodontist to predict growth potentiality, yet, the predictability has not always been correct. According to him, if during a years' time the cranial base (measured at cephalometric points Nasion and Basion) of a Class III patient increases by 1 mm or less and the facial axis (line vertical to the cranial base passing from cephalometric point Pogonion) at the area of menton increases by 3.5 mm or more this is considered as a severe Class III case which should be treated surgically. In a non surgical Class III case, while the cranial base increases 1 mm or more during a years' time, the facial axis increases only approximately 2.6 mm, which is still bigger than the average Class I patient, yet is smaller than the surgical Class III case. Sugawara (2005) described a protocol following some clinical practice guidelines for developing Class III malocclusion, which enables the clinician to switch between camouflage and surgery in the second phase of treatment, if a Class III patient has already undergone first phase conventional Class III treatment (Figure 2).

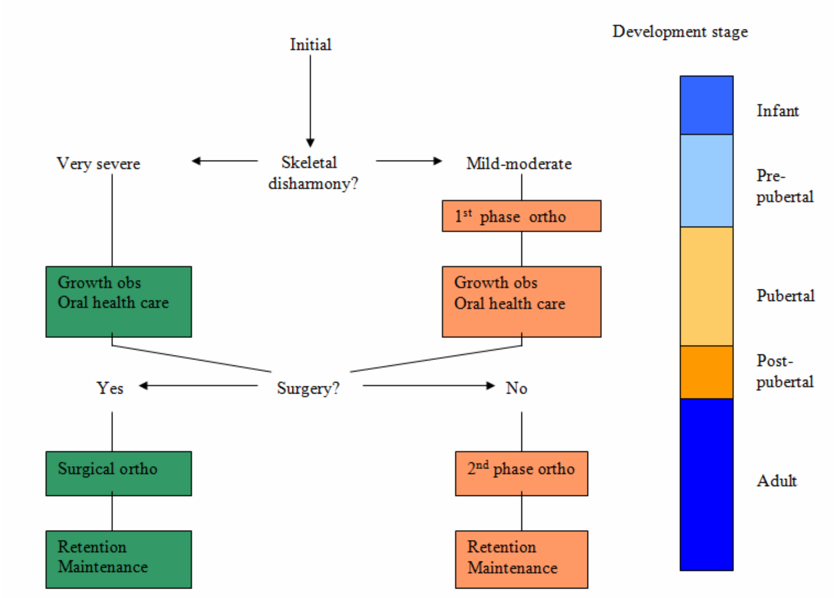

Figure 2. Clinical practice guidelines for developing Class III malocclusion. *[Following permission, from: Sugawara (2005)]*

1.5. Appliances used for the Class III malocclusion

A plethora of appliances are available for the treatment of Class III malocclusion which can be used with judgment depending on patient's age and phase of treatment. The following categories can be discerned:

1) Appliances for the deciduous or mixed dentition.

Functional appliances: the Fränkel III appliance (Fränkel and Fränkel, 1989) and the Class III activator (Rakosi, 1997) used early in age for the first (orthopedic) phase of conventional Class III treatment.

Extraoral appliances: the facemask (Delaire, 1971; Petit, 1982), the chin cup (Ritucci and Nanda, 1986; Sugawara et al., 1997), the Tubingen facial mask (Williams and Bundgaard, 1985) and the reverse headgear (Rakosi, 1997; Proffit and Fields, 2000) also used early in age for the first (orthopedic) phase of conventional Class III treatment.

2) Appliances for the permanent dentition.

Fixed appliances in conjunction with class III elastics (Proffit and Fields, 2000) usually used in the second phase of the conventional treatment.

3) Appliances after surgery.

Fixed appliances in conjunction with class III elastics usually used to prevent relapse after surgery (Proffit and Fields, 2000).

4) Skeletal appliances.

Mini-implants, mini-plates and distraction osteogenesis devices used to enhance the effectiveness of conventional orthodontic treatment in Class III malocclusion and to avoid surgery in borderline cases (Ngan, 2001).

Three traditional orthopedic appliances, the Fränkel-III appliance, the face mask and the chin cup will be discussed bellow, together with their specific indications and contra-indications.

1.5.1. The Fränkel-III appliance

The Fränkel-III, or Functional Regulator type III (FR-3) appliance is an orthopedic appliance used in growing patients with mild maxillary deficiency and/or mandibular forward shift (Fränkel et al., 1989). It is designed to counteract the muscle forces acting on the maxillary complex (Fränkel et al., 1989). The vestibular shields in the depths of the sulcus are placed approximately 2 mm away from the alveolar buccal plates of the maxilla to stretch the periosteum and allow for forward development of the maxilla (Fränkel et al., 1989). These shields are, however, fitted as closely to the alveolar process of the mandible as possible to hold or redirect growth posteriorly (Fränkel et al., 1989). Fränkel (1974) theorizes that the soft tissue matrix, formed by the cheeks, lips and tongue, has an important influence on dental structure development. He supports the opinion that the apical extension of the shield into the vestibule places tension on the buccinators muscle fibbers and dentoalveolar periosteum stimulating bone deposition (Fränkel, 1974). The “periostal tension hypothesis” has also been supported by Harvold (1975) one year later while other authors, much later (Kerr and Ten Have, 1987; Ulgen and Firalti, 1994) became adherent to this theory accepting that the buccal shields produce great passive and adaptive transversal expansion of the dental arches (Harvold, 1975; Fränkel, 1974; Kerr and Ten Have, 1987; Ulgen and Firalti, 1994). Another factor, independent from the “periostal tension hypothesis” which is believed to attribute largely on the correction of Class III malocclusion is the downward and backward rotation of the mandible caused by the FR-3 appliance (Kerr and Ten Have, 1987; Ulgen and Firalti, 1994).

Kalavritinos et al. (2005) on their clinical study found that following treatment with the FR-3 appliance in 14 subjects with mixed dentition for an average of 2.4±0.6 years, three variables concerning vertical relationships of the skeletal profile (N-Me, N-ANS, ANS-Me) were significantly increased, demonstrating a significant increase in skeletal overall anterior, upper and lower facial height. Following evaluation of the cephalometric variables associated with dentoalveolar changes, they observed a significant increase of the overjet and overbite (Kalavritinos et al., 2005). The soft tissues changed most profoundly. A significant increase was also noted in facial convexity (G'–A'/A'– Pg') and in nose prominence (Ns–N'). Regarding the soft tissue vertical relationships, there was an increase in the soft tissue lower facial height (Sn–Me') and in the heights of upper (Sn–St.mean) and lower (St.mean–Me') lips. In addition, thickness of the upper lip was increased (Sn–A, A–A', Sn–St.sup), as well as of the lower lip (Li–Lifac, St.inf–Me'). The decrease observed in the lower lip inclination and depth of the inferior labial sulcus (Li–B'/FH, B'/Pg'–Ls) demonstrated a significant normalization in the lower lip position (Kalavritinos et al., 2005).

The Fränkel III appliance is suggested to be used during the early mixed dentition stage to correct Class III malocclusion characterized by maxillary retrusion (McNamara, 1985), while it is also quite successful in patients with Class III malocclusion presenting a functional shift on closure (Ulgen and Firatli, 1994). It can additionally be applied as a retentive device, following maxillary protraction treatment (Ngan, 2001).

1.5.2. The facemask

The facemask is generally used in patients with maxillary deficiency (McNamara, 2005). It is made of two pads, which contact the soft tissue in the forehead and chin region (McNamara, 2005). A midline framework connects the two pads, which are adjustable through the loosening and tightening of a set-screw (McNamara, 2005). For a downward and forward pull on the maxilla, an adjustable anterior wire with hooks is also connected to the midline framework to accommodate the corresponding elastics (McNamara, 2005). These protraction elastics are attached on the anterior wire with hooks located near the canines so that the opening of the bite is minimized (McNamara, 2005). They exercise a downward and forward pull of 30 degrees to the occlusal plane (McNamara, 2005). According to McNamara (2005), for maxillary protraction 300-600 gr are required per side, depending on patient age, and 20 hours use per day is recommended.

The facemask, also know as Delaire Mask, was originally addressed to growing patients with maxillary deficiency, retrusive maxilla, mild to moderate skeletal Class III pattern and a hypodivergent maxilla (Delaire, 1971; Ngan, 2001; Petit, 1982). Oppenheim (1944) much earlier was the first to suggest that since the growth or the anterior displacement of the mandible could not easily be controlled, moving the maxilla forward in an attempt to counterbalance mandibular protrusion could be an alternative. McNamara (2005) claimed to use facemask in combination with bonded maxillary expansion appliance as the appliance of choice in most of Class III orthodontically treated patients.

Treatment effects when using the facemask include the following (Proffit and Fields, 2000, Macdonald et al., 1999; McNamara, 2005; Tortop et al., 2007; Yüksel et al., 2001):

- Forward displacement of the maxilla.
- Downward and backward rotation of the mandible.
- Decrease in facial depth and in profile concavity.
- Clockwise rotation of the occlusal plane.
- Labial inclination of the upper incisors.
- Lingual inclination of the lower incisors.
- Further eruption of the maxillary and mandibular molars.

The cephalometric values supporting the aforementioned changes following facemask treatment can be summarized as (Macdonald et al., 1999; Yüksel et al., 2000):

- Increase of the SNA, ANB and Mp-SN angles,
- Decrease of SNB angle,
- Increase of ANS-Me, Co-Gn and Go-Me distances,
- Increase of Wits' appraisal,
- Increase of Overjet,
- Decrease of Overbite,
- Increase of U1-SN angle,
- Decrease of L1-Mp angle, and
- Increase of U6-SN and L6-SN distances

(where U1=upper central incisor, L1=lower central incisor, U6=upper first molar and L6=lower first molar).

Ideal patients for treatment with this approach, are pre-adolescents, who would present (a) a retrusive maxilla, (b) normally positioned or retrusive but not protrusive maxillary teeth, and (c) normal or short (but not long) anterior facial vertical dimensions (Proffit and Fields, 2000).

According to Ngan (2001), optimal time to intervene to a Class III patient with a facemask is at the time of initial eruption of the maxillary central incisors. Anterior crossbites at that age can be corrected within 3-4 months given an acceptable patient cooperation (McNamara, 2005; Ngan, 2001). An additional 9-12 months of retention period is suggested before removing the facemask, to avoid quick relapse (McNamara, 2005; Ngan, 2001). In preadolescent patients, in whom the maxillary sutures start to fuse together, the facemask is preferably combined with Rapid Maxillary Expansion (RME) appliances in an effort to break the sutures and "disarticulate" the maxilla (McNamara, 2005; Ngan, 2001). At the time of discontinuation of the facemask, a 4-5 mm overjet should have been achieved to allow for some latter Class III relapse until the start of the second phase of conventional or camouflage treatment in the permanent dentition (McNamara, 2005; Sugawara, 2005). If, however, the relapse extends to a degree that recreates anterior crossbite, the patient should be left untreated until the completion of growth, when combined orthodontic-orthognathic surgery should be attempted (McNamara, 2005; Mitchell, 1996; Ngan, 2001; Proffit and Fields, 2000; Rakosi, 1997; Sugawara, 2005).

1.5.3. The chin cup

Chin cup appliances can be divided in two subcategories: the vertical pull chin cup and the occipital pull chin cup.

<u>*1.5.3.1. The vertical pull chin cup (Figure 3)*</u>

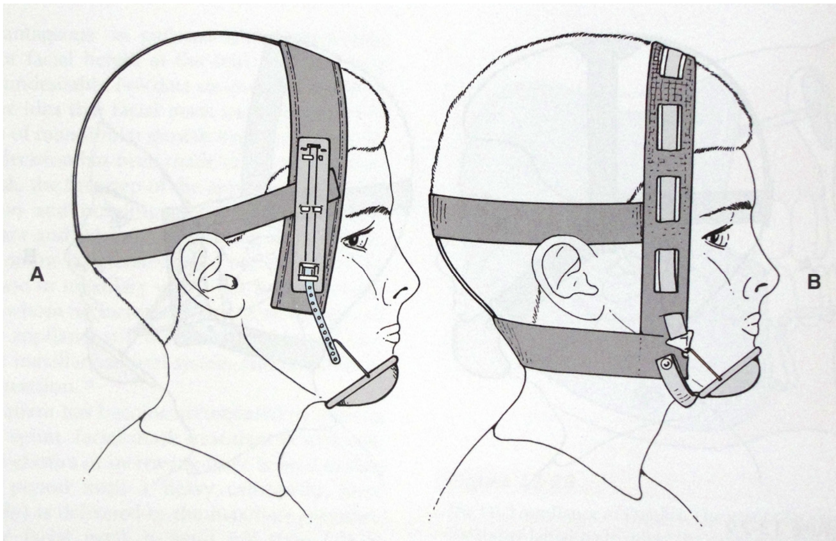

Figure 3. The vertical pull chin cup. **(A)** Uniteck design. **(B)** Summit design. *[Following permission, from: McNamara (2000)]*

The vertical pull chin cup is an orthopedic extraoral appliance designed and used to address problems related with the increased vertical dimension of the face, and more specifically, with increased lower face height, and/or open bite, which can accompany Class III, or Class I malocclusions (Figure 3) (McNamara, 2005). Because of problems encountered in the anchorage from the cranium, the creation of a true vertical force to the chin is difficult (McNamara, 2005). Two designs of vertical chin cup are available in the market, the Unitek design (Figure 3A) and the Summit design (Figure 3B). In the first one, a padded band extends coronally and is secured to the posterior part of the cranium with a cloth strap (McNamara, 2005). A spring mechanism is activated by pulling the tab inferiorly and attaching the pad to a hook adjusted on the hard chin cup (McNamara, 2005). The Summit design incorporates a cloth head cap that curves around the crown of the head and also is secured posteriorly with two horizontal straps. The appliance is additionally secured with a throat strap to the head of the patient (McNamara, 2005). A vertical force of approximately 450-550 gr per side is suggested and the appliance should be used 14 to 16 hours per day (Pedrin et al., 2006).

The vertical pull chin cup seems to pull the chin upwards using as a buttress the anterior part of the cranium (Pedrin et al., 2006). No statistically significant changes have been found in skeletal components (Buschang et al., 2002; McNamara, 2005; Pedrin et al., 2006, Sugawara and Mitani, 1997) with the use of the vertical pull chin cup. Dentoalveolar effects associated with this appliance are the increase of overbite, the correction of the axial inclination of both the upper and lower incisors relatively to the maxillary and mandibular osseous base respectively, the intrusion of the maxillary molars and the extrusion of the mandibular molars (Pearson, 1978; Pearson 1986; Pedrin et al., 2006). When the vertical pull chin cup is used in combination with bonded maxillary expansion appliances, Baccetti et al. (2008) found a slight increase in the posterior facial height and remodeling of the mandible and the temporomandibular joint. These treatment effects are observed immediately after the removal of the appliance, however, there is a tendency for the patient to return to his/her previous vertical growth pattern in the long term (Pedrin et al., 2006; Torres et al., 2006).

In contrast with the other extraoral appliances, where early age plays a dominant role, the vertical pull chin cup should be worn at the circumpubertal period (i.e. shortly before, during and shortly after the pubertal period) as at that time more favorable dentoalveolar treatment effects are observed in comparison with the prepubertal period (Baccetti et al., 2008). Suggested duration of treatment ranges between one and two years depending on the severity of the problem and the patient co-operation (McNamara, 2005). The treatment effects of the vertical pull chin cup can be maintained in the permanent dentition with the use of vertical elastics (McNamara, 2005).

1.5.3.2. The occipital pull chin cup (Figures 4, 5)

The occipital pull chin cup is another orthopedic extraoral device, which is differentiated from the vertical pull chin cup as it addresses problems mainly in the anteroposterior dimension and more specifically with Class III malocclusion (Figures 4, 5) (McNamara, 2005).

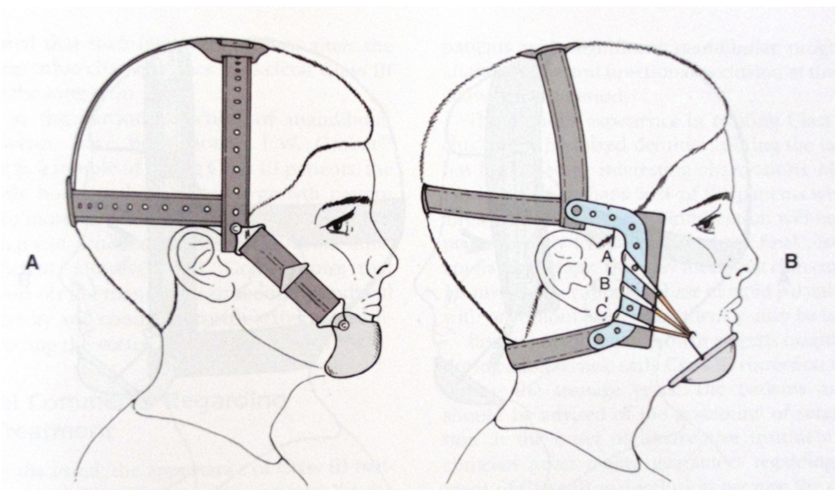

Figure 4. The occipital pull chin cup. **(A)** Soft elastic appliance. **(B)** Hickman type headgear, used as anchorage for a hard chin cup. *[Following permission, from: McNamara (2000)]*

By the use of the occipital pull chin cup and through the application of compressive forces at the condyle, restriction of mandibular growth can be attempted (McNamara, 2005). Two are the principal forms of the occipital pull chin cup: the soft and the hard chin cup (McNamara, 2005). In the first one (Figure 4A), which is a soft elastic appliance, a padded band extends coronally and a soft cup is worn on the chin (McNamara, 2005). Elastic traction bands connect the head cup in the cranium with the cup in the chin (McNamara, 2005). In this version, the position of the head cap determines the direction of the force (McNamara, 2005, Proffit and Fields, 2000). In the second one (Figure 4B), a Hickmans type headgear is used as anchorage for a hard chin cup (McNamara, 2005). It is constructed with padded bands extending coronally and cervically and bearing a plastic force guide in the condylar area, which is used as an anchorage (McNamara, 2005). A hard plastic chin cup is applied on the chin (Langlade, 1981; McNamara, 2005). The direction of the force is adjustable according to the placement of the elastics between the chin cup and the Hickmans type force guide (McNamara, 2005; Williams and Bundgaard, 1985). A combination of the coronal padded band (without the Hickmans type guide) with a hard cup to the chin is used by some clinicians (Ngan, 2001; Sugawara, 2005) and at the orthodontic clinic of the Aristotle University in Thessaloniki in Greece as well (Figure 5). Because the hard cup is not always tolerable by the facial skin, Langlade (1981) suggests creating little holes for better respiration of the skin and covering the inside with suet or castor leather. He also suggests the construction of the hard cup by the orthodontist himself for better fitting of the appliance (Langlade, 1981).

Figure 5. Frontal and lateral view of occipital pull chin cup in a 6 years old patient of the orthodontic clinic of the Aristotle University of Thessaloniki.

There are two main approaches to the occipital pull chin cup therapy according to the direction of force against the mandible (Proffit and Fields, 2000). In the first one (Figure 6), the force is aimed directly at the condylar area to impede mandibular growth, in exactly the same way that the extraoral force against the maxilla impedes its growth (Proffit and Fields, 2000). However, the nature of the temporomandibular joint, the stiffness of the mandibular bone and the difference between maxillary and mandibular growth affect significantly the results, which are not that impressive in the mandible (Proffit and Fields, 2000). In the second approach (Figure 6), the force is aimed bellow the condyle to produce a clockwise rotation of the mandible (Proffit and Fields, 2000). What actually happens is a decrease of the prominence of the chin exchanged with an increase of the anterior facial height (Proffit and Fields, 2000).

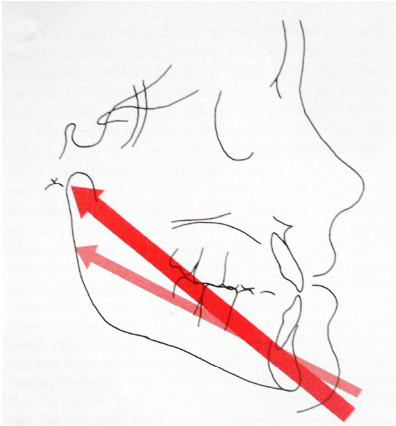

Figure 6. The two main approaches to the occipital pull chin cup therapy. ***(Wide vector)*** Force aimed directly at the condylar area. ***(Narrow vector)*** Force aimed bellow the condyle to produce a clockwise rotation of the mandible. [Following permission, *from: Proffit and Fields (2000)]*

Short-term orthopedic effects of the occipital pull chin cup in Class III may include the following: (a) vertical redirection of the mandibular growth, (b) backward repositioning of the mandible, (c) retardation of mandibular growth and (d) remodeling of the mandible with closure of the gonial angle (Chang et al., 2005; Ngan, 2001; Proffit and Fields, 2000). Alarcón et al. (2011) also supports the concomitant temporomandibular joint remodeling through this process. Langlade (1981), being more optimistic, additionally expects the chin cup application: (e) to retard the vertical development of the posterior part of the maxilla and to exercise a clockwise rotational force to the maxillary plane, (f) to change the growth direction of the condyle from a horizontal to a vertical one by pressing the anterior part of the condyle in the glenoid fossa and by creating free space in its upper and posterior area through the forced clockwise rotation of the mandible, (g) to ameliorate the occlusal relationships of the canines and the molars, (h) not to change at all the axial inclination of the lower incisors and the area of the menton and (i) to reduce the profile concavity. Other authors (Oppenheim, 1944; McNamara, 2005; Sugawara and Mitani, 1997; Thilander, 1965) however, question the ability of the chin cup to restrict mandibular growth both in the ramus and in the corpus of the mandible. It is also observed that the axial inclination of the lower incisors changes and tends to be more labial with the chin cup wear (Abdelnaby and Nassar, 2010; Arman et al., 2004, Barrett et al., 2010; Chang et al., 2005; Wahabuddin, 2005). This phenomenon seems to be more accentuated with the soft rather than the hard chin cup (Arman et al., 2004, Chang et al., 2005).

Several cephalometric variables are found to be affected in the short term by the chin cup use (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Barrett et al., 2010; Gökalp and Kurt, 2005; Tuncer et al., 2009). Reduction of the SNB and the gonial angle, increase in the anterior facial height (N-Me), slight reduction of the Co-Gn and Co-Go mandibular dimensions, increase of the Wits appraisal, reduction of the interincisal angle, increase of the overjet and slight reduction of the overbite are some of them observed by different authors and in different extend (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Barrett et al., 2010; Gökalp and Kurt, 2005; Tuncer et al., 2009).

The long-term effects of the occipital pull chin cup are more rarely investigated. What is generally observed is a tendency of relapse of the clinically promising short-term effects (Sugawara and Mitani, 1997; Deguchi et al., 2002). In other patients, the improvement continues to occur but in a much slower degree than in the short-term (Abu Alhaija and Richardson, 1999; Sugawara and Mitani, 1997; Deguchi et al., 2002).

The relapse observed depends largely on the genetic predetermination of the individual (Alexander, 2007), but also on the facial type (short, long, average), the muscular tone and the tongue habits (Deguchi et al., 1999). More specifically, due to the further increase in the anterior lower face height following chin cup use, dolichocephalic patients are not always quite an indication for this appliance, as they may end up with lip incompetence or even anterior open bite (Gökalp and Kurt, 2005; McNamara, 2005;

Proffit and Fields, 2000). Chin cup application seems to present better esthetic results in horizontal growth patterns than in vertical ones (Alexander, 2007; Langlade, 1981; Ngan et al., 1996; Proffit and Fields, 2000). Indicated may be the conservative correction of their malocclusion in brachycephalic patients, relapse tends to be grater than in dolichocephalic ones due to increased muscular tone (Deguchi et al., 1999). Therefore, it is often expected vertical growth patterns not to be indicated for chin cup use whereas horizontal growth patterns, although indicated, to present grate resistance in the attempted correction (Alexander, 2007; Deguchi et al., 1999; Langlade, 1981; Ngan et al., 1996; Proffit and Fields, 2000). Mesocephalic patients are somewhere in the middle between the previous two categories (Deguchi et al., 1999).

The occipital pull chin cup is generally used for mild to moderate Class III problems where a relatively normal maxilla and a moderately protrusive mandible coexist (Ngan, 2001; Proffit and Fields, 2000; Wahabuddin, 2005). Best age of use for maximum orthopedic effects is in the primary or early mixed dentition, 4 to 9 years of age, whereas in older patients the dentoalveolar effects are more dominant in comparison with the skeletal ones (McNamara, 2005; Ngan, 2001; Sugawara, 2005; Thilander, 1965; Wahabuddin, 2005). It can also be used for the retention period following facemask treatment (McNamara, 2005). Contra-indications for its use are vertical growth pattern with increased lower anterior facial height, Class III problem due to maxillary retrusion and problems in the temporomandibular joints (Gökalp and Kurt, 2005; McNamara, 2005; Proffit and Fields, 2000). Severe Class III cases could also benefit in the 1st phase of their treatment from the chin cup, but generally it is preferred to be left untreated until the completion of their growth, when the skeletal problem can be corrected in a surgical way (McNamara, 2005; Ngan, 2001; Proffit and Fields, 2000; Rakosi and Nanda, 1986; Sugawara, 2005).

A thorough and in depth investigation of the literature data on vertical and occipital chin cup, however, reveals controversies and contradictions regarding both its appropriate way of use and its clinical effectiveness. These are extensively and analytically presented on the "research hypothesis" part of the study (please see pg 59) leading to a justification for the conduction of this meta-analysis.

1.6. Evidence Based medicine/orthodontics

The incorporation of Evidence-Based Orthodontics into clinical care requires a basic understanding of the main research designs underlying the published evidence (Forrest and Miller, 2008; Straus et al., 2005). Inherent characteristics of research designs classify their level of evidence, which may vary from weak to strong (Forrest and Miller, 2008; Straus et al., 2005). This hierarchy is often graphically depicted as a pyramid (Figure 7) (Forrest and Miller, 2008; Straus et al., 2005). The pyramid represents the current bibliographic data both qualitatively, depicting the research design by levels, and quantitatively showing the quantity of each study design in the body of published literature (Forrest and Miller, 2008; Straus et al., 2005). Systematic reviews and meta-analyses (higher quality) are the most difficult, demanding and time-intensive articles to synthesize, being therefore more infrequent than other types of studies (Forrest and Miller, 2008; Straus et al., 2005), but providing the higher level of evidence (Papadopoulos, 2007). While quality of evidence increases as one ascends the levels of the pyramid, databases search massively and indiscreetly reveal all data available on a topic from case-controls up to meta-analytic data (Forrest and Miller, 2008). Problems erupt not just from the massive volume of literature data available at once, which is difficult to be handled, but also from the fact that in the vast majority of it [expert opinions, case reports, case series, case controlled studies, prospective or retrospective control clinical trials, randomized controlled trials (RCTs)] the information provided is not filtered (Forrest and Miller, 2008). Filtered information is provided only by critically appraised individual articles or topics and mainly by systematic reviews and meta-analyses which form part of the formers (Forrest and Miller, 2008).

Raw data for the conduction of systematic reviews and meta-analysis can be derived from randomized clinical trials (RCTs) and from prospective or sometimes even from retrospective controlled clinical trials (pCCTs & rCCTs) (Moher et a.l, 1999; Moher et al., 2010). The CCTs are also known as cohort studies and as depicted in Figure 7 they are lower in quality than the RCTs (Forrest and Miller, 2008). Therefore, the escalation of the quality and validity of the raw data for performing a systematic review or a meta-analysis with descending order is: RCTs, pCCTs, and only under specific conditions rCCTs (Forrest and Miller, 2008; Moher et al., 1999; Moher et al., 2010).

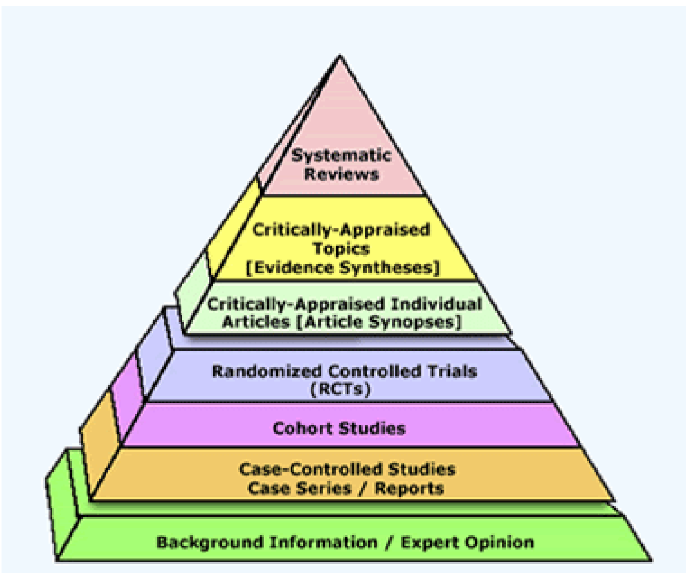

Figure 7. Pyramid depicting the hierarchy of levels of evidence. *[Following permission, from: http://www.ebmpyramid.org/images/pyramid.gif (2006)]*

1.6.1. Randomized control clinical trials (RCTs)

Randomized controlled clinical trials (RCTs) or randomized comparative trials are study designs of original research aimed to measure in the most unbiased way the efficacy of an intervention within a patient population (Gelbach, 1992; Moher et al., 2010; Porta, 2008). The main characteristic of the RCTs, which distinguishes them from cohort studies, is that following recruitment of the appropriate subjects but before the intervention begins the subjects are randomly allocated to two or more groups, one of them being a control group (Moher et al., 2010). After randomization, the two (or more) groups of subjects are followed in exactly the same way and the only differences between the care they receive should be those related to the intervention factor being compared (Moher et al., 2010).

Up to date, RCTs are believed to form the golden standard of original studies (Papadopoulos, 2003). Although recruitment of “suitable” subjects for the study introduces some kind of selection bias by itself (Gelbach, 1992), the random allocation is generally believed to reduce bias in comparison with cohort studies, ensuring a balance between known and unknown confounding factors and thus being higher in the pyramid of the levels of evidence (Gelbach, 1992; Moher et al., 1999; Moher et al., 2010).

RCTs are not always easy to be performed, as they are both cost expensive and time consuming, therefore the compromise of performing a prospective or retrospective controlled clinical trial instead is often accepted (Moher et al., 2010).

1.6.2. Prospective controlled clinical trials (pCCTs)

Prospective controlled clinical trials (pCCTs) or prospective cohort studies are also original studies aiming to investigate an intervention factor determined at present while its outcome lies in the future but the allocation of subjects to the treated or the control group is not random (Gelbach, 1992; Porta, 2008). To do so, like the RCTs, they are ideally characterized by consecutive selection of patients and the control sample is by definition matching (Gelbach, 1992). Therefore, pCCTs begin in the present and march forward in a way very similar to the RCTs but the methodology between the two differs in that the non-randomized selection and allocation of the sample subjects introduces high levels of bias for not balancing known and unknown factors that may interfere in the future and thus being lower than the RCTs in the quality of evidence provided (Forrest and Miller, 2008; Gelbach, 1992; Porta, 2008). Prospective CCTs are however considered to be higher in quality than retrospective studies, where bias is believed to be even stronger, and can be used to "feed" a SR or a MA (Forrest and Miller, 2008; Gelbach, 1992; Porta, 2008).

1.6.3. Retrospective controlled clinical trials (rCCTs)

Retrospective controlled clinical trials (rCCTs) or retrospective cohort studies are studies in which a very particular and specific outcome is decided to be examined retrospectively due to a precise intervention that also happened in the past and the relevant records of the collected data form a control and a treated sample (Gelbach, 1992; Porta, 2008). These two ideally are alike in many ways but differ by certain characteristics related to this particular outcome (matching samples) (Gelbach, 1992; Porta, 2008). Retrospective CCTs have the same methodology with pCCTs but different time focus (Gelbach, 1992). They both begin in the present but in the first ones data collection is performed from past records and does not follow patients up, as is the case with a prospective study (Gelbach, 1992; Porta, 2008).

Being based upon past records, rCCTs are more susceptible to both selection and misclassification or information bias from the previously described study designs (RCTs and pCCTs) (Dickersin, 1990; Gelbach, 1992). Also calibrated and accurate measurements and comparisons between the exposed and non-exposed subsamples are difficult to be performed since investigators need to rely on others for this (Dickersin, 1990; Gelbach, 1992). For all these reasons, rCCTs are considered lower than pCCTs concerning the quality of evidence provided (Forrest and Miller, 2008). If, however, they are properly designed with consecutive selection of patients and matching samples they can constitute a possible source of raw data for a SR or a MA since their information can still be regarded as valuable and reliable (Forrest and Miller, 2008).

There are also inherent characteristics in rCCTs that makes them useful for elaborating data and deriving outcomes, especially where prospective forms of study designs are difficult and unrealistic to be performed (Gelbach, 1992; Porta, 2008). They are less time consuming, they can better analyze multiple a priori determined outcomes and they can better address diseases of low incidence whereas proper data collection in a prospective study would be extremely difficult (Gelbach, 1992; Porta, 2008). They are also less expensive than pCCT while providing the same methodology (Gelbach, 1992; Porta, 2008).

1.6.4. Systematic reviews (SRs)

Systematic reviews are summaries of the medical literature that use strict and explicit methods to perform a comprehensive literature search and critical appraisal of individual studies (Gehlbach, 1993; Straus et al., 2005). In systematic reviews an attempt is made by the authors to find all literature data available on a topic (Gehlbach, 1993; Straus et al., 2005). Systematic reviews, being different than review articles, aim to answer a specific clinical foreground question providing filtered information, while, review articles provide an overview on a topic to answer background questions with more or less unfiltered information (Gehlbach, 1993; Straus et al., 2005). Other differences between the two is that for review articles the literature search does not attempt to find all existing knowledge on a topic and is not submitted to strict quality evaluation and inclusion-exclusion criteria applied both in the selection and in the assessing stage (Gehlbach, 1993; Oxford Centre of Evidence Based Medicine). Review articles present rather the trend of an era while systematic reviews focus on the little verified knowledge remained from peer-reviewed publications on a very specific area after removal of all unnecessary or questionable raw data (Gehlbach, 1993; Oxford Centre of Evidence Based Medicine; Straus et al., 2005). Systematic reviews also differs form meta-analyses because while having both similar qualitative analysis only the last ones (meta-analyses) include a quantitative summary of the results (Oxford Centre of Evidence Based Medicine).

1.6.5. Meta-analyses (MAs)

Meta-analyses, also sometimes misleadingly named "statistical overviews" or "systematic reviews" (Papadopoulos, 2003) are important components of systematic reviews that attempt to combine and summarize not only qualitative, but also quantitative data from multiple studies using sophisticated statistical methodology (Greenhalgh and Taylor, 1997; Oxford Centre of Evidence Based Medicine). Such a strategy strengthens evidence since, when combinable data can be derived, it enlarges significantly the sample size (Greenhalgh and Taylor, 1997), giving the results more statistical power and, therefore, more credibility than the individual studies (Greenhalgh and Taylor, 1997).

The term "meta-analysis" was firstly introduced by Glass (Glass, 1976). He actually defined three different levels of analysis: (a) the primary analysis, where the researcher collects and statistically elaborates original data by himself, (b) the secondary analysis, where the statistical elaboration is performed by a different person from the researcher who collects the original data and, (c) the meta-analysis, where the results of other, "combinable", studies are elaborated by an independent investigator (Glass, 1976; Huque, 1988). Besides, the term meta-analysis derives from the Greek words "μετά" and "ανάλυση" which mean analyze afterwards.

Meta-analyses are even more restrictive than systematic reviews, as only compatible data may be combined into a larger data set (Gehlbach, 1993; Greenhalgh and Taylor, 1997). Similar measures for the comparable data are systematically applied and a combined effect size is derived when possible (Dickersin and Berlin, 1992). Meta-analysts have been criticized in the past for not elaborating combinable data and mixing “apples” with “oranges” (Jones, 1992). When properly designed and performed though, a meta-analysis codes “apples” as “apples” and “oranges” as “oranges” (Green and Hall, 1986).

Due to their inherent characteristics and the very nature of the methodology they require, meta-analyses are rather susceptible to present biased results (Gehlbach, 1993; Greenhalgh and Taylor, 1997). In fact, it is difficult for a meta-analysis not to be biased (Gehlbach, 1993; Greenhalgh and Taylor, 1997). Bias can invade during the screening or the selection procedure, when original sample sizes are small or when the number of included studies is limited, when statistical processing is performed based on previous proper or not proper raw data or even when unexpected or uninspected confounding factors erupt (Antczack et al., 1986; Gehlbach, 1993; Jadad et al., 1996). Publication bias is present quite often due to language restrictions (not all languages are as easily included as English) or even due to the fact that original articles reporting statistically significant results are more easily publicized than the ones without statistically significant outcomes (Dickersin, 1990; Egger et al., 1997c). The first ones are also more likely to be cited by other authors (Dickersin, 1990; Egger et al., 1997c).

Despite its possible sources of bias, meta analysis still is considered nowadays the gold standard of evidence and is indicated in multiple and demanding circumstances (Greenhalgh and Taylor, 1997). Indications to perform meta-analysis have erupted not just from the need to filter the almost unlimited literature “debris” but also to gain time, knowledge and wisdom in clinical practice. More specifically, Victor (1995) recommends its use when: (a) there is a necessity of an urgent decision and if the performance of a new trial is impossible, (b) there are reported side effects of drugs and other therapies and generally on evaluation of safety aspects (c) when literature upon a theme is based on non-conclusive studies with small sample sizes and execution of new original clinical trials are unrealistic in view of needed sample size and time and, finally (d) when literature data is confusing and contradictory upon a subject (Victor, 1995). Straus et al. (2005) have even evolved the usefulness of meta-analysis by incorporating its results in systems. Systems are automated evidence based data banks where one reports the specific clinical symptoms of a patient and the system, by correlating the data, automatically makes diagnosis and proposes treatment (Straus et al., 2005). Gehlbach (1993) however seeks attention to the fact that statistically significant results are not equivalent with clinically significant ones, and vice versa. A result might be statistically significant but so small in magnitude that becomes clinically insignificant (Gehlbach, 1993). On the other hand, clinically significant results might be missed because statistical restrictions such as the sample size deprive the researchers from the ability to reject a null hypothesis (Gehlbach, 1993).

According to Papadopoulos (2003), it is of paramount importance to create a precise and accurate protocol to work with, before conducting a meta-analytic study. An acceptable protocol should aim in advance: (1) to define the response variables, (2) to describe analytically the methods of literature searching for the individual studies to be included in the analysis, (3) to present the measures taken to identify and reduce publication bias, (4) to present and justify the inclusion and exclusion criteria for the studies to be included in the analysis (selection bias), (5) to explain the data extraction procedure and (6) to be characterized by adequate statistical analysis (Papadopoulos, 2003).

2. RESEARCH HYPOTHESIS

Chin cup is probably the oldest of the orthopedic approaches to treat growing patients with Class III malocclusion (McNamara, 2005) and is still considered a basic tool when dealing with Class III patients (Alexander, 2007). Yet, the clinical effectiveness of the chin cup has been a subject of questions and controversies in the literature.

The suggested appropriate age for use varies from 4 years of age (Turpin, 1981) up to 14 years of age (Sugawara, 2005). Patients' sex could be a parameter to consider since females mature earlier than males. Regarding the force magnitude, it is generally believed that the younger the patient the less the force magnitude should be (McNamara, 2005; Ngan, 2001). In addition, it should be increased gradually, but the suggested force at the centre of the chin cup ranges in the literature from 150 gr (Barrett et al., 2010) to 1200 gr (Abdelnaby and Nassar, 2010). Suggested hours of wear also range between 8 hours per day (Barrett et al., 2010) and 18 hours per day (Gökalp and Kurt, 2005). Treatment duration varies between 0.8 years and 4 years, depending on the severity of the malocclusion and patients' age (Ngan, 2001; Tuncer et al., 2009). Concomitant use of additional appliances like bite planes and maxillary protraction appliances could significantly affect the results (Abdelnaby and Nassar, 2010; Deguchi and McNamara, 1999; Tuncer et al., 2009).

Clinical results achieved with the chin cup could also be a field of confusion. Retardation or even sometimes restriction of mandibular growth is supported by some authors (Chang et al., 2005; Ngan, 2001; Proffit and Fields, 2000), while others seriously question the ability of the chin cup to restrict mandibular growth both in the ramus and in the corpus of the mandible (Oppenheim, 1944; McNamara, 2005; Sugawara and Mitani, 1997; Thilander, 1965). Change in the axial inclination of the frontal teeth due to chin cup wear is rejected by Langlade (1981) but it is supported by other authors (Abdelnaby and Nassar, 2010; Arman et al., 2004, Barrett et al., 2010; Chang et al., 2005; Wahabuddin, 2005).

Since no standard protocol has been followed from the various clinicians concerning the chin cup use, it is evident that the effectiveness of the appliance varies according to the exact and individualized way of use and it ranges between minimal effectiveness (Ngan, 2001) and grate effectiveness (Altuğ, 1989; Wahabuddin, 2005). Although it has been investigated for many years now, yet, there is little evidence concerning its clinical effectiveness under the scope of the evidence-based science, in form of data derived from systematic reviews (SRs) or meta-analyses (MAs). According to Victor (1995), this kind of approach is recommended, since existing literature data is both contradictory and confusing and since clinical benefit could be derived. The null hypothesis, therefore, investigated in this study and submitted to potential rejection is that the chin cup has no clinical effectiveness on Class III malocclusion.

On January 2011, shortly after the present study was initiated, a systematic review was published investigating the efficacy of the chin cup therapy on Class III (Liu et al., 2011). The systematic review of Liu et al. (2011) mainly examined the efficacy of the chin cup in combination with maxillary protraction appliances and as discussed later (please see pages 169-171) the inclusion and exclusion criteria differ between the present study and that one. Besides, the clinical effectiveness of the solo use of the chin cup can not be evaluated, being always confused with the additional effects of the maxillary protraction appliances. Furthermore, weaknesses observed in that study, such as including primary studies with not matching control sample or using the polled estimate Mean Difference instead of the Standard Difference in Means in lineal measurements with possibly varying magnification factors call for a more careful and properly designed approach of the subject.

3. AIMS

The purpose, therefore, of this meta-analysis was to investigate in an evidence based manner data derived from individual randomized clinical trials (RCTs), prospective control clinical trials (pCCTs) and retrospective control clinical trials (rCCTs) reporting short or long term treatment effects of the chin cup utilized for the treatment of growing patents with Class III malocclusion measured on lateral cephalometric radiographs, dental casts or reported with the aid of electromyography and to compare them with the changes observed in matching untreated individuals with Class III malocclusion.

The main question (PICO question) this meta analysis aims to answer is: for growing patients presenting Class III malocclusion, could chin cup, as compared with no treatment at all, be beneficial for the improvement of their facial, skeletal and dentoalveolar characteristics?

4. MATERIALS AND METHODS

The present meta-analysis is based on the guidelines provided by the PRISMA Statement for reporting systematic reviews and meta-analyses of studies evaluating health care interventions (Liberati et al., 2009) and the Cochrane Handbook for Systematic Reviews of Interventions version 5.1.0 (Higgins and Green, 2011).

4.1. Data sources and searches

Systematic searches were conducted for published, unpublished and ongoing studies up to September 2011 to identify potentially relevant studies reporting data from growing patients with Class III malocclusion having received treatment with chin cup appliance (occipital or vertical) for the improvement of their facial, skeletal and dentoalveolar characteristics. Every effort to minimize any possible bias in the location of studies was made and citations to potentially relevant studies from journal articles, dissertations or conference proceedings were located by searching the corresponding electronic databases.

In addition to the electronic searches, manual searching was also performed for the journals: "American Journal of Orthodontics and Dentofacial Orthopedics", "Angle Orthodontist", "European Journal of Orthodontist", "Journal of Orofacial Orthopedics" and "Orthopedics and Craniofacial Search", as well as in the reference list of the full-text articles eligible for inclusion, in an effort to identify and retrieve all possible relevant data. Existing systematic reviews and meta-analyses relevant to this study were identified and their reference lists were also scanned for additional trials. Conference abstracts were additionally searched and inquired on their current status. Articles published in journals, dissertations, and conference proceedings were located from several electronic databases following use of an appropriately adjusted search strategy for each individual database (Table I).

No restrictions were applied concerning publication year, language or status. "Grey literature" was also included in our search. If additional information was needed, authors were to be contacted. Translation was arranged for languages not commonly used in the country of the author of this study (Turkish, Polish, Chinese, Japanese e.c.t.).

Table I. The electronic databases searched, the search strategies used and the corresponding results.

Electronic database	Search strategy used	Extend of search	Hits
MEDLINE searched via PubMed (1950 – 19.09.2011) www.ncbi.nlm.nih.gov/pubmed/	"chin cup"[All Fields] OR "chin cups"[All Fields] OR "chin cap"[All Fields] OR "chin caps"[All Fields] OR "chincup"[All Fields] OR "chincups"[All Fields] OR "chincap"[All Fields] OR "chincaps"[All Fields]	In all fields	167
EMBASE searched via ScienceDirect (0.-09.1945 – 19.09.2011) www.embase.com	'chin'/exp AND 'cup' OR ('chin'/exp AND 'cups') OR ('chin'/exp AND 'cap') OR ('chin'/exp AND 'caps') OR 'chincup' OR 'chincups' OR 'chincap' OR 'chincaps' AND [humans]/lim* AND [1-9-1945]/sd NOT [20-9-2011]/sd *limitation=humans	In all fields	71
Cochrane Database of Systematic Reviews searched via The Cochrane Library on 19.09.2011 www.thecochranelibrary.com	("chin cup" OR "chin cups" OR "chin cap" OR "chin caps" OR “chincup” OR “chincups” OR “chincap” OR “chincaps”)	In all fields	6
Cochrane Central Register of Controlled Trials searched via The Cochrane Library on 19.09.2011 www.thecochranelibrary.com	("chin cup" OR "chin cups" OR "chin cap" OR "chin caps" OR “chincup” OR “chincups” OR “chincap” OR “chincaps”)	In all fields	13
Google Scholar Beta searched on 19.09.2011 www.scholar.google.com	“chin cup” “chin cups” “chin cap” “chin caps” “chincup” “chincups” “chincap”	All in title	73 2 82 4 33 0 30

	“chincaps”		0
Web of Science searched on 19.09.2011 http://scientific.thomson.com/products/wos/	“chin cup”	In all fields	0
	“chin cups”		0
	“chin cap”		0
	“chin caps”		0
	“chincup”		0
	“chincups”		0
	“chincap”		0
	“chincaps”		0
	Filtered by healthcare		
Evidence-Based Medicine searched on 19.09.2011 http://ebm.bmjjournals.com	"chin cup" OR "chin cups" OR "chin cap" OR "chin caps" OR “chincup” OR “chincups” OR “chincap” OR “chincaps”	In all fields	0
Scopus searched on 19.09.2011 www.scopus.com	("chin cup" OR "chin cups" OR "chin cap" OR "chin caps" OR “chincup” OR “chincups” OR “chincap” OR “chincaps”)	In all fields	187
LILACS database searched on 19.09.2011 http://bases.bvs.br	Abstract=(chin and cup) or (chin and cups) or (chin and cap) or (chin and caps) or (chincup) or (chincups) or (chincap) or (chincaps) AND Group=Humans* *Limitation=humans	In abstract	3
Bibliografia Brasileira de Odontologia searched on 19.09.2011 http://pesquisa.bvs.br/brasil/index.php	(chin and cup) or (chin and cups) or (chin and cap) or (chin and caps) or (chincup) or (chincups) or (chincap) or (chincaps)	In abstract	1
Ovid database searched via Heal-link on 19.09.2011 http://ovidsp.ovid.com/autologin.html	chin-cup or chin-cups or chin-cap or chin-caps or chincup or chincups or chincap or chincaps	In all fields	0
Bandolier searched on 19.09.2011	("chin cup", "chin cups", "chin cap", "chin caps", “chincup”, “chincups”, “chincap”, “chincaps”)	In all fields	0

http://www.medicine.ox.ac.uk/bandolier			
Atypon Link searched on 19.09.2011 www.atypon-link.com	((chin cup) OR (chin cups) OR (chin cap) OR (chin caps) OR chincup OR chincups OR chincap OR chincaps)	In title, keywords, abstract	0
African Journals Online searched on 19.09.2011 www.ajol.info	“chin cup” OR “chin cups” OR “chin cap” OR “chin caps” OR “chincup” OR “chincups” OR “chincap” OR “chincaps”	In all fields	0
Digital Dissertations searched via UMI ProQuest on 19.09.2011 http://proquest.umi.com/pqdweb?RQT = 302&cfc = 1	("chin cup" or "chin cups" or "chin cap" or "chin caps" or “chincup” or “chincups” or “chincap” or “chincaps”)	In all fields (Databases: Dissertations & Theses)	29
Conference Paper Index searched via Cambridge Scientific Abstracts (1982 – 19.09.2011) http://journals.cambridge.org/action/search	(chin cup) OR (chin cups) OR (chin cap) OR (chin caps) OR chincup OR chincups OR chincap OR chincaps	In all fields	0
metaRegister of Controlled Trials (all registers active and archived) searched on 19.09.2011 www.controlled-trials.com/mrct/	(chin cup) OR (chin cups) OR (chin cap) OR (chin caps) OR chincup OR chincups OR chincap OR chincaps	In all fields	1
German National Library of Medicine (ZB MED) searched on 19.09.2010 http://www.medpilot.de	"chin cup" OR "chin cups" OR "chin cap" OR "chin caps" OR “chincup” OR “chincups” OR “chincap” OR “chincaps”	In all fields	168
Sum			**870**

4.2. Selection of studies

Studies eligible to be included in this MA were (a) randomized controlled trials (RCTs), (b) prospective controlled clinical trials, (pCCTs) and (c) retrospective controlled clinical trials with matching control sample (rCCTs) investigating the clinical effectiveness of the chin cup used for orthodontic/orthopedic treatment alone or in combination with removable dissocclusion or expansion appliances (such as maxillary bite planes, mandibular bite planes, removable palatal cribs or Quadhelises) in patients being in the pubertal or prepubertal growth spurt (6-14 years old) at the start of their treatment. Studies were expected to provide measurements of skeletal, dental or soft tissue profile and muscular changes as recorded by means of lateral cephalometric analyses or dental cast analyses or even electromyographic analyses before and after chin cup treatment in the short- and/or long-term. All other clinical or non-clinical study designs were excluded and any investigation not fulfilling the aforementioned criteria in reporting data was also excluded from further evaluation. Duplicate records, such as published articles presented also in conferences, studies with multiple publications, as well as dissertations published also as journal articles, were excluded. Trials referring to human studies investigating patients being by the end or after the pubertal growth spurt (15+ years old) at the start of their treatment were excluded. Also trials with inadequate sample size groups (sample size <10 or not reported at all) were excluded. Finally, studies with no matching control sample were excluded as well. Table II presents in detail the eligibility criteria used in this meta-analysis.

If the reviewers could not decide on the eligibility of a study by examining the title and the abstract, the full text of the article was retrieved. For the remaining articles, the corresponding full texts were retrieved for further evaluation.

All the above-mentioned processes of data selection were not performed blinded, because scientific evidence does not strongly recommend masked assessment (Moher et al., 1999). Non-blinded processes where also conducted for data extraction and quality analysis.

Screening of titles, abstracts and full-text citations was performed by two reviewers (the author of this thesis [MC] and one of the supervisors, Dr. I. Ioannidou [II]) independently. Any disagreement was resolved after consulting the second supervisor of this thesis Dr. M. A. Papadopoulos [MAP] after discussion and mutual agreement. Inter-reviewer agreement on study eligibility was assessed by means of the Cohen's kappa statistic using the SPSS 19.0 software (IBM Company, United States). Levels of agreement were classified as poor (Kappa < 0.00), slight (Kappa between 0.00 and 0.20), fair (Kappa between 0.21 and 0.40), moderate (Kappa between 0.41 and 0.60), substantial (Kappa between 0.61 and 0.80) and almost perfect (Kappa between 0.81 and 1.00).

Table II. Eligibility criteria used in this meta-analysis.

Criteria category	*Inclusion criteria*	*Exclusion criteria*
Outcome	Studies investigating the clinical effectiveness of chin cup on growing patients with Class III malocclusion, alone or in combination with removable dissocclusion or expansion appliances (measured with lateral cephalometric analyses or dental cast analyses)	Investigations not relevant to the study
Study design	Randomized clinical trials (RCTs) Prospective controlled clinical trials (pCCTs) Retrospective controlled clinical trials with matching control samples (rCCTs)	Unsupported opinions of expert Replies (to author/editor) Editor's choices Books' abstracts Conferences' abstracts Protocol of clinical procedures Technique description Cross-sectional surveys Uncontrolled clinical trials (prospective or retrospective) Case-control observational studies Case series without a control Case reports Reviews Systematic reviews Meta-analysis In vitro studies Animal studies/testing

		Studies on molecular biology, histology, genetics or engineering Studies on cleft lip and palate and craniofacial anomalies Studies on Class I malocclusion Studies on mandibular or maxillary protraction appliances with or without simultaneous use of chin cup Treatment outcomes given after full orthodontic treatment including chin cup and fixed appliances Geometric or morphometric assessment without cephalometric measurements Studies with no English abstract or no abstract at all
Participants' characteristics	Trials referring to human studies investigating patients during/before pubertal growth spurt (6-14 years old) at the start of treatment	Trials referring to human studies investigating patients by the end or after the pubertal growth spurt (15+ years old) at the start of treatment Studies with no matching control sample Clinical trials with inadequate sample size groups, i.e., - studies of less than 10 participants - studies not reporting the size of the examined sample
Principal outcome measures	Studies providing measurements of skeletal, dental or soft tissue profile changes as recorded by means of lateral cephalometric analyses, dental cast analyses or electromyographic analyses before and after chin cup treatment in the short- and/or long-term	Studies not providing measurements of skeletal, dental or soft tissue profile changes as recorded by means of lateral cephalometric analyses, dental cast analyses or electromyographic analyses before and after chin cup treatment in the short- and/or long-term

4.3. Data extraction and management

Two reviewers (MC and II) independently extracted study characteristics and outcomes from the included studies using pre-defined data extraction forms. Any disagreements were resolved by discussion after consulting the second supervisor of this thesis (MAP). The Cohen's kappa statistic was used again to assess the level of agreement between the two reviewers.

4.4. Assessment of risk of bias of included studies (quality analysis)

The risk of bias (quality analysis) for all included studies was performed independently by two reviewers (MC and II), with respect to pre-established characteristics. Two separate methods were used, applying both in RCTs and in non-RCTs, the one described by Antczak et al. (1986) and Jadad et al. (1996) and the Downs and Black (1998) scale. Again, any disagreements were resolved by discussion after consulting the second supervisor of this thesis (MAP).

For the Antczak et al. (1986) and Jadad et al. (1986) method the following eight domains were considered: (a) study design, (b) sample size, (c) selection description, (d) valid measurement methods, (e) method error analysis, (f) blinding in measurements, (g) adequate statistics provided and (h) confounding factors. Answers are scored 0 or 1, except for one item in the study design domain, which is scored 0 to 3. Therefore, the total maximum score a trial can receive is 10. Studies are categorized as of low (0–5 points), medium (6–8 points), or high (9 or 10 points) quality.

For the Downs and Black (1998) method, the following five domains were examined: (a) reporting (10 items), (b) external validity

(3 items), (c) internal validity – bias (7 items), (d) internal validity – confounding (selection bias) (6 items), and (e) power (1 item). Answers are scored 0 or 1, except for one item in the reporting domain, which is scored 0 to 2, and the single item on power, which is scored 0 to 5. Therefore, the total maximum score a trial can receive is 31, with a higher score indicating higher methodological quality.

Inter-reviewer agreement on quality assessment for both methods was evaluated by the Cohen's kappa statistic.

4.5. Data synthesis and analysis

Data were summarized and considered suitable for pooling if the corresponding RCTs, prospective controlled clinical trials (pCCTs), or retrospective controlled clinical trials (rCCTs) used similar interventions in the same way and reported similar outcomes. Alterations produced in growing patients following vertical or occipital pull chin cup treatment for Class III malocclusion and expressed by means of skeletal or dentoalveolar changes depicted on lateral cephalometric radiographs were pooled and estimated. Standard Difference in Means (SDM) and their corresponding 95% Confidence Intervals (CI) was applied for their statistic elaboration, since possibly different magnification factors for the original lateral cephalometric radiographs might have been used or since cephalometric points used in the primary studies for the common variables examined might have not been identical. Standard difference in means instead of Mean Difference (MD) is generally preferred when raw data derived from the primary studies are neither identically measured nor calibrated (Deeks et al., 2001; Janket et al., 2005). Soft tissue, cast model and perioral muscular electromyography data analysis were also to be performed, where data available, and to examine muscular adaptation to new position measured by means of dimension and strength, lips competence or incompetence, soft tissue profile alterations, 3-dimentional (3D) full mouth teeth position evaluation and subsequent occlusal alterations. If included studies were less than 5, due to inadequacy in quantity of the pooled data, exploratory analysis was to be preformed instead and its derivative was regarded as secondary in validity and must be viewed with caution (Anello, 1995; Papageorgiou et al., 2012).

Weighting of the pooled estimates, was performed with the random-effects model since it was expected that one or more non-RCTs would be included (Borenstein et al. 2009; DerSimonian and Laird, 1986). This model was used because the random effects model takes into account the heterogeneity of the data and it is considered more conservative in terms of the significance intervals (Borenstein et al., 2009). In addition, since the observed effect was expected to differ across studies due to sample and implementation differences, the use of this model was seemed as more appropriate (Janket et al., 2005; Kolokitha et al., 2008; Papageorgiou et al., 2012).

Analyses were performed by means of the statistical software “Comprehensive Meta-Analysis” version 2.0 (Biostat Inc., Englewood, NJ, USA) using the random-effects model. The pooled estimate (SMDs) of the examined variables and the corresponding 95% CIs were used to construct a forest plot. Level of significance was set at $P \leq 0.05$ except for heterogeneity where it was set at $P \leq 0.10$. *P*-values were two-sided.

4.6. Assessment of publication bias

The funnel plots of the examined variables were visually inspected in an effort to identify and evaluate publication bias (Light and Pillemer, 1984). If asymmetry was observed, the presence of publication bias was additionally tested statistically using the Begg and Mazumdar's (1994) test (rank correlation method), as well as the Egger's (1997) test (weighted regression).

4.7. Heterogeneity assessment

To identify the extent of heterogeneity, the Cochran Q test for homogeneity with the corresponding P-value was calculated (Higgins et al., 2003; Huedo-Medina, 2006; Pigott, 2012). Q test, being a type of chi-square test, follows the normal distribution and can theoretically take values from $-\infty$ to $+\infty$ but its results are regarded as statistically significant if the corresponding P-value is higher than 0.10 (Huedo-Medina, 2006; Pigott, 2012). However, the Cochran Q test has poor power to detect heterogeneity if a few number of studies are included in the meta-analysis and, conversely, has very high power to detect negligible variability if many studies are included (Higgins et al., 2003; Huedo-Medina, 2006). Therefore, the additional use of the I^2 index was deemed necessary. I^2 index measures the extent of true heterogeneity by dividing the difference between the result of the Q test and its degrees of freedom ($k - 1$) by the Q value itself, and multiplied by 100 (Huedo-Medina, 2006). I^2 index is, therefore, given in percentages and, due to its nature, it does not depend upon the number of studies, the type of outcome data or the choice of treatment effect that has been used to quantify the impact of heterogeneity and assess inconsistency within studies (Higgins et al., 2003). The I^2 index not only identifies the existence of heterogeneity but also presents an estimate of its extend (Higgins et al., 2003; Huedo-Medina, 2006). According to Higgins et al. (2003), a value of 0% for the I^2 index equals with no observed heterogeneity, while values of 25%, 50% and 75% equal with low, medium and high heterogeneity respectively.

Due to the random effects model application, T^2 was also used to quantify the impact of heterogeneity between studies (Borenstein, 2009). I^2 index and T^2 are directly related (Huedo-Medina, 2006; Pigott, 2012). For a high value of heterogeneity (I^2=75%) T^2 will be high as well (T^2=3v, where v represents the within studies variance) while for I^2=50% T^2=3v and for I^2=50% T^2=1/3v respectively (Pigott, 2012).

With a retrospective view on the meta-analysis model chosen to be applied, one could say that high values of I^2 and T^2 indexes can justify the selection of the random effects model (Borenstein, 2009; Higgins et al., 2003; Huedo-Medina, 2006; Pigott, 2012).

4.8. Subgroup analysis

Where concomitant, well-described modifiers (factors) were identified, expressed by means of different protocols in the chin cup application, subgroup analysis was conducted as proposed by Deeks et al. (2001). The following two modifiers, which were set as categorical, led to the formation of two respective subgroups: (a) force level of the chin cup appliance and (b) simultaneous additional appliance used.

The procedures followed were those of Comprehensive Meta-analysis (www.meta-analysis.com) as suggested by Borenstein et al. (2009). More specifically, for the subgroup analysis a mixed effects model was applied: studies within each subgroup were combined with the use of a random effects model, while the subgroups were combined with a fixed effect model and yielded the overall effect (Borenstein et al., 2009). The between studies variance (T^2) was not assumed to be the same for all subgroups, so it was computed within subgroups and not pooled across subgroups (Borenstein et al., 2009).

5. RESULTS

5.1. Literature flow

Among 1238 initially identified relevant records (870 through electronic searching and 368 through manual searching of the aforementioned journals) 229 unique citations remained after duplicate removal. One hundred eighty three of them were excluded on the basis of title, abstract, non-existing abstract or non English abstract and 46 articles remained to be assessed on a full text basis. Hand searching on the reference list of these 46 full text evaluated articles revealed 5 additional relevant records. Of them, 3 were excluded due to non-English abstract and a sum of 48 articles remained for full text eligibility assessment. Application of the previously reported inclusion and exclusion criteria ended in the exclusion of another 43 articles (Appendix Table 1). The numbers of excluded articles, according to the exclusion criteria are listed in summary in Table III. In total, 5 studies (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Barrett et al., 2010; Gökalp and Kurt, 2005; Tuncer et al., 2009) remained for qualitative and quantitative synthesis. These were categorized according to study design into 4 pCCTs and 1 rCCTs. No RCT was found to be eligible for inclusion in this meta-analysis, despite the established usefulness of the chin cup appliance in clinical practice and the plethora of publications on chin cup. The flow diagram of the retrieved studies is presented in Figure 8.

The Cohen's kappa between the two investigators before reconciliation for data selection was 0.838 (standard error: 0.042) and for data extraction 0.898 (standard error: 0.100), both of which represent an almost perfect agreement.

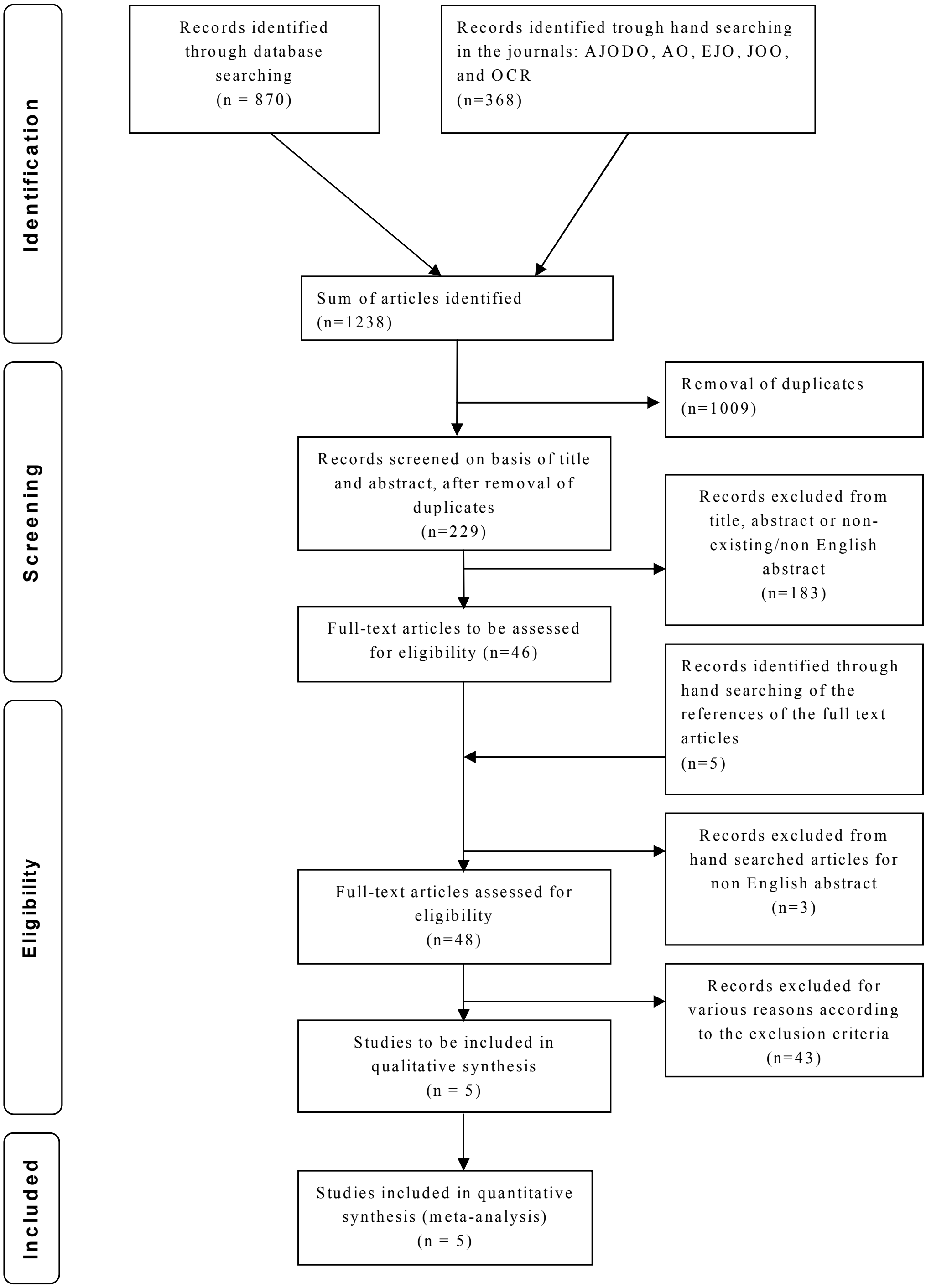

Figure 8. The PRISMA flowchart of the retrieved studies indicating the selection process.

Table III. Number of the excluded articles according to the exclusion criteria.

Exclusion Criteria	*Number of Excluded Articles*
Investigations not relevant to the study	54
Unsupported opinions of expert	3
Reviews	1
Protocol of clinical procedures	4
Technique description	4
Uncontrolled clinical trials (prospective or retrospective)	32
Case series without a control	12
Case reports	15
Systematic reviews	4
Meta-analysis	1
In vitro studies	4
Animal studies/testing	3
Studies on molecular biology, histology, genetics or engineering	5
Studies on cleft lip and palate and craniofacial anomalies	1
Studies on Class I malocclusion	3
Studies on mandibular or maxillary protraction appliances with or without simultaneous use of chin cup	22
Treatment outcomes given after full orthodontic treatment including chin cup and fixed appliances	7
Geometric or morphometric assessment without cephalometric measurements	2
Studies with no English abstract or no abstract at all	38
Studies not providing measurements of skeletal, dental or soft tissue profile changes as recorded by means of lateral cephalometric analyses or dental cast analyses before and after chin cup treatment in the short- and/or long-term	5
Studies with no matching control sample	4
Total	***224***

5.2. Description of studies and baseline characteristics

The characteristics of the included studies (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Barrett et al., 2010; Gökalp and Kurt, 2005; Tuncer et al., 2009) in this meta-analysis are presented in Table IV. As reported above, not even one RCT was possible to be included in current evaluation whereas in total, 4 *p*CCTs and 1 *r*CCT were included.

In detail, Altuğ et al. (1989) examined in a prospective controlled clinical trial (*p*CCT) performed in Ankara University the effects of chin cup during a period of 0.7 years and compared them with an untreated Class III control group. Mean age at the start of treatment ranged between 9.7 and 11 years and measurements were performed on lateral cephalometric radiographs. The authors did not report anything about the force magnitude and the suggested hours of chin cup wear.

Gökalp and Kurt (2005) investigated in a *p*CCT the results of the chin cup for Class III patients mainly in relation to the TMJ, condylar growth pattern and disc position, but also conventional measurements were made on lateral cephalometric radiographs. The treatment group was matched with a corresponding Class III untreated control sample. Mean age ranged between 8.90 and 9.06 years at the start of treatment, while treatment duration was 1.7 years. The applied force magnitude was 600 gr, however a specification if this was in total or per side was missing and following relevant question to the authors no reply was received. Patients were expected to wear the appliance 18 hours per day. Measurements were made on lateral cephalometric radiographs and on magnetic resonance images (MRIs) in a closed mouth position. Only comparable measurements from lateral cephalometric radiographs were examined quantitatively in this meta-analysis.

Abdelnaby and Nassar (2010) conducted a prospective controlled clinical trial (*p*CCT) and examined in two separate Class III groups of patients the effects of the light force chin cup, where the force magnitude applied was 300 gr per side, and the heavy force chin cup, where the force magnitude applied was 600 gr per side. The concomitant use of an upper occlusal bite plane was aiming to dissocclude the anterior teeth, where cross bite was present. These two groups of patients, being independent, were examined separately as different study groups in the meta-analysis. Appliance wear was suggested to be 14 hours per day. The control group consisted also of Class III patients and received no treatment during the examination period, which was 1 year. Mean age of patients at the start of treatment was 9-10 years and all of them had not passed the peak of pubertal growth spurt. Measurements were made on lateral cephalometric radiographs.

Barrett et al. (2010) performed a *p*CCT where they examined the effects of light force chin cup (150-250 gr in total, or 75-125 gr per side) in Class III patients for a period of 1 year and compared them with a matched control group. Twelve of the 26 treated patients also received a quad helix appliance and, therefore, the treated group was divided in two subgroups, one group consisting of 14 patients who received the chin cup only and one group consisting of 12 patients who received the chin cup and the quad helix. These two subgroups were examined separately as being independent in current meta-analysis. Patients were advised to wear the appliance at night only. Mean age at the start of treatment ranged between 7.3 and 8.5 years. Measurements were performed on lateral cephalometric radiographs.

Finally, Tuncer et al. (2009) performed a retrospective controlled clinical trial (*r*CCT) and compared the effects of the chin cup used with a force magnitude of 600 gr in total (300 gr per side) in Class III patients for

0.8 years with a Class III control group observed for 0.95 years. Similarly with Abdelnaby and Nassar, Tuncer et al. in their study also described the placement of an upper removable bite plane to produce dissocclusion in the anterior area of the dental arch to all treated patients. Mean age at the start of treatment ranged between 8.5 and 11.5 years of age, and all subjects, being before the peak of growth spurt, were suggested to wear the appliance 14-16 hours per day. Measurements were performed on lateral cephalometric radiographs.

Taking into consideration the characteristics and the data provided from the included articles although the initial plan was to investigate the effects of both the occipital pull and the vertical pull chin cup on their short- and long-term effects, only the short-term occipital pull chin cup effects were finally examined since no study with vertical chin cup appliance and no study with long-term effects from occipital pull chin cup was finally included. Consequently, where the term "chin cup" is used thereafter, it is referred to the occipital pull chin cup and where the term "clinical effects" is used is limited to the short term ones.

Soft tissue, cast model and perioral muscular electromyography data analysis also was not possible to be performed for no such data has been retrieved and considered as appropriate for inclusion and analysis in the present study. Therefore, treatment effect comparisons between the experimental groups were considered just for skeletal and dentoalveolar alterations as measured on lateral cephalometric radiographs.

Table IV. Characteristics of the studies included in the meta-analysis.

Nr	***1***		***2***	***3***		***4***	***5***
Study*	***Abdelnaby and Nassar (2010)a***	***Abdelnaby and Nassar (2010)b***	***Altuğ et al. (1989)***	***Barrett et al. (2010)a***	***Barrett et al. (2010)b***	***Gökalp and Kurt 2005***	***Tuncer et al. 2009***
Study design	pCCT		pCCT	pCCT		pCCT	rCCT
Treated Sample origin and/or characteristics	Growing patients with skeletal Class III pattern, protrusive mandible and anterior cross bite. Origin not defined. All patients had not passed the peak of pubertal growth spurt.		Patients with skeletal Class III pattern from the Orthodontic Department in Ankara University.	Consecutively treated patients with Class III pattern and Witts >-2 Origin not defined.		Patients with mandibular prognathism. Origin not defined.	Patients with skeletal Class III pattern mandibular prognathism and dental Class III malocclusion, without forward functional displacement, optimal mandibular plane angle and no congenital anomalies. All patients had not passed the peak of pubertal growth spurt. Origin not defined.
Treated sample size	40		21	26		13	20
Subgroups	Heavy force occipital pull chin	Light force occipital pull	Not specified	Light force chin cup (N=14)	Light force chin cup and	No	No

	cup and upper anterior bite plane (N=20)	chin cup and upper bite plane (N=20)			quadhelix (N=12)		
Gender (M / F)	10/10	9/11	11/10	–	–	3/10	10/10
Mean age	9.6	10.1	11	8.5	8.5	9.06	10.31
Control sample origin and/or characteristics	The same origin with the treated group		Patients with skeletal Class III pattern form from the Orthodontic Department in Ankara University	Patients with Class III malocclusion (matched with the treated sample) from the University of Florence and the University of Michigan Growth Study.		Patients with mandibular prognathism. Origin not defined.	Patients with skeletal Class III pattern and dental Class III malocclusion, mandibular prognathism without forward functional displacement, optimal mandibular plane angle and no congenital anomalies. Patients matched with the treated sample according to skeletal developmental stage. Origin not defined.
Control sample size	10		9	20		7	18
Gender (M / F)	5/5		4/5	6/14		1/6	10/8
Mean age	9.2		9.7	7.3		8.9	9.89
Diagnosis	Class III		Class III	Class III		Class III	Class III
Method of measurement	Lateral cephalometric radiographs, hand-wrist films.		Lateral cephalometric radiographs.	Lateral cephalometric radiographs.		Lateral cephalometric radiographs, hand-	Lateral cephalometric radiographs, hand-wrist films

						wrist films, unilateral-left sagittal -oblique TMJ magnetic resonance images (MRIs) in a closed mouth position.	
Appliance used	Heavy force occipital pull chin cup (600 gr per side).	Light force occipital pull chin cup (300 gr per side).	Occipital pull chin cup. Force not specified.	Light force occipital pull chin cup (75-125 gr per side).	Light force occipital pull chin cup (75-125 gr per side).	Chin cup (600 gr).	Light force occipital pull chin cup (300 gr per side).
Additional appliance used	Acrylic upper bite plane.	Acrylic upper bite plane.	–	–	Quadhelix	–	Lower bite plane.
Suggested hours of use	14h/day		Not defined.	At night only.		18 hrs/day	14-16h/day
Treatment follow-ups	1 year		0.7years	1 year		1.7 years	0.8 years
Treatment duration	Not defined.		Not defined.	2.6 years for boys, 2.4 years for girls.		1.7 years	0,8 years
Control observation	1 year		0.7 years	2.4 years		1.7 years	0,95 years
Reported effects	Significant increase in ANB, Wits appraisal, anterior face height and mandibular	Significant increase in ANB, Wits appraisal, anterior face height and	Significant increase in ANB and in Wits appraisal. Significant de-	Significant increase in ANB. Slight increase in Wits appraisal. Slight decrease in	Significant increase in ANB. and in Wits appraisal. Significant decrease in	Significant increase in ANB. Significant decrease in SNB and gonial angle. Increase in man-	Significant increase in ANB, anterior face height, mandibular plane angle, maxillary forward position and in ef-

	plane angle. Significant decrease in SNB and ramus height. Significant increase in retroclination of mandibular incisors.	mandibular plane angle. Significant decrease in SNB but less decrease in ramus height. Further retroclination of mandibular incisors.	crease in SNB and maxillary plane angle. Decrease in mandibular dimensions Pg-ramus inferior tangent point, Pg-ramus superior tangent point and Pg-Cd.	SNB. Significant decrease in maxillary plane angle. Significant increase in retroclination of mandibular incisors.	SNB, gonial angle and in mandibular length and maxillary plane angle. Significant increase in protrusion of maxillary incisors. Less increase in retroclination of mandibular incisors.	dibular corpus and ramus length. Sagittal mandibular growth was inhibited. Sagittal maxillary growth was activated. Significant increase in overjet.	fective maxillary and mandibular length. Significant decrease in SNB and clockwise rotation of the mandible. Increase in pharyngeal dimension and downward movement of the hyoid bone.

**Authors in alphabetical order; M/F, male / female; RCT, Randomized controlled trial; pCCT, Prospective controlled clinical trial; rCCT, Retrospective controlled clinical trial.*

5.3. Cephalometric variables assessed in the original studies

The common cephalometric variables retrieved from the 5 included studies, and thus were possible to be examined in current meta-analysis were the following:

a) <u>*Skeletal variables*</u>:

αα) in the sagittal plane:

i) **SNA** (^{o}): angle formed between the cephalometric points S, N and A

ii) **SNB** (^{o}): angle formed between the cephalometric points S, N and B

iii) **ANB** (^{o}): angle formed between the cephalometric points A, N and B

iv) **Wits appraisal** (mm): the perpendicular distance of the cephalometric points A and B to the functional occlusal plane

v) **Co-Gn** (mm): distance between the cephalometric points Co and Gn

ab) in the vertical plane:

vi) **SN-ML** (^{o}): angle formed between the cephalometric planes SN and Go-Me or Go-Gn

vii) **Gonial angle** (^{o}): angle formed between the cephalometric points Ar, Go and Me

viii) **N-Me** (mm): distance between the cephalometric points N and Me

ix) **UFH** (mm): distance between the cephalometric points N and ANS

x) **LAFH** (mm): distance between the cephalometric points ANS and Me

xi) **Co-Go** (mm): distance between the cephalometric points Co and Go

b) Dentoalveolar variables:

i) **Overjet** (mm): horizontal distance between the cephalometric points “Upper incisor tip” and “Lower incisor tip”

ii) **Overbite** (mm): vertical distance between the cephalometric points “Upper incisor tip” and “Lower incisor tip”

The cephalometric points and the variables examined are presented in Figures 9 and 10 respectively.

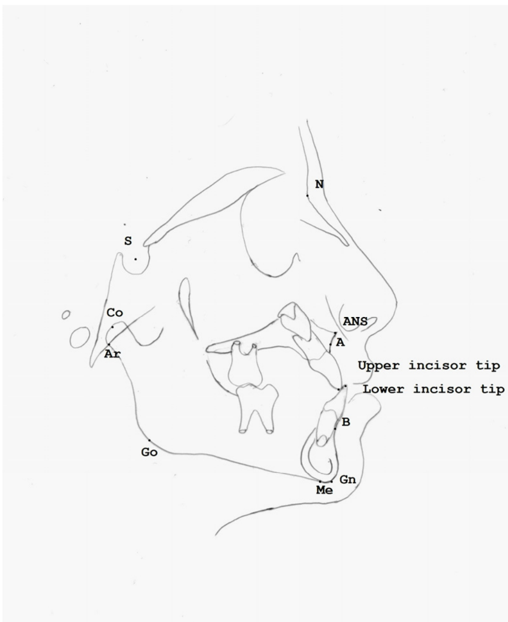

Figure 9. Cephalometric points used for the variables examined in current investigation.

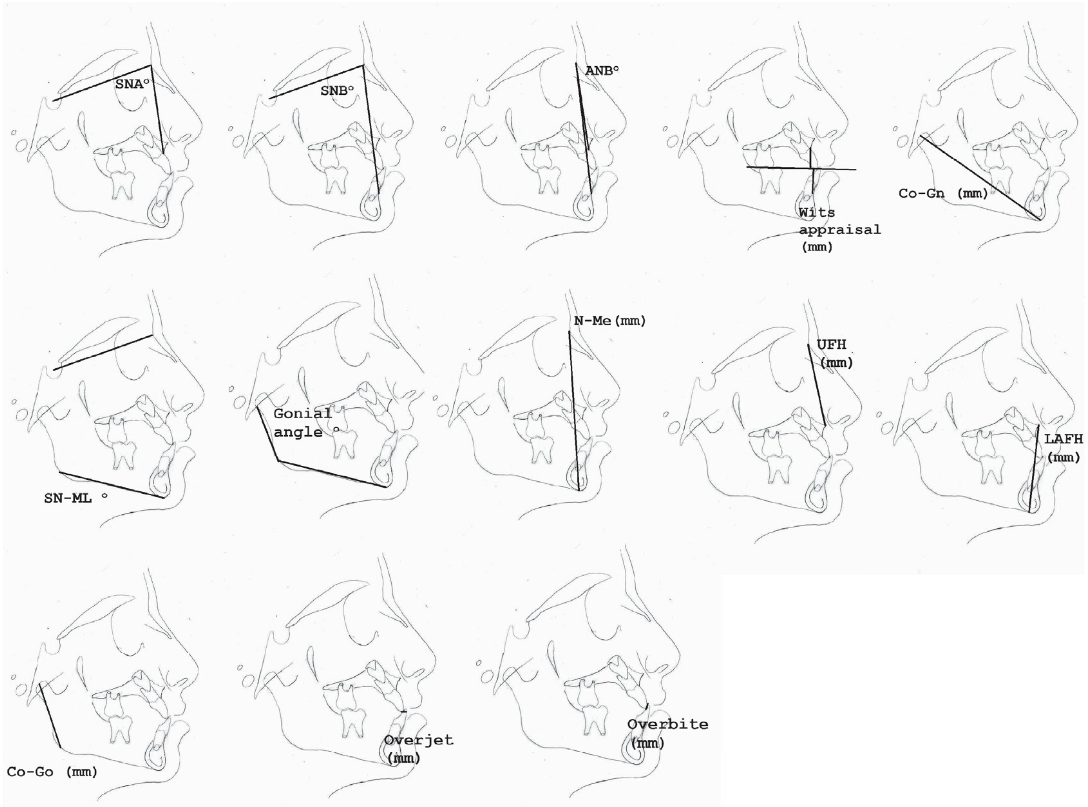

Figure 10. Cephalometric variables examined in current investigation.

The contribution of the original studies to the investigation of each individual cephalometric variable assessed in current meta-analysis is presented in Table V. For the cephalometric variables SNA, SNB and ANB all 7 treated groups from all 5 studies (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Barrett et al., 2010; Gökalp and Kurt, 2005; Tuncer et al., 2009) participated in the formation of the results. For the Wits appraisal 5 treated groups from 3 studies (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Barrett et al., 2010) were assessed qualitatively and quantitatively, while for Co-Gn data from 4 treated groups of 3 studies (Barrett et al., 2010; Gökalp and Kurt, 2005; Tuncer et al., 2009) have been taken into consideration. The results for the angular cephalometric variable SN-ML were derived from 5 treated groups of 4 included studies (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Gökalp and Kurt, 2005; Tuncer et al., 2009) and for the Gonial angle they were based on 6 treated groups from 4 participating studies (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Barrett et al., 2010; Gökalp and Kurt, 2005). For the linear cephalometric variable N-Me outcomes were derived from 2 studies (Abdelnaby and Nassar, 2010; Tuncer et al., 2009) that formed 3 treated groups and the variables UFH and LAFH they were also derived from 2 studies (Barrett et al., 2010; Tuncer et al., 2009) or 3 treated groups respectively. Finally, 2 studies (Barrett et al., 2010; Gökalp and Kurt, 2005) forming 3 treated groups, contributed to the results for the cephalometric variables Co-Go, Overjet and Overbite.

Table V. Contribution of the original studies to the investigation of the individual cephalometric variables assessed in current meta-analysis.

	*Study**								
	Abdelnaby and Nassar 2010a	**Abdelnaby and Nassar 2010b**	**Altuğ et al. 1989**	**Barrett et al. 2010a**	**Barrett et al. 2010b**	**Gökalp and Kurt 2005**	**Tuncer et al. 2009**	***Sum of studies***	***Sum of separate treated groups***
Force level of the chin cup	Heavy	Light	Not specified	Light	Light	Heavy	Light		
Additional appliance used	Bite plane	Bite plane	No	No	Quad helix	No	Bite plane		
SNA (^{o})	yes	yes	yes	yes	yes	yes	yes	5	7
SNB (^{o})	yes	yes	yes	yes	yes	yes	yes	5	7
ANB (^{o})	yes	yes	yes	yes	yes	yes	yes	5	7
Wits appraisal (mm)	yes	yes	yes	yes	yes	no	no	3	5
Co-Gn (mm)	no	no	no	yes	yes	yes	yes	3	4
SN-ML (^{o})	yes	yes	yes	no	no	yes	yes	3	5
Gonial angle (^{o})	yes	yes	yes	yes	yes	yes	no	4	6
N-Me (mm)	yes	yes	no	no	no	no	yes	2	3
UFH (mm)	no	no	no	yes	yes	no	yes	2	3
LAFH (mm)	no	no	no	yes	yes	no	yes	2	3
Co-Go (mm)	no	no	no	yes	yes	yes	no	2	3
Overjet (mm)	no	no	no	yes	yes	yes	no	2	3
Overbite (mm)	no	no	no	yes	yes	yes	no	2	3

**Authors in alphabetical order.*

5.4. Risk of bias analysis (Quality evaluation of the included studies)

The risk of the bias analysis of the included studies according to the Antczack et al. (1986) and Jadad et al. (1996) method is presented in Table VI, while according to the Downs and Black (1998) scale is presented in Table VII.

In detail, according to the Antczack et al. (1986) and Jadad et al. (1996) method (Table VI), selection description was adequate in 3 out of the 5 studies (Abdelnaby and Nassar, 2010; Barrett et al., 2010; Tuncer et al., 2009), partly adequate in 1 study (Gökalp and Kurt, 2005), and inadequate in another one (Altuğ et al., 1989). The measurement methods and statistics used to retrieve data of the cephalometric variables were valid and adequate in 4 out of the 5 studies (Abdelnaby and Nassar, 2010; Barrett et al., 2010; Gökalp and Kurt, 2005; Tuncer et al., 2009), whereas in the fifth one (Altuğ et al., 1989) they were partly adequate. Only one study (Barrett et al., 2010) reported blinding in measurements and none of the studies reported confounding factors. According to this evaluation 3 studies were found to present medium quality (Abdelnaby and Nassar, 2010; Barrett et al., 2010; Tuncer et al., 2009) and 2 to present low quality (Altuğ et al., 1989; Gökalp and Kurt, 2005), while the overall quality of the 5 included studies was low.

Further, according to the Downs and Black (1998) scale (Table VII), in the reporting part, only two studies (Abdelnaby and Nassar, 2010; Tuncer et al., 2009) clearly described the characteristics of the patients included whereas no study described the characteristics of patients lost to follow-up. Only one study (Altuğ et al., 1989) fulfilled the requirements of external validity by recruiting and presenting a sample representative of the entire population. In internal validity-bias, only one study (Abdelnaby and Nassar, 2010) reported blinding of the subjects submitted to intervention and one study (Barrett et al., 2010) described an attempt to blind those measuring the main outcome of the intervention, whereas no study reported whether compliance with the extraoral appliance used was reliable or not. In internal validity-confounding (selection bias) only one study (Abdelnaby and Nassar, 2010) assigned randomly the subjects to intervention groups and in no study were losses of patients to follow-up taken into account.

According to the classification factor sample size (sample size=N), the power of the included studies to detect a clinically important effect, where the probability value for a difference being due to chance is less than 5%, range in a scale from 1 to 5 (Downs and Black, 1998). The correspondence between the sample sizes (N) and the power of the study ranks as following: where sample size less than 10, the power of the study equals to zero (N<10=0), for sample sizes between 10 and 12, the power of the study equals to 1 (N:10-12=1), for sample sizes between 13 and 15, the power of the study equals to 2 (N:13-15=2), for sample sizes between 16 and 18, the power of the study equals to 3 (N:16-18=3), for sample sizes between 19 and 21, the power of the study equals to 4 (N:19-21=4), and for sample sizes larger than 21, the power of the study equals to 5 (N>21=5). In the present meta analysis, no study had a power of 5. Four treated groups from 3 studies (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Tuncer et al., 2009) had a power of 4 which is generally good, but the rest two studies (Barrett et al., 2010; Gökalp and Kurt, 2005) had a very small power of 2 or less.

In total, according to the Downs and Black (1998) scale, the 5 studies scored an overall score between 15 and 21, which is far away from the maximum score of 31, and thus their overall quality could be considered as low to medium.

The Cohen's kappa between the two investigators for quality evaluation was for the method according to Antczack et al. (1986) and Jadad et al. (1996) 0.737 (standard error: 0.200) (Table VIII), which is considered substantial. For the method according to Downs and Black (1998) was 0.750 (standard error: 2.100) for reporting, which is substantial, and 1 (standard error: 0.000) for external validity, internal validity, and power, which is perfect (Table IX).

Table VI. Risk of the bias analysis of the included studies according to the method of Antczack et al. (1986) and Jadad et al. (1996).

Nr.	*Study**	*Study design*	*Sample size*	*Selection description*	*Valid measurement methods*	*Method error analysis*	*Blinding in measurements*	*Adequate statistics provided*	*Confounding factors*	*Score*	*Judged quality standard*
1	Abdelnaby and Nassar 2010	1	0.5	1	1	1	0	1	0	5.5	Medium
2	Altuğ et al. 1989	0	0	0.5	0.5	0	0	0.5	0	1.5	Low
3	Barrett et al. 2010	1	0.5	1	1	1	1	1	0	6.5	Medium
4	Gökalp and Kurt 2005	1	0	0	1	1	0	1	0	4	Low
5	Tuncer et al. 2009	1	0.5	1	1	1	0	1	0	5.5	Medium
Overall estimate										**4.6**	**Low**

**Authors in alphabetical order.*

Table VII. Quality analysis of the included studies according to the Downs and Black (1998) scale.

Nr.		**1**	**2**	**3**	**4**	**5**
Study*		**Abdelnaby and Nassar 2010**	**Altuğ et al. 1989**	**Barrett et al. 2010**	**Gökalp and Kurt 2005**	**Tuncer et al. 2009**
Reporting	*Is the hypothesis/aim/objective of the study clearly described?*	1	1	1	1	1
	Are the main outcomes to be measured clearly described in the Introduction or Methods section?	1	1	1	1	1
	Are the characteristics of the patients included clearly described?	1	0	0	0	1
	Are the functional appliances used clearly described?	1	0	1	0	1
	Are the distributors of principal confounders in each group of subjects to be compared clearly described	1	0	1	1	1
	Are the main findings of the study clearly described?	1	1	1	1	1
	Does the study provide estimates of the random variability in the data for the main outcomes?	1	1	1	1	1
	Have all important adverse events that may be a consequence of functional appliances been reported?	1	0	0	1	1
	Have the characteristics of patients lost to follow-up been described?	0	0	0	0	0
	Have actual probability values been reported for the main outcomes except where the probability value is less than 0.001?	0	0	1	1	1
External validity	*Were the patients asked to participate in the study representative of the entire population from which they were recruited?*	0	1	0	0	0
	Were those subjects who were prepared to participate representative of the entire population from which they were recruited?	0	1	0	0	0
	Were the staff, places and facilities where the patients were treated representative of the treatment the majority of patients receive?	0	1	1	1	0

	Was an attempt made to blind study subjects for the intervention they had received?		1	0		0	0	0
Internal validity - bias	*Was an attempt made to blind those measuring the main outcome of the intervention?*		0	0		1	0	0
	If any of the results of the study were based on "data dredging", was that made clear?		1	1		1	1	1
	In trials, do the analyses adjust for different lengths of follow-up of patients, or in case control studies, is the time and period between the intervention and outcome the same for cases and controls?		1	1		0	1	1
	Were the statistical tests used to assess the main outcomes appropriate?		1	1		1	1	1
	Was compliance with the extraoral appliance used reliable?		0	0		0	0	0
	Were the main outcome measures used accurate (valid and reliable)?		1	1		1	1	1
Internal validity – confounding (selection bias)	*Were the patients in different intervention groups recruited from the same population?*		1	1		0	0	1
	Were study subjects in different intervention groups recruited over the same period of time?		1	1		1	1	1
	Were study subjects randomized to intervention groups?		1	0		0	0	0
	Was the randomised intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?		0	0		0	0	0
	Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?		1	0		1	1	1
	Were losses of patients to follow-up taken into account?		0	0		0	0	0
		a	**b**		**a**	**b**		
Power	*Did the study (or the independent treated groups of the study where this applicable) have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?*	4	4	4	2	1	2	4
Sum		**21**	**21**	**17**	**16**	**15**	**16**	**20**

***Authors in alphabetical order; yes=1, no=0, unable to determine=0.*

Table VIII. Kappa scores measuring levels of agreement between the two reviewers in assessing quality scores of the included articles according to the method of Antczack et al. (1986) and Jadad et al. (1996).

Nr.	***Study****	***Reviewer #1***	***Reviewer #2***
1	Abdelnaby and Nassar 2010	5.5	5.5
2	Altuğ et al. 1989	1.5	1.5
3	Barrett et al. 2010	6.5	6.5
4	Gökalp and Kurt 2005	3.5	4.5
5	Tuncer et al. 2009	5.5	5.5
	Overall estimate		**0.737 (Standard error 0.200)**

**Authors in alphabetical order.*

Table IX. Kappa scores measuring levels of agreement between the two reviewers in assessing quality scores of the included articles according to the Downs and Black (1998) scale.

		Reporting		***External validity***		***Internal validity - bias***		***Internal validity – con-founding (selection bias)***		***Power***	
Nr.	*Study**	*Re-viewer #1*	*Re-viewer #2*	*Re-viewer #1*	*Re-viewer #2*	*Re-viewer #1*	*Re-viewer #2*	*Reviewer #1*	*Reviewer #2*	*Re-viewer #1*	*Re-viewer #2*
1	Abdelnaby and Nassar 2010	8	8	0	0	5	5	4	4	4	4
2	Altuğ et al. 1989	4	4	2	2	5	5	2	2	4	4
3	Barrett et al. 2010	7	7	0	0	5	5	2	2	1.5	1.5
4	Gökalp and Kurt 2005	6	7	0	0	5	5	2	2	2	2
5	Tuncer et al. 2009	9	9	0	0	4	4	3	3	4	4
Overall estimate		**0.750 (Standard error 2.100)**		**1 (Standard error 0.000)**		**1 (Standard error 0.000)**		**1 (Standard error 0.000)**		**1 (Standard error 0.000)**	

**Authors in alphabetical order.*

5.5 Assessment of publication bias

Indications of asymmetry were observed in the funnel plots of all the variables examined (Figures 11-23). Due to the observed asymmetry in the funnel plots, which was present at different extend in different variables and also due to the small number of included studies, this method was deemed partly adequate to trace existent publication bias. Thus, publication bias was then further assessed with the Begg and Mazumdar's (1994) test and the Egger's (1997) test to appraise the statistical significance of the observed publication bias (level of significance: $P \leq 0.01$) (Table X). In summary, statistically significant publication bias was found for the variables SNB (Figure 12), ANB (Figure 13) Wits appraisal (Figure 14), Gonial angle (Figure 17), N-Me (Figure 18) and Co-Go (Figure 21). No statistically significant was the publication bias observed for the variables SNA (Figure 11), Co-Gn (Figure 15), SN-ML (Figure 16), UFH (Figure 19), LAFH (Figure 20), Overjet (Figure 22) and Overbite (Figures 23).

Table X. Publication bias of the variables examined in the meta-analysis.

Variables	***N of treated groups***	***Observed asymmetry in the corresponding funnel plot***	***Publication bias***	
			Begg and Mazumdar's test (P-value)	***Eggers' regression (P-value)***
SNA (^{o})	7	Yes	-0.238 (0.453)	-9.217 (0.133)
SNB (^{o})	7	Yes	-0.238 (0.453)	-10.188 (0.010)*
ANB (^{o})	7	Yes	0.905 (0.004)**	10.055 (0.003)**
Wits appraisal (mm)	5	Yes	1.000 (0.014)*	10.673 (0.001)**
Co-Gn (mm)	4	Yes	-0.500 (0.308)	-5.193 (0.216)
SN-ML (^{o})	5	Yes	0.200 (0.624)	0.088 (0.994)
Gonial angle (^{o})	6	Yes	-0.600 (0.091)	-11.108 (0.024)*
N-Me (mm)	3	Yes	1.000 (0.117)	9.463 (0.030)*
UFH (mm)	3	Yes	-0.333 (0.602)	-18.209 (0.422)
LFAH (mm)	3	Yes	-0.333 (0.602)	-17.769 (0.374)
Co-Go (mm)	3	Yes	-1.000 (0.117)	-2.740 (0.002)**
Overjet (mm)	3	Yes	1.000 (0.117)	-3.633 (0.656)
Overbite (mm)	3	Yes	-0.333 (0.602)	-6.702 (0.076)

**statistically significant at $P \leq 0.05$, ** statistically significant at $P \leq 0.01$, ***statistically significant at $P \leq 0.00$*

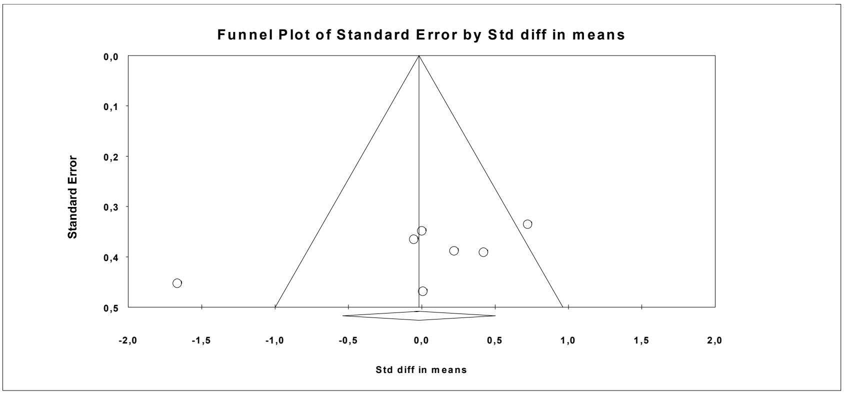

Figure 11. Funnel plot of the cephalometric variable SNA.

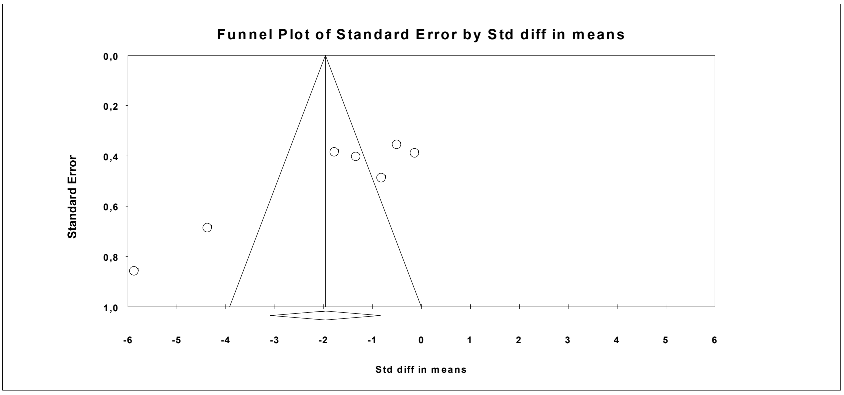

Figure 12. Funnel plot of the cephalometric variable SNB.

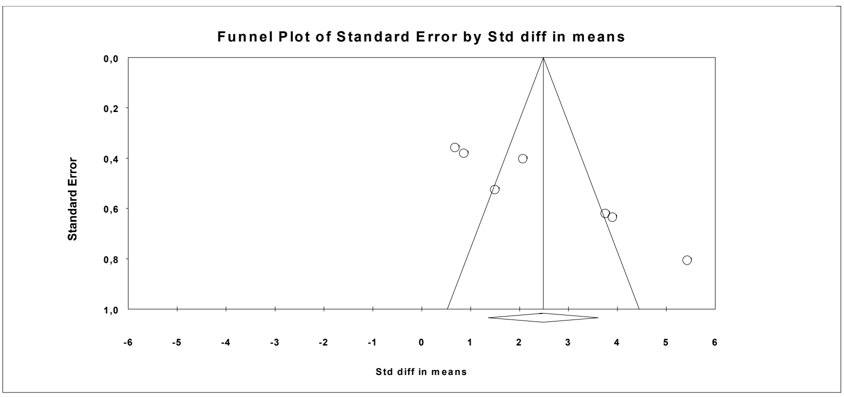

Figure 13. Funnel plot of the cephalometric variable ANB.

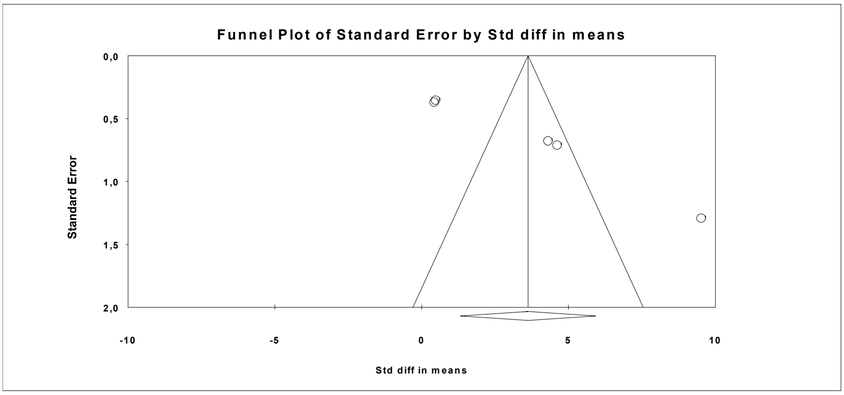

Figure 14. Funnel plot of the cephalometric variable Wits appraisal.

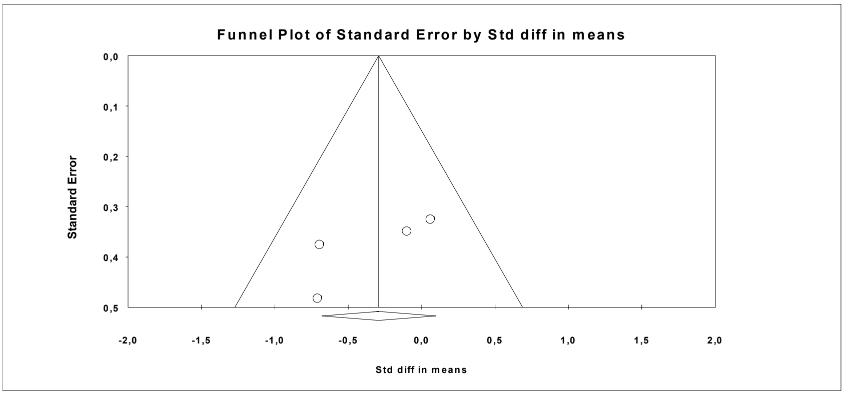

Figure 15. Funnel plot of the cephalometric variable Co-Gn.

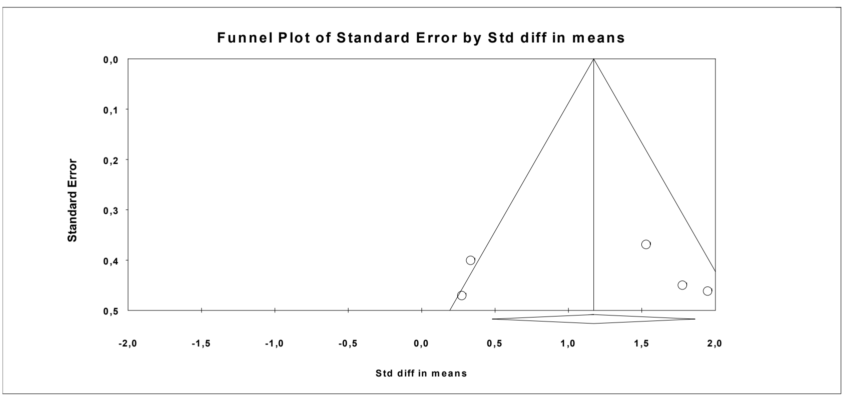

Figure 16. Funnel plot of the cephalometric variable SN-ML.

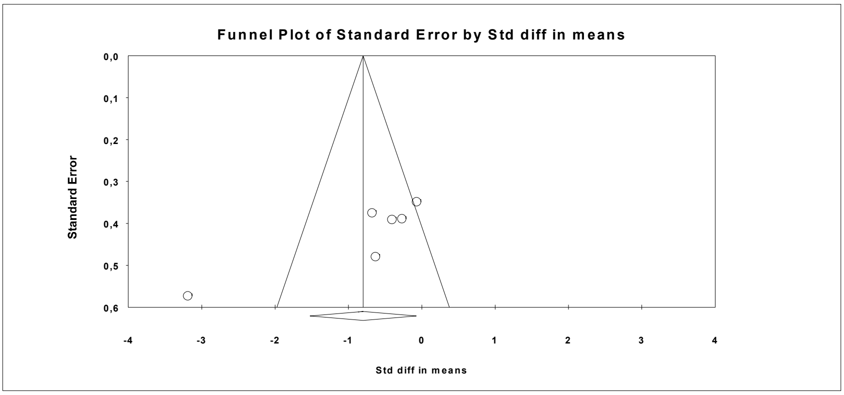

Figure 17. Funnel plot of the cephalometric variable Gonial angle.

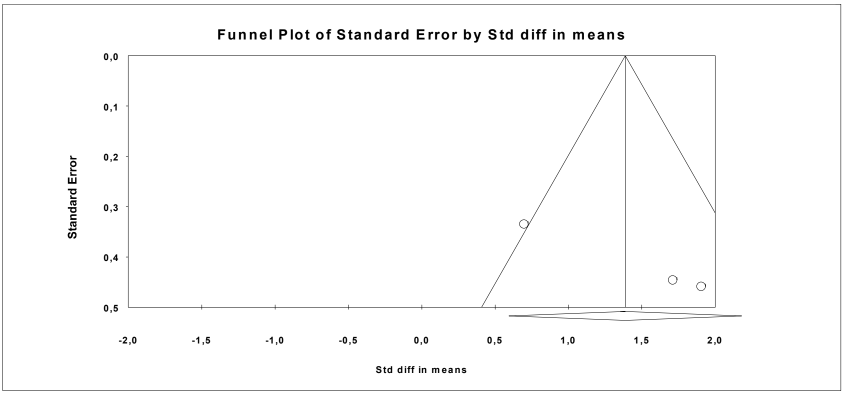

Figure 18. Funnel plot of the cephalometric variable N-Me.

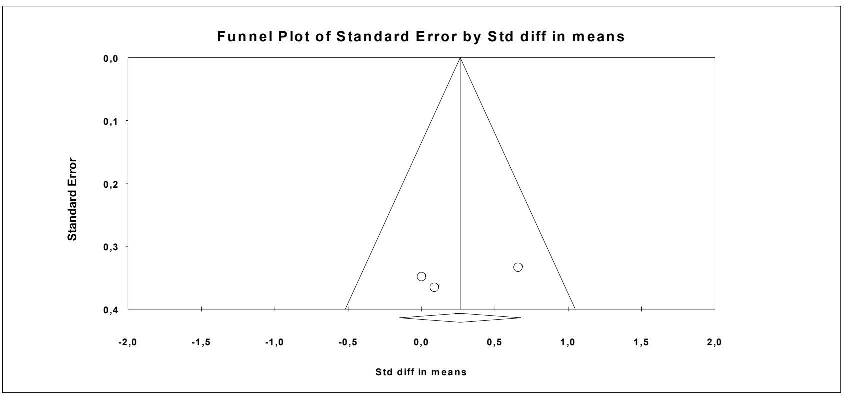

Figure 19. Funnel plot of the cephalometric variable UFH.

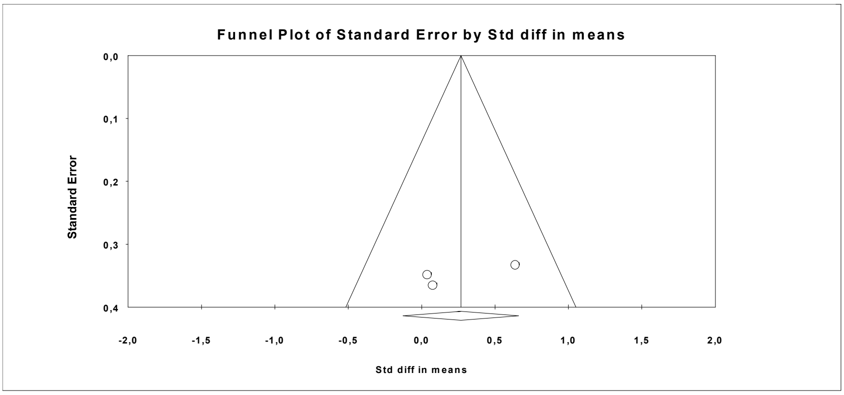

Figure 20. Funnel plot of the cephalometric variable LAFH.

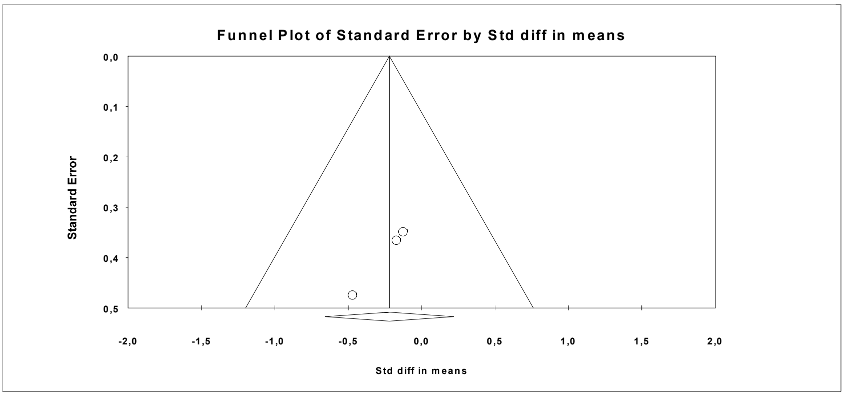

Figure 21. Funnel plot of the cephalometric variable Co-Go.

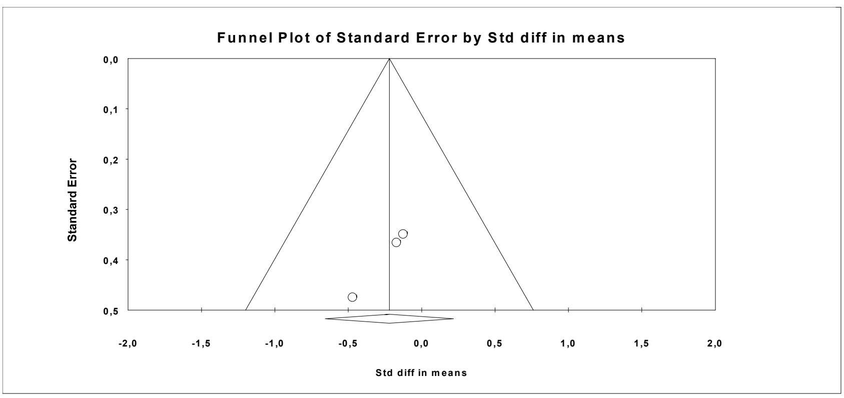

Figure 22. Funnel plot of the cephalometric variable Overjet.

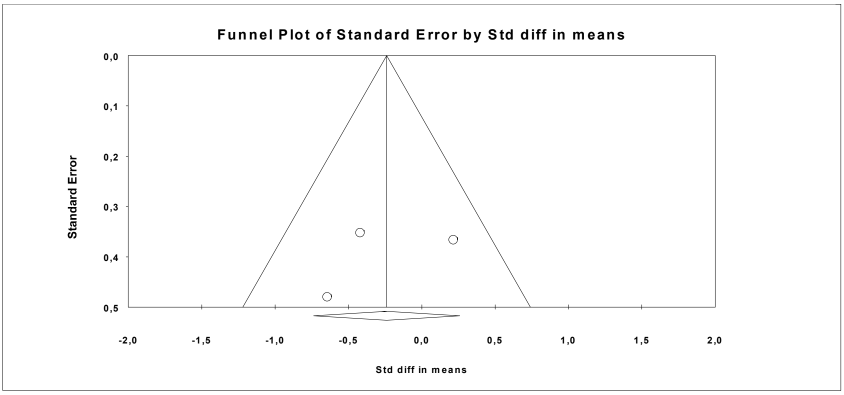

Figure 23. Funnel plot of the cephalometric variable Overbite.

5.6. Cephalometric changes

Meta-analysis were performed for the variables SNA, SNB, ANB, Wits appraisal, SN-ML and Gonial angle, where 5 or more treated groups participated in the assessment of the results (Table XI). For the rest of the variables, namely Co-Gn, N-Me, UFH, LAFH, Co-Go, Overjet and Overbite, where 4 or less treated groups were included (Table XI), exploratory analyses were performed instead and the relevant results are regarded as secondary and must be viewed with caution.

The summary of pooled estimates of all cephalometric variables under investigation performed with the random effects model is presented in Table XI and the corresponding forest plots in Figures 24-36. Meta-analysis of the original data revealed a statistically significant decrease for the variables SNB (Figure 25) and Gonial angle (Figure 30), and a statistically significant increase for the variables ANB (Figure 26), Wits appraisal (Figure 27), and SN-ML (Figure 29). In addition, exploratory analysis revealed a statistically significant increase for the variables N-Me (Figure 31) and Overjet (Figure 35).

Subgroup meta analysis using as modifier the applied “force level” (i.e., light or heavy) (Table XII), indicated significant decrease for the variables SNB (Figure 25a) and Gonial angle (Figure 30a) and significant increase for the variable ANB (Figure 26a), while subgroup exploratory analysis with the same modifier showed less reliable decrease for the variable N-Me (Figure 31a), and increase for the variable Overjet (Figure 35a).

Respectively, subgroup meta-analysis examining the variables in relation with the modifier “additional appliances used” (Table XIII) indicated statistically significant decrease for the variables SNB (Figure 25b) and Gonial angle (Figure 30b) and statistically significant increase for the variables ANB (Figure 26b) and Wits appraisal (Figure 27b). Exploratory analysis with the same modifier revealed less reliable decrease for the variables SN-ML (Figure 29b) and N-Me (Figure 31b) and increase for the variable Overjet ((Figure 35b).

More specifically the detailed results of the statistical elaboration are presented below.

Table XI. Summary of pooled estimates for the meta-analyses performed with the random effects model, and analysis of heterogeneity.

Variables	***N of treated groups***	***Kind of analysis performed, meta analysis (MA) or exploratory analysis (EA)***	***Heterogeneity***			***Test of null Z (P-value)***	***Standard Difference in Means***	***95% Confidence Intervals***	***Significance (P-value)***
			I^2	***Q (P-value)***	T^2				
SNA (º)	7	MA	69.362%	19.584 (0.003)	0.338	-0.070 (0.944)	-0.019	-0.538 to 0.501	0.944
SNB (º)	7	MA	90.800%	65.214 (0.000)	2.029	-3.429, (0.001)***	-1.966	-3.089 to -0.842	0.001***
ANB (º)	7	MA	89.614%,	57.767 (0.000)	2.029	4.311 (0.000)***	2.484	1.355 to 3.613	0.000***
Wits appraisal (mm)	5	MA	95.645%	91.844 (0.000)	6.364	3.083, (0.002)**	3.622	1.319 to 5.925	0.002**
Co-Gn (mm)	4	EA	11.164%	3.377 (0.337)	0.018	-1.483, (0.138)	-0.293	-0.681 to 0.094	0.138
SN-ML (º)	5	MA	70.569%	13.591 (0.009)	0.437	3.321 (0.001)***	1.171	0.480 to 1.863	0.001***
Gonial angle (º)	6	MA	78.915%	23.714 (0.000)	0.636	-2.164, (0.030)*	-0.799	-1.523 to 0.075	0.030*
N-Me (mm)	3	EA	65.651%	5.823 (0.054)	0.321	3.428, (0.001)***	1.387	0.594 to 2.180	0.001***

UFH (mm)	3	EA	9.200%	2.203 (0.332)	0.012	1.251, (0.211)	0.264	-0.150 to 0.678	0.211
LFAH (mm)	3	EA	0%	1.944 (0.378)	0	1.328, (0.184)	0.267	-0.127 to 0.661	0.184
Co-Go (mm)	3	EA	0%	0.369 (0.832)	0	-0.987, 0.324	-0.220	-0.657 to 0.217	0.324
Overjet (mm)	3	EA	84.912%	13.256 (0.001)	1.548	3.291 (0.001)***	2.623	1.061 to 4.185	0.001***
Overbite (mm)	3	EA	20.552%	2.517 (0.284)	0.040	-0.943 (0.345)	-0.239	-0.737 to 0.258	0.345

**statistically significant at P*≤0.05, ** statistically significant at *P*≤0.01, ***statistically significant at *P*≤0.001

Table XII. Summary of pooled estimates for the meta-analyses performed with the random effects model for subgroups classified according to force level, and analysis of heterogeneity.

Variables	***Subgroup: force level***	***N of treated groups***	***Heterogeneity Q (between subgroups) (P-value)***	***Test of null Z (P-value)***	***Standard Difference in Means***	***95% Confidence Intervals***	***Significance (P-value)***
SNA (º)	heavy	2	0.171 (0.679)	0.446 (0.655)	0.133	-0.453 to 0.720	0.655
	light	4		1.478 (0.140)	0.280	-0.091 to 0.651	0.140
	overall	6		1.487 (0.137)	0.238	-0.076 to 0.552	0.137
SNB (º)	heavy	2	0.806 (0.369)	-1.446 (0.148)	-2.570	-6.053 to 0.913	0.148
	light	4		-2.508 (0.012)*	-0.939	-1.674 to -0.205	0.012*
	overall	6		-2.752 (0.006)**	-1.009	-1.727 to -0.290	0.006**
ANB (º)	heavy	2	0.464 (0.496)	2.223 (0.026)*	2.669	0.316 to 5.023	0.026*
	light	4		2.983 (0.003)**	1.758	0.603 to 2.913	0.003**
	overall	6		3.657 (0.000)***	1.935	0.898 to 2.972	0.000***
Wits appraisal (mm)	heavy	1	6.265 (0.012)	6.493 (0.000)***	0.710	3.220 to 6.004	0.000***
	light	3		1.710 (0.087)	0.955	-0.239 to 3.505	0.087
	overall	4		6.230 (0.000)***	3.551	2.434 to 4.668	0.000***

Co-Gn (mm)	heavy	1	0.863 (0.353)	-1.476 (0.140)	-0.711	-1.656 to 0.233	0.140
	light	3		-0.976 (0.329)	-0.218	-0.655 to 0.219	0.329
	overall	4		-1.506 (0.132)	-0.305	-0.702 to 0.092	0.132
SN-ML (^{o})	heavy	2	0.339 (0.560)	1.329 (0.184)	0.113	-0.528 to 2.755	0.184
	light	2		5.705 (0.000)***	1.629	1.069 to 2.188	0.000***
	overall	4		5.829 (0.000)***	1.575	1.045 to 2.105	0.000***
Gonial angle (^{o})	heavy	2	0.015 (0.903)	-1.371 (0.170)	-0.414	-1.006 to 0.178	0.170
	light	3		-1.726 (0.084)	-0.369	-0.788 to 0.050	0.084
	overall	5		-2.201 (0.028)	-0.384	-0.726 to -0.042	0.028
N-Me (mm)	heavy	1	0.362 (0.547)	3.836 (0.000)***	1.710	0.836 to 2.584	0.000***
	light	2		2.091 (0.037)*	1.259	0.079 to 2.440	0.037*
	overall	3		4.327 (0.000)***	1.551	0.848 to 2.253	0.000***
UFH (mm)	light	3	***Not applicable***				
	overall	3					

LFAH (mm)	light	3	***Not applicable***				
	overall	3					
Co-Go (mm)	heavy	1	0.361 (0.548)	-0.994 (0.320)	-0.472	-1.402 to 0.459	0.320
	light	2		-0.590 (0.555)	-0.149	-0.644 to 0.346	0.555
	overall	3		-0.987 (0.324)	-0.220	-0.657 to 0.217	0.324
Overjet (mm)	heavy	1	11.818 (0.001)	5.467 (0.000)***	5.097	3.269 to 6.924	0.000***
	light	2		5.332 (0.000)***	1.708	1.080 to 2.335	0.000***
	overall	3		6.819 (0.000)***	2.065	1.472 to 2.659	0.000***
Overbite (mm)	heavy	1	0.861 (0.353)	-1.345 (0.179)	-0.645	-1.585 to 0.295	0.179
	light	2		-0.350 (0.727)	-0.111	-0.734 to 0.512	0.727
	overall	3		-1.034 (0.301)	-0.274	-0.793 to 0.245	0.301

**statistically significant at* $P\leq0.05$, ** statistically significant at $P\leq0.01$, ***statistically significant at $P\leq0.001$

Table XIII. Summary of pooled estimates for the meta-analyses performed with the random effects model for subgroups classified according to additional appliances used, and analysis of heterogeneity.

Variables	*Subgroup: additional appliance*	*N of treated groups*	*Heterogeneity Q (between subgroups) (P-value)*	*Test of null Z (P-value)*	*Standard Difference in Means*	*95% Confidence Intervals*	*Significance (P-value)*
SNA (º)	Bite plane	3	4.010 (0.135)	2.262 (0.024)*	0.482	0.899 to 2.262	0.024*
	Nothing	3		-0.997 (0.319)	0.538	0.520 to -0.997	0.319
	Quadhelix	1		-0.149 (0.881)	-0.055	0.661 to -0.149	0.881
	Overall	7		1.457 (0.145)	0.254	0.595 to 1.457	0.145
SNB (º)	Bite plane	3	0.789 (0.674)	-1.951 (0.051)	-2.030	-4.070 to 0.010	0.051
	Nothing	3		-1.818 (0.069)	-2.272	-4.722 to 0.178	0.069
	Quadhelix	1		-3.347 (0.001)***	-1.345	-2.133 to -0.557	0.001***
	Overall	7		-4.186 (0.000)***	-1.503	-2.207 to -0.799	0.000***
ANB (º)	Bite plane	3	9.894 (0.007)	4.823 (0.000)***	3.164	1.878 to 4.450	0.000***
	Nothing	3		2.062 (0.039)*	2.427	4.734 to 2.062	0.039*
	Quadhelix	1		2.255 (0.024)*	0.858	1.604 to 2.255	0.024*
	Overall	7		4.765 (0.000)***	1.511	2.132 to 4.765	0.000***

Wits appraisal (mm)	Bite plane	2	43.551 (0.000)	9.079 (0.000)***	0.490	3.489 to 5.410	0.000***
	Nothing	2		1.085 (0.278)	4.523	-3.956 to 13.775	0.278
	Quadhelix	1		1.144 (0.253)	0.369	-0.301 to 1.145	0.253
	Overall	5		6.431 (0.000)***	0.294	1.315 to 2.468	0.000***
Co-Gn (mm)	Bite plane	1	2.332 (0.312)	0.179 (0.858)	0.058	-0.579 to 0.695	0.858
	Nothing	2		-1.103 (0.270)	-0.312	-0.865 to 0.242	0.270
	Quadhelix	1		-1.857 (0.063)	-0.697	-1.433 to 0.039	0.063
	Overall	4		-1.539 (0.124)	-0.285	-0.649 to 0.078	0.124
SN-ML (^{o})	Bite plane	3	13.053 (0.000)	7.071 (0.000)***	1.717	1.241 to 2.193	0.000***
	Nothing	2		1.010 (0.313)	0.308	-0.290 to 0.906	0.313
	Overall	5		6.162 (0.000)***	1.171	0.798 to 1.543	0.000***
Gonial angle (^{o})	Bite plane	2	1.302 (0.522)	-1.227 (0.220)	-0.338	-0.879 to 0.202	0.220
	Nothing	3		-1.424 (0.154)	-1.252	-2.974 to 0.471	0.154
	Quadhelix	1		-1.810 (0.070)	-0.679	-1.413 to 0.056	0.070
	Overall	6		-2.347 (0.019)*	-0.505	-0.927 to -0.083	0.019*

N-Me (mm)	Bite plane	3	***Not applicable***				
	Overall	3					
UFH (mm)	Bite plane	1	2.203 (0.332)	1.973 (0.049)*	0.658	0.004 to 1,312	0.049*
	Nothing	1		0.000 (1.000)	0.000	-0.683 to 0.683	1.000
	Quadhelix	1		0.239 (0.811)	0.087	-0.629 to 0.803	0.811
	Overall	3		1.321 (0.186)	0.266	-0.128 to 0.660	0.186
LFAH (mm)	Bite plane	1	1.944 (0.378)	1.911 (0.056)	0.636	-0.016 to 1.289	0.056
	Nothing	1		0.106 (0.916)	0.037	-0.646 to 0.720	0.916
	Quadhelix	1		0.204 (0.838)	0.075	-0.641 to 0.791	0.838
	Overall	3		1.328 (0.184)	0.267	-0.127 to 0.661	0.184
Co-Go (mm)	Nothing	2	0.027 (0.870)	-0.882 (0.378)	-0.248	-0.799 to 0.303	0.378
	Quadhelix	1		-0.472 (0.637)	-0.173	-0.890 to 0.544	0.637
	Overall	3		0.987 (0.324)	-0.220	-0.657 to 0.217	0.324

Overjet (mm)	Nothing	2	0.336 (0.562)	1.724 (0.085)	3.162	-0.433 to 6.757	0.085
	Quadhelix	1		4.623 (0.000)***	2.068	1.191 to 2.945	0.000***
	Overall	3		4.900 (0.000)***	2.130	1.278 to 2.982	0.000***
Overbite (mm)	Nothing	2	2.376 (0.123)	-1.759 (0.079)	-0.500	-1.056 to 0.057	0.079
	Quadhelix	1		0.586 (0.558)	0.215	-0.503 to 0.932	0.558
	Overall	3		-1.031 (0.302)	-0.231	-0.671 to 0.208	0.302

**statistically significant at* $P \leq 0.05$, ** statistically significant at $P \leq 0.01$, ***statistically significant at $P \leq 0.001$

5.6.1. Skeletal variables in the sagittal plane

SNA (°)

Meta-analysis performed for this variable, based on 7 treated groups, revealed no statistically significant results (SDM=-0.019, CI= -0.538 to 0.501, *P*-value=0.944) (Table XI) (Figure 24). The heterogeneity of the included studies was between moderate and high (I^2=69.362%, Q=19.584, *P*-value=0.003), and the between studies variance was moderate (T^2=0.338) (Table XI). Subgroup analysis performed according to the categorical modifier "force level" (heavy or light force) and, based on 6 treated groups, showed no particular tendency for SNA (Table XII) (Figure 24a). The same resulted from subgroup analysis with the modifier "additional appliances used" based on 7 treated groups (Table XIII) (Figure 24b).

Model	Study name	Sample size: Control	Sample size: Treated	Statistics for each study: Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed): Weight (Fixed)	Relative weight	Weight (Random): Weight (Random)	Relative weight
	Abdelnaby and Nassar, 2010a	10	20	0,220	0,388	0,151	-0,541	0,981	0,567	0,571	6,63	14,04	2,04	14,37
	Abdelnaby and Nassar, 2010b	10	20	0,420	0,391	0,153	-0,346	1,187	1,075	0,282	6,54	13,84	2,04	14,31
	Altug et al., 1989	9	21	-1,667	0,453	0,205	-2,554	-0,779	-3,681	0,000	4,88	10,33	1,84	12,94
	Barrett et al., 2010a	20	14	0,000	0,348	0,121	-0,683	0,683	0,000	1,000	8,24	17,44	2,17	15,29
	Barrett et al., 2010b	20	12	-0,055	0,365	0,133	-0,770	0,661	-0,149	0,881	7,50	15,88	2,12	14,90
	Gökalp and Kurt , 2005	7	13	0,007	0,469	0,220	-0,912	0,926	0,016	0,988	4,55	9,63	1,79	12,60
	Tuncer et al., 2009	18	20	0,722	0,335	0,112	0,064	1,379	2,152	0,031	8,90	18,84	2,22	15,60
Fixed				0,045	0,146	0,021	-0,240	0,330	0,309	0,757				
Random				-0,019	0,265	0,070	-0,538	0,501	-0,070	0,944				

Std diff in means and 95% CI

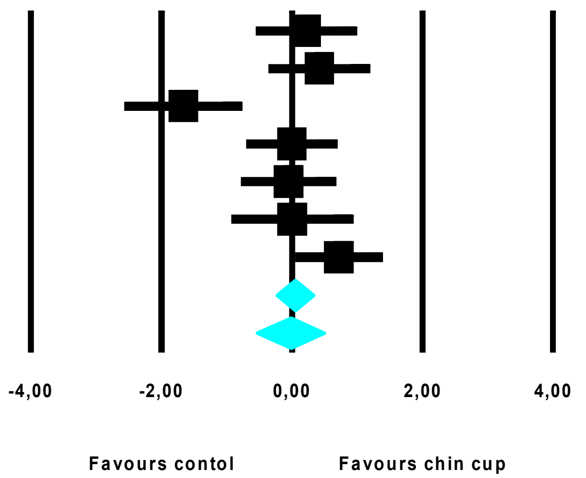

Figure 1. Forest plot of the cephalometric variable SNA.

Model	Group by Force level	Study name	Sample size: Control	Sample size: Treated	Statistics for each study: Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed): Weight (Fixed)	Relative weight	Weight (Separate tau): Weight (Separate tau)	Relative weight
	heavy	Abdelnaby and Nassar, 2010a	10	20	0,220	0,388	0,151	-0,541	0,981	0,567	0,571	6,63	59,31	6,63	59,31
	heavy	Gökalp and Kurt , 2005	7	13	0,007	0,469	0,220	-0,912	0,926	0,016	0,988	4,55	40,69	4,55	40,69
Fixed	heavy				0,133	0,299	0,089	-0,453	0,720	0,446	0,655				
Random	heavy				0,133	0,299	0,089	-0,453	0,720	0,446	0,655				
	light	Abdelnaby and Nassar, 2010b	10	20	0,420	0,391	0,153	-0,346	1,187	1,075	0,282	6,54	20,98	5,96	21,36
	light	Barrett et al., 2010a	20	14	0,000	0,348	0,121	-0,683	0,683	0,000	1,000	8,24	26,42	7,33	26,30
	light	Barrett et al., 2010b	20	12	-0,055	0,365	0,133	-0,770	0,661	-0,149	0,881	7,50	24,06	6,74	24,18
	light	Tuncer et al., 2009	18	20	0,722	0,335	0,112	0,064	1,379	2,152	0,031	8,90	28,54	7,85	28,16
Fixed	light				0,281	0,179	0,032	-0,070	0,632	1,569	0,117				
Random	light				0,280	0,189	0,036	-0,091	0,651	1,478	0,140				
Fixed	Overall				0,242	0,154	0,024	-0,059	0,543	1,575	0,115				
Random	Overall				0,238	0,160	0,026	-0,076	0,552	1,487	0,137				

Std diff in means and 95% CI

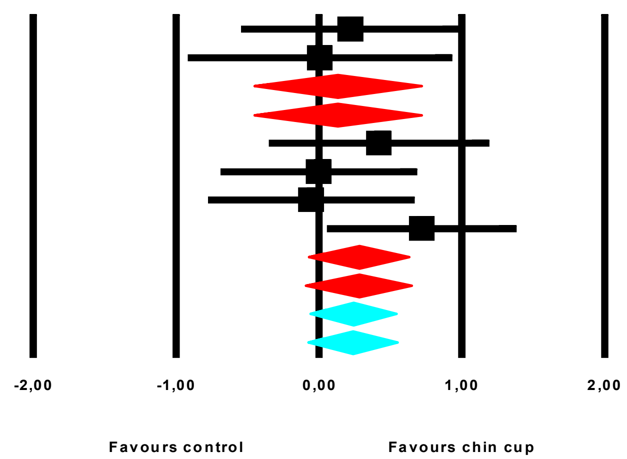

Figure 24a. Subgroup analysis of cephalometric variable SNA with regard to the factor force level.

Model	Group by Additional appliance	Study name	Sample size: Control	Sample size: Treated	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed): Weight (Fixed)	Weight (Fixed): Relative weight	Weight (Separate tau): Weight (Separate tau)	Weight (Separate tau): Relative weight
	bite plane	Abdelnaby and Nassar, 2010a	10	20	0,220	0,388	0,151	-0,541	0,981	0,567	0,571	6,63	30,05	6,63	30,05
	bite plane	Abdelnaby and Nassar, 2010b	10	20	0,420	0,391	0,153	-0,346	1,187	1,075	0,282	6,54	29,63	6,54	29,63
	bite plane	Tuncer et al., 2009	18	20	0,722	0,335	0,112	0,064	1,379	2,152	0,031	8,90	40,32	8,90	40,32
Fixed	bite plane				0,482	0,213	0,045	0,064	0,899	2,262	0,024				
Random	bite plane				0,482	0,213	0,045	0,064	0,899	2,262	0,024				
	nothing	Altug et al., 1989	9	21	-1,667	0,453	0,205	-2,554	-0,779	-3,681	0,000	4,88	27,61	1,11	32,40
	nothing	Barrett et al., 2010a	20	14	0,000	0,348	0,121	-0,683	0,683	0,000	1,000	8,24	46,63	1,23	35,72
	nothing	Gökalp and Kurt , 2005	7	13	0,007	0,469	0,220	-0,912	0,926	0,016	0,988	4,55	25,76	1,10	31,88
Fixed	nothing				-0,458	0,238	0,057	-0,925	0,008	-1,926	0,054				
Random	nothing				-0,538	0,540	0,291	-1,595	0,520	-0,997	0,319				
	quadhelix	Barrett et al., 2010b	20	12	-0,055	0,365	0,133	-0,770	0,661	-0,149	0,881	7,50	100,00	7,50	100,00
Fixed	quadhelix				-0,055	0,365	0,133	-0,770	0,661	-0,149	0,881				
Random	quadhelix				-0,055	0,365	0,133	-0,770	0,661	-0,149	0,881				
Fixed	Overall				0,045	0,146	0,021	-0,240	0,330	0,309	0,757				
Random	Overall				0,254	0,174	0,030	-0,088	0,595	1,457	0,145				

Std diff in means and 95% CI

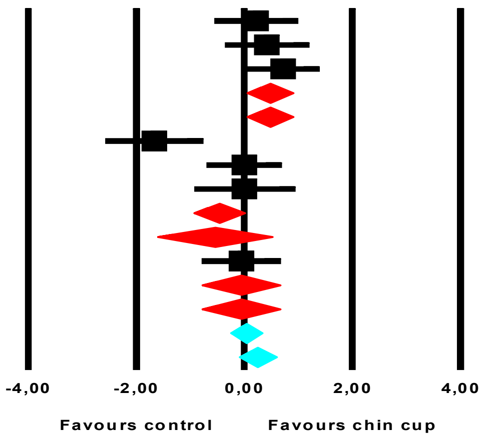

Figure 24b. Subgroup analysis of cephalometric variable SNA with regard to the factor additional appliance used.

SNB (^{o})

Following meta-analysis performed on 7 treated groups, it was revealed that there was statistically significant reduction of the SNB angle due to chin cup use (SDM=-1.966, CI=-3.089 to -0.842, *P*-value=0.001) (Table XI) (Figure 25). The observed heterogeneity was high (I^2=90.800%, *Q*=65.214, *P*-value=0.000) (Table XI) and the between studies variance was accordingly high (T^2=2.029) (Table XI). Subgroup analysis performed on 6 treated groups with regard to the factor "force level" indicated a trend of SNB to reduce further with heavier forces (Table XII) (Figure 25a), while the same variable, when examined in a subgroup analysis based on the modifier "additional appliances used", in which 7 treated groups participated revealed no specific trend (Table XIII) (Figure 25b).

Model	Study name	Sample size		Statistics for each study							Weight (Fixed)		Weight (Random)	
		Treated	Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Random)	Relative weight
	Abdelnaby and Nassar, 2010a	20	10	-4,381	0,685	0,470	-5,724	-3,037	-6,391	0,000	2,13	6,03	0,40	13,15
	Abdelnaby and Nassar, 2010b	20	10	-0,142	0,388	0,150	-0,902	0,618	-0,366	0,715	6,65	18,84	0,46	15,08
	Altug et al., 1989	21	9	-5,880	0,857	0,735	-7,560	-4,200	-6,859	0,000	1,36	3,85	0,36	11,89
	Barrett et al., 2010a	14	20	-0,509	0,354	0,125	-1,203	0,185	-1,438	0,150	7,98	22,61	0,46	15,26
	Barrett et al., 2010b	12	20	-1,345	0,402	0,162	-2,133	-0,557	-3,347	0,001	6,19	17,52	0,46	15,00
	Gökalp and Kurt , 2005	13	7	-0,826	0,487	0,237	-1,779	0,128	-1,697	0,090	4,22	11,96	0,44	14,51
	Tuncer et al., 2009	20	18	-1,785	0,384	0,147	-2,538	-1,033	-4,649	0,000	6,78	19,20	0,46	15,10
Fixed				-1,310	0,168	0,028	-1,639	-0,980	-7,782	0,000				
Random				-1,966	0,573	0,329	-3,089	-0,842	-3,429	0,001				

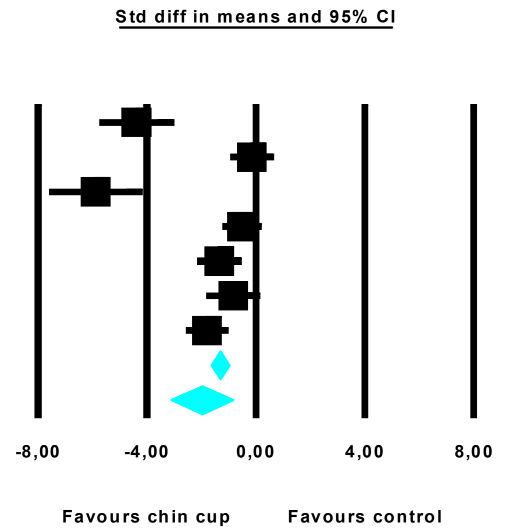

Figure 25. Forest plot of the cephalometric variable SNB.

Model	Group by Force level	Study name	Sample size		Statistics for each study							Weight (Fixed)		Weight (Separate tau)	
			Control	Treated	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Separate tau)	Relative weight
	heavy	Abdelnaby and Nassar, 2010a	10	20	-4,381	0,685	0,470	-5,724	-3,037	-6,391	0,000	2,13	33,51	0,16	49,08
	heavy	Gökalp and Kurt , 2005	7	13	-0,826	0,487	0,237	-1,779	0,128	-1,697	0,090	4,22	66,49	0,16	50,92
Fixed	heavy				-2,017	0,397	0,157	-2,795	-1,239	-5,083	0,000				
Random	heavy				-2,570	1,777	3,158	-6,053	0,913	-1,446	0,148				
	light	Abdelnaby and Nassar, 2010b	10	20	-0,142	0,388	0,150	-0,902	0,618	-0,366	0,715	6,65	24,10	1,77	24,80
	light	Barrett et al., 2010a	20	14	-0,509	0,354	0,125	-1,203	0,185	-1,438	0,150	7,98	28,93	1,85	25,95
	light	Barrett et al., 2010b	20	12	-1,345	0,402	0,162	-2,133	-0,557	-3,347	0,001	6,19	22,42	1,73	24,32
	light	Tuncer et al., 2009	18	20	-1,785	0,384	0,147	-2,538	-1,033	-4,649	0,000	6,78	24,56	1,78	24,93
Fixed	light				-0,921	0,190	0,036	-1,295	-0,548	-4,841	0,000				
Random	light				-0,939	0,375	0,140	-1,674	-0,205	-2,508	0,012				
Fixed	Overall				-1,126	0,172	0,029	-1,463	-0,790	-6,564	0,000				
Random	Overall				-1,009	0,367	0,134	-1,727	-0,290	-2,752	0,006				

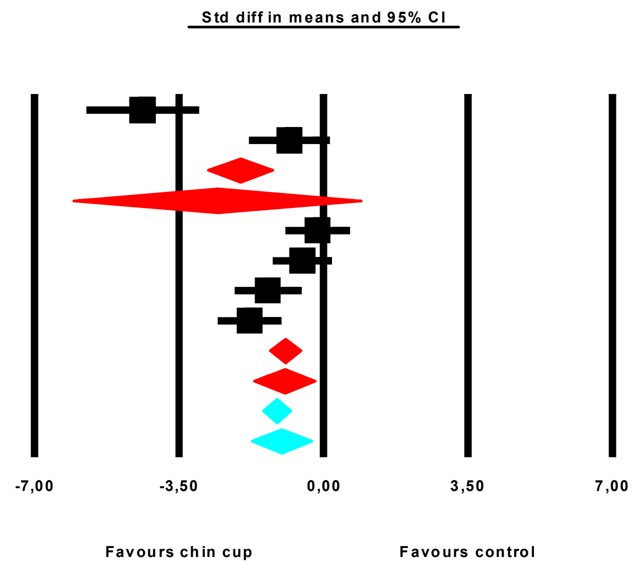

Figure 25a. Subgroup analysis of cephalometric variable SNB with regard to the factor force level.

Model	Group by Additional appliance	Study name	Sample size: Treated	Sample size: Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Separate tau)	Relative weight
	bite plane	Abdelnaby and Nassar, 2010a	20	10	-4,381	0,685	0,470	-5,724	-3,037	-6,391	0,000	2,13	13,68	0,29	31,21
	bite plane	Abdelnaby and Nassar, 2010b	20	10	-0,142	0,388	0,150	-0,902	0,618	-0,366	0,715	6,65	42,75	0,32	34,38
	bite plane	Tuncer et al., 2009	20	18	-1,785	0,384	0,147	-2,538	-1,033	-4,649	0,000	6,78	43,57	0,32	34,41
Fixed	bite plane				-1,438	0,254	0,064	-1,935	-0,941	-5,671	0,000				
Random	bite plane				-2,030	1,041	1,083	-4,070	0,010	-1,951	0,051				
	nothing	Altug et al., 1989	21	9	-5,880	0,857	0,735	-7,560	-4,200	-6,859	0,000	1,36	10,03	0,20	30,81
	nothing	Barrett et al., 2010a	14	20	-0,509	0,354	0,125	-1,203	0,185	-1,438	0,150	7,98	58,85	0,22	35,02
	nothing	Gökalp and Kurt , 2005	13	7	-0,826	0,487	0,237	-1,779	0,128	-1,697	0,090	4,22	31,12	0,22	34,17
Fixed	nothing				-1,146	0,271	0,074	-1,678	-0,614	-4,222	0,000				
Random	nothing				-2,272	1,250	1,562	-4,722	0,178	-1,818	0,069				
	quadhelix	Barrett et al., 2010b	12	20	-1,345	0,402	0,162	-2,133	-0,557	-3,347	0,001	6,19	100,00	6,19	100,00
Fixed	quadhelix				-1,345	0,402	0,162	-2,133	-0,557	-3,347	0,001				
Random	quadhelix				-1,345	0,402	0,162	-2,133	-0,557	-3,347	0,001				
Fixed	Overall				-1,310	0,168	0,028	-1,639	-0,980	-7,782	0,000				
Random	Overall				-1,503	0,359	0,129	-2,207	-0,799	-4,186	0,000				

Std diff in means and 95% CI

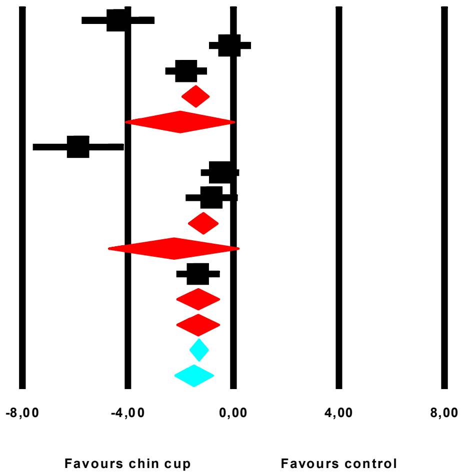

Figure 25b. Subgroup analysis of cephalometric variable SNB with regard to the factor additional appliance used.

ANB (°)

There was a statistical increase of the ANB angle following chin cup use, as revealed from the meta-analysis performed on 7 treated groups (SDM=2.484, CI=1.355 to 3.613, *P*-value=0.000) (Table XI) (Figure 26). The heterogeneity of the included studies was high (I^2=89.614%, *Q*=57.767, *P*-value=0.000) followed by an also high between studies variance (T^2=2.029) (Table XI). Subgroup analysis based on the force magnitude included 6 treated groups indicated a trend of ANB to increase further with heavier forces (Table XII) (Figure 26a). The same variable when examined in a subgroup analysis comprised of 7 treated groups according to the factor “additional appliances used” showed a slight tendency of ANB towards grater increase with the simultaneous use of bite planes, less when no additional appliances were used and even lesser with quad-helices (Table XIII) (Figure 26b).

Model	Study name	Sample size		Statistics for each study							Weight (Fixed)		Weight (Random)	
		Treated	Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Random)	Relative weight
	Abdelnaby and Nassar, 2010a	20	10	3,897	0,635	0,403	2,652	5,141	6,138	0,000	2,48	7,97	0,41	13,65
	Abdelnaby and Nassar, 2010b	20	10	3,751	0,620	0,385	2,536	4,966	6,049	0,000	2,60	8,35	0,41	13,75
	Altug et al., 1989	21	9	5,427	0,806	0,650	3,847	7,007	6,733	0,000	1,54	4,94	0,37	12,39
	Barrett et al., 2010a	14	20	0,677	0,358	0,128	-0,025	1,378	1,890	0,059	7,80	25,06	0,46	15,39
	Barrett et al., 2010b	12	20	0,858	0,381	0,145	0,112	1,604	2,255	0,024	6,90	22,17	0,46	15,27
	Gökalp and Kurt , 2005	13	7	1,495	0,525	0,276	0,466	2,524	2,847	0,004	3,63	11,65	0,43	14,40
	Tuncer et al., 2009	20	18	2,067	0,402	0,162	1,279	2,855	5,139	0,000	6,18	19,85	0,46	15,15
Fixed				1,836	0,179	0,032	1,485	2,188	10,248	0,000				
Random				2,484	0,576	0,332	1,355	3,613	4,311	0,000				

Std diff in means and 95% CI

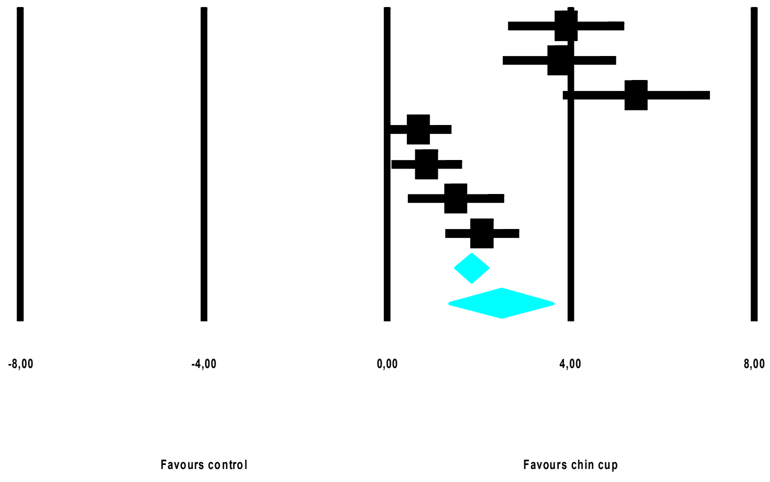

Figure 26. Forest plot of the cephalometric variable ANB.

Model	Group by Force level	Study name	Sample size		Statistics for each study							Weight (Fixed)		Weight (Separate tau)	
			Control	Treated	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Separate tau)	Relative weight
	heavy	Abdelnaby and Nassar, 2010a	10	20	3,897	0,635	0,403	2,652	5,141	6,138	0,000	2,48	40,61	0,34	48,90
	heavy	Gökalp and Kurt , 2005	7	13	1,495	0,525	0,276	0,466	2,524	2,847	0,004	3,63	59,39	0,35	51,10
Fixed	heavy				2,470	0,405	0,164	1,677	3,263	6,106	0,000				
Random	heavy				2,669	1,201	1,442	0,316	5,023	2,223	0,026				
	light	Abdelnaby and Nassar, 2010b	10	20	3,751	0,620	0,385	2,536	4,966	6,049	0,000	2,60	11,07	0,63	22,04
	light	Barrett et al., 2010a	20	14	0,677	0,358	0,128	-0,025	1,378	1,890	0,059	7,80	33,22	0,76	26,32
	light	Barrett et al., 2010b	20	12	0,858	0,381	0,145	0,112	1,604	2,255	0,024	6,90	29,39	0,75	25,99
	light	Tuncer et al., 2009	18	20	2,067	0,402	0,162	1,279	2,855	5,139	0,000	6,18	26,32	0,74	25,66
Fixed	light				1,436	0,206	0,043	1,032	1,841	6,961	0,000				
Random	light				1,758	0,589	0,347	0,603	2,913	2,983	0,003				
Fixed	Overall				1,650	0,184	0,034	1,289	2,010	8,975	0,000				
Random	Overall				1,935	0,529	0,280	0,898	2,972	3,657	0,000				

Std diff in means and 95% CI

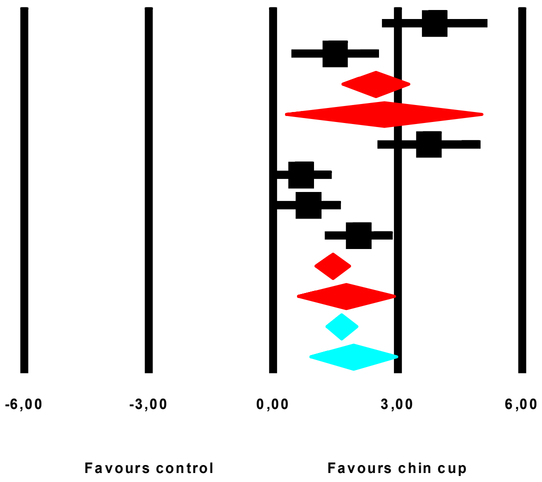

Figure 26a. Subgroup analysis of cephalometric variable ANB with regard to the factor force level.

Model	Group by Additional appliances	Study name	Sample size: Control	Sample size: Treated	Statistics for each study: Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed): Weight (Fixed)	Relative weight	Weight (Separate tau): Weight (Separate tau)	Relative weight
	bite plane	Abdelnaby and Nassar, 2010a	10	20	3,897	0,635	0,403	2,652	5,141	6,138	0,000	2,48	22,03	0,72	31,02
	bite plane	Abdelnaby and Nassar, 2010b	10	20	3,751	0,620	0,385	2,536	4,966	6,049	0,000	2,60	23,09	0,73	31,44
	bite plane	Tuncer et al., 2009	18	20	2,067	0,402	0,162	1,279	2,855	5,139	0,000	6,18	54,88	0,87	37,55
Fixed	bite plane				2,859	0,298	0,089	2,275	3,443	9,595	0,000				
Random	bite plane				3,164	0,656	0,430	1,878	4,450	4,823	0,000				
	nothing	Altug et al., 1989	9	21	5,427	0,806	0,650	3,847	7,007	6,733	0,000	1,54	11,87	0,22	31,02
	nothing	Barrett et al., 2010a	20	14	0,677	0,358	0,128	-0,025	1,378	1,890	0,059	7,80	60,16	0,25	35,12
	nothing	Gökalp and Kurt , 2005	7	13	1,495	0,525	0,276	0,466	2,524	2,847	0,004	3,63	27,97	0,24	33,86
Fixed	nothing				1,469	0,278	0,077	0,925	2,013	5,291	0,000				
Random	nothing				2,427	1,177	1,386	0,120	4,734	2,062	0,039				
	quadhelix	Barrett et al., 2010b	20	12	0,858	0,381	0,145	0,112	1,604	2,255	0,024	6,90	100,00	6,90	100,00
Fixed	quadhelix				0,858	0,381	0,145	0,112	1,604	2,255	0,024				
Random	quadhelix				0,858	0,381	0,145	0,112	1,604	2,255	0,024				
Fixed	Overall				1,836	0,179	0,032	1,485	2,188	10,248	0,000				
Random	Overall				1,511	0,317	0,101	0,889	2,132	4,765	0,000				

Std diff in means and 95% CI

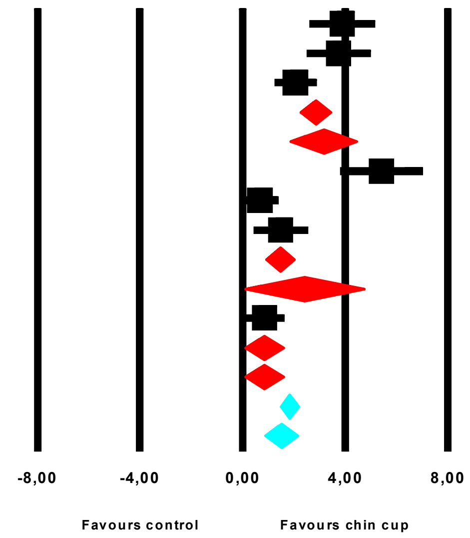

Figure 26b. Subgroup analysis of cephalometric variable ANB with regard to the factor additional appliance used.

Wits appraisal (mm)

The increase observed in the Wits appraisal following chin cup use is supported by statistical significance, as resulted from the current meta-analysis, in which 5 treated groups participated (SDM=3.622, CI=1.319 to 5.924, *P*-value=0.002) (Table XI) (Figure 27). The accompanying heterogeneity of the included studies was very high (I^2=95.645%, *Q*=91.844, *P*-value=0.000) while the between studies variance was also very high (T^2=6.364) (Table XI). Exploratory subgroup analysis comprised of 4 treated groups was performed examining the variable in relation with the applied force level and revealed a trend of Wits appraisal to increase further with heavier forces (Table XII) (Figure 27a). Subgroup meta-analysis performed on 5 treated groups and examining Wits appraisal according to the modifier "additional appliances used" showed a trend towards further increase of the variable with the concomitant use of bite planes, whereas quad-helices or absence of additional appliances did not seem to have any contribution at all (Table XIII) (Figure 27b).

Model	Study name	Sample size		Statistics for each study							Weight (Fixed)		Weight (Random)	
		Treated	Control	Standard error	Std diff in means	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Random)	Relative weight
	Abdelnaby and Nassar, 2010a	20	10	0,710	4,612	0,504	3,220	6,004	6,493	0,000	1,98	9,85	0,15	20,09
	Abdelnaby and Nassar, 2010b	20	10	0,677	4,301	0,458	2,974	5,628	6,353	0,000	2,18	10,84	0,15	20,23
	Altug et al., 1989	21	9	1,292	9,519	1,669	6,987	12,051	7,368	0,000	0,60	2,98	0,12	17,18
	Barrett et al., 2010a	14	20	0,353	0,471	0,125	-0,222	1,163	1,333	0,183	8,02	39,84	0,15	21,27
	Barrett et al., 2010b	12	20	0,369	0,422	0,136	-0,301	1,145	1,144	0,253	7,35	36,50	0,15	21,23
Fixed				0,223	1,545	0,050	1,108	1,982	6,932	0,000				
Random				1,175	3,622	1,380	1,319	5,924	3,083	0,002				

Std diff in means and 95% CI

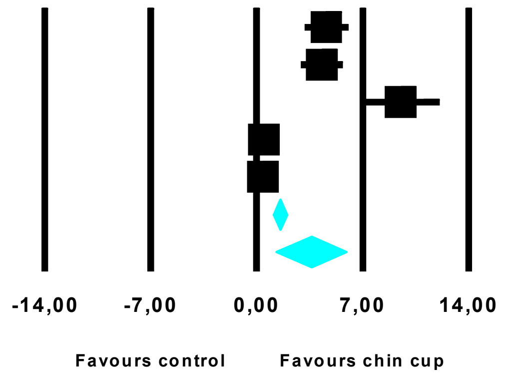

Figure 27. Forest plot of the cephalometric variable Wits appraisal.

Model	Group by Force level	Study name	Sample size		Statistics for each study							Weight (Fixed)		Weight (Separate tau)	
			Control	Treated	Standard error	Std diff in means	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Separate tau)	Relative weight
	heavy	Abdelnaby and Nassar, 2010a	10	20	0,710	4,612	0,504	3,220	6,004	6,493	0,000	1,98	100,00	1,98	100,00
Fixed	heavy				0,710	4,612	0,504	3,220	6,004	6,493	0,000				
Random	heavy				0,710	4,612	0,504	3,220	6,004	6,493	0,000				
	light	Abdelnaby and Nassar, 2010b	10	20	0,677	4,301	0,458	2,974	5,628	6,353	0,000	2,18	12,43	0,34	30,78
	light	Barrett et al., 2010a	20	14	0,353	0,471	0,125	-0,222	1,163	1,333	0,183	8,02	45,70	0,38	34,69
	light	Barrett et al., 2010b	20	12	0,369	0,422	0,136	-0,301	1,145	1,144	0,253	7,35	41,87	0,38	34,54
Fixed	light				0,239	0,926	0,057	0,459	1,394	3,881	0,000				
Random	light				0,955	1,633	0,912	-0,239	3,505	1,710	0,087				
Fixed	Overall				0,226	1,300	0,051	0,857	1,744	5,747	0,000				
Random	Overall				0,570	3,551	0,325	2,434	4,668	6,230	0,000				

Std diff in means and 95% CI

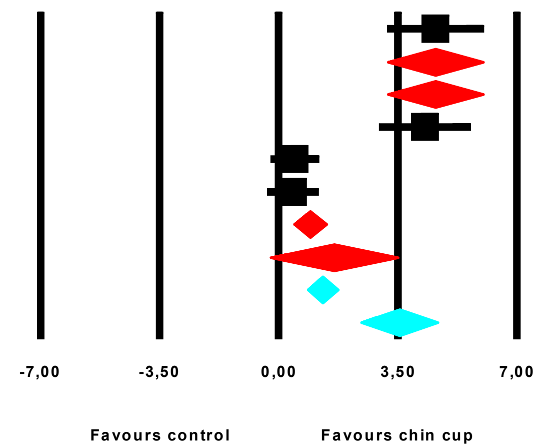

Figure 27a. Subgroup analysis of cephalometric variable Wits appraisal with regard to the factor force level.

Model	Group by Additional appliance	Study name	Sample size: Treated	Sample size: Control	Standard error	Std diff in means	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight (Fixed)	Weight (Separate tau)	Relative weight (Separate tau)
	bite plane	Abdelnaby and Nassar, 2010a	20	10	0,710	4,612	0,504	3,220	6,004	6,493	0,000	1,98	47,61	1,98	47,61
	bite plane	Abdelnaby and Nassar, 2010b	20	10	0,677	4,301	0,458	2,974	5,628	6,353	0,000	2,18	52,39	2,18	52,39
Fixed	bite plane				0,490	4,449	0,240	3,489	5,410	9,079	0,000				
Random	bite plane				0,490	4,449	0,240	3,489	5,410	9,079	0,000				
	nothing	Altug et al., 1989	21	9	1,292	9,519	1,669	6,987	12,051	7,368	0,000	0,60	6,95	0,02	49,06
	nothing	Barrett et al., 2010a	14	20	0,353	0,471	0,125	-0,222	1,163	1,333	0,183	8,02	93,05	0,02	50,94
Fixed	nothing				0,341	1,100	0,116	0,432	1,767	3,228	0,001				
Random	nothing				4,523	4,909	20,460	-3,956	13,775	1,085	0,278				
	quadhelix	Barrett et al., 2010b	12	20	0,369	0,422	0,136	-0,301	1,145	1,144	0,253	7,35	100,00	7,35	100,00
Fixed	quadhelix				0,369	0,422	0,136	-0,301	1,145	1,144	0,253				
Random	quadhelix				0,369	0,422	0,136	-0,301	1,145	1,144	0,253				
Fixed	Overall				0,223	1,545	0,050	1,108	1,982	6,932	0,000				
Random	Overall				0,294	1,892	0,087	1,315	2,468	6,431	0,000				

Std diff in means and 95% CI

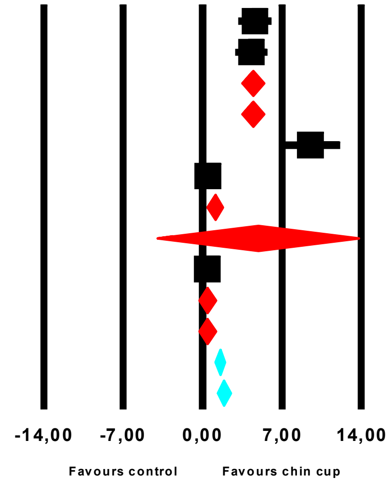

Figure 27b. Subgroup analysis of cephalometric variable Wits appraisal with regard to the factor additional appliance used.

Co-Gn (mm)

According to the results of the explanatory analysis performed on 4 included treated groups no statistical significance of the slight decrease observed of the variable Co-Gn was found due to chin cup wear, (SDM=-0.293, CI=-0.681 to 0.094, *P*-value=0.138) (Table XI) (Figure 28). Both the heterogeneity of the included studies (I^2=11.164%, Q=3.377, *P*-value=0.337) and the between studies variance were low (T^2=0.018) (Table XI). Exploratory subgroup analysis did not reveal any significance in the trends correlating Co-Gn with force magnitude (Table XII) (Figure 28a). The same was observed when additional appliances were considered (Table XIII) (Figure 28b).

Model	Study name	Sample size		Statistics for each study							Weight (Fixed)		Weight (Random)	
		Treated	Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Random)	Relative weight
	Barrett et al., 2010a	14	20	-0,103	0,349	0,122	-0,786	0,581	-0,294	0,769	8,22	28,27	7,18	28,05
	Barrett et al., 2010b	12	20	-0,697	0,375	0,141	-1,433	0,039	-1,857	0,063	7,10	24,39	6,30	24,63
	Gökalp and Kurt , 2005	13	7	-0,711	0,482	0,232	-1,656	0,233	-1,476	0,140	4,30	14,79	4,00	15,62
	Tuncer et al., 2009	20	18	0,058	0,325	0,106	-0,579	0,695	0,179	0,858	9,47	32,55	8,11	31,69
Fixed				-0,285	0,185	0,034	-0,649	0,078	-1,539	0,124				
Random				-0,293	0,198	0,039	-0,681	0,094	-1,483	0,138				

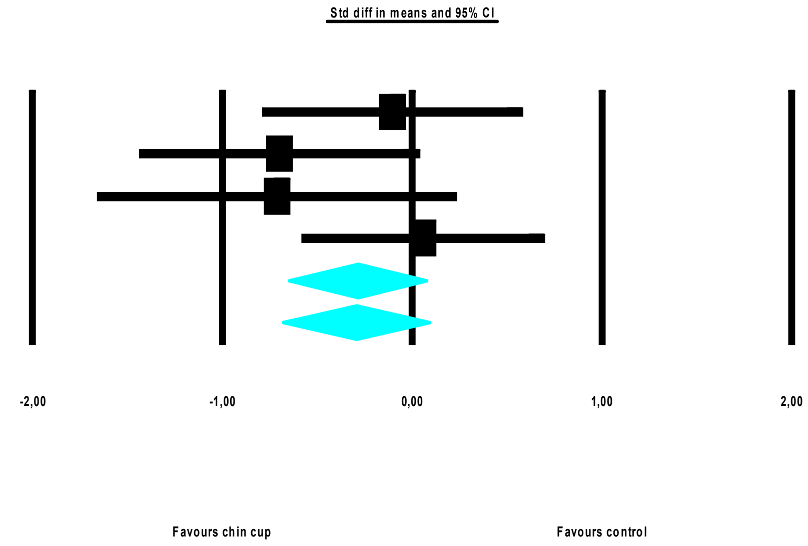

Figure 28. Forest plot of the cephalometric variable Co-Gn.

Model	Group by Force level	Study name	Sample size		Statistics for each study							Weight (Fixed)		Weight (Separate tau)	
			Treated	Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Separate tau)	Relative weight
	heavy	Gökalp and Kurt , 2005	13	7	-0,711	0,482	0,232	-1,656	0,233	-1,476	0,140	4,30	100,00	4,30	100,00
Fixed	heavy				-0,711	0,482	0,232	-1,656	0,233	-1,476	0,140				
Random	heavy				-0,711	0,482	0,232	-1,656	0,233	-1,476	0,140				
	light	Barrett et al., 2010a	14	20	-0,103	0,349	0,122	-0,786	0,581	-0,294	0,769	8,22	33,18	6,68	33,28
	light	Barrett et al., 2010b	12	20	-0,697	0,375	0,141	-1,433	0,039	-1,857	0,063	7,10	28,62	5,92	29,47
	light	Tuncer et al., 2009	20	18	0,058	0,325	0,106	-0,579	0,695	0,179	0,858	9,47	38,20	7,48	37,26
Fixed	light				-0,211	0,201	0,040	-0,605	0,182	-1,052	0,293				
Random	light				-0,218	0,223	0,050	-0,655	0,219	-0,976	0,329				
Fixed	Overall				-0,285	0,185	0,034	-0,649	0,078	-1,539	0,124				
Random	Overall				-0,305	0,202	0,041	-0,702	0,092	-1,506	0,132				

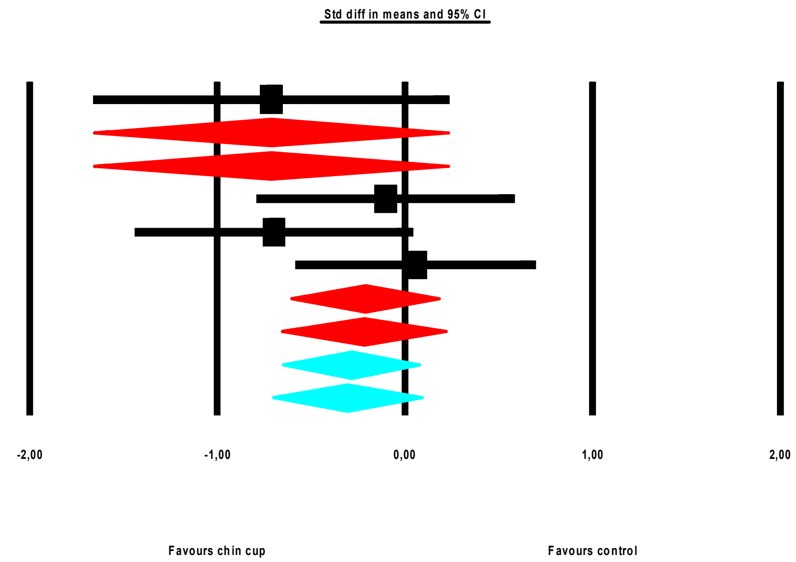

Figure 28a. Subgroup analysis of cephalometric variable Co-Gn with regard to the factor force level.

Model	Group by Additional appliance	Study name	Sample size: Treated	Sample size: Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Separate tau)	Relative weight
	bite plane	Tuncer et al., 2009	20	18	0,058	0,325	0,106	-0,579	0,695	0,179	0,858	9,47	100,00	9,47	100,00
Fixed	bite plane				0,058	0,325	0,106	-0,579	0,695	0,179	0,858				
Random	bite plane				0,058	0,325	0,106	-0,579	0,695	0,179	0,858				
	nothing	Barrett et al., 2010a	14	20	-0,103	0,349	0,122	-0,786	0,581	-0,294	0,769	8,22	65,66	7,69	64,95
	nothing	Gökalp and Kurt , 2005	13	7	-0,711	0,482	0,232	-1,656	0,233	-1,476	0,140	4,30	34,34	4,15	35,05
Fixed	nothing				-0,312	0,283	0,080	-0,865	0,242	-1,103	0,270				
Random	nothing				-0,316	0,291	0,084	-0,885	0,253	-1,088	0,277				
	quadhelix	Barrett et al., 2010b	12	20	-0,697	0,375	0,141	-1,433	0,039	-1,857	0,063	7,10	100,00	7,10	100,00
Fixed	quadhelix				-0,697	0,375	0,141	-1,433	0,039	-1,857	0,063				
Random	quadhelix				-0,697	0,375	0,141	-1,433	0,039	-1,857	0,063				
Fixed	Overall				-0,285	0,185	0,034	-0,649	0,078	-1,539	0,124				
Random	Overall				-0,286	0,188	0,035	-0,654	0,081	-1,527	0,127				

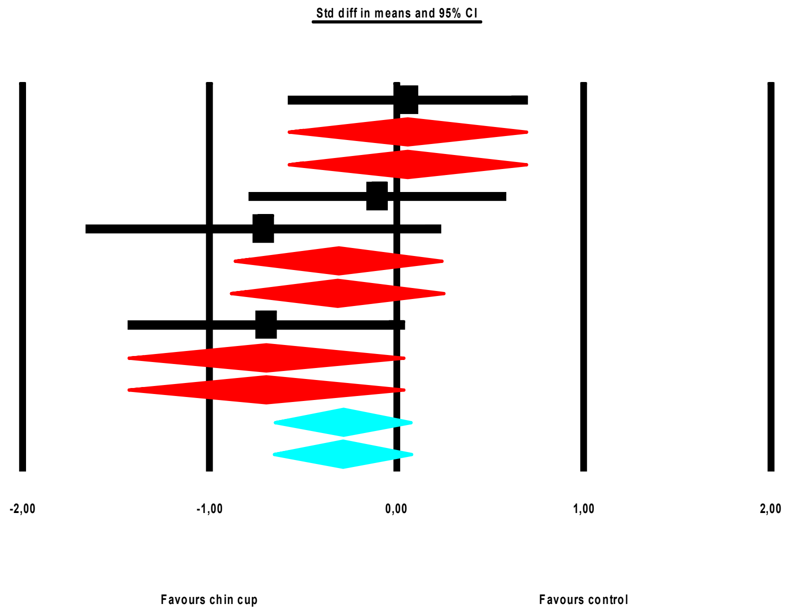

Figure 28b. Subgroup analysis of cephalometric variable Co-Gn with regard to the factor additional appliance used.

5.6.2. Skeletal variables in the vertical plane

SN-ML (^{o})

The results of the meta-analysis performed on 5 included treated groups was that SN-ML angle increased following chin cup use with statistical significance (SDM=1.171, CI=0.480 to 1.863, *P*-value=0.001) (Table XI) (Figure 29). The observed heterogeneity of the included studies was moderate to high (I^2=70.569%, *Q*=13.591, *P*-value=0.009), and the between studies variance was moderate as well (T^2=0.437) (Table XI). Exploratory subgroup analysis, in which 4 treated groups participated, indicated a tendency of light forces to allow for more increase of the SN-ML angle in comparison with heavier ones (Table XII) (Figure 29a). Concomitant use of bite planes, as derived from the second subgroup meta analysis, based on 5 treated groups, also seemed to permit further increase of SN-ML, when compared with no simultaneous use of any additional appliances at all (Table XIII) (Figure 29b).

Model	Study name	Sample size: Treated	Sample size: Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight (Fixed)	Weight (Random)	Relative weight (Random)
	Abdelnaby and Nassar, 2010a	20	10	1,948	0,462	0,213	1,043	2,854	4,219	0,000	4,69	16,93	1,54	19,13
	Abdelnaby and Nassar, 2010b	20	10	1,777	0,450	0,203	0,895	2,659	3,947	0,000	4,94	17,82	1,56	19,45
	Altug et al., 1989	21	9	0,333	0,401	0,161	-0,452	1,119	0,832	0,406	6,23	22,48	1,67	20,82
	Gökalp and Kurt , 2005	13	7	0,273	0,471	0,222	-0,650	1,196	0,580	0,562	4,51	16,29	1,52	18,89
	Tuncer et al., 2009	20	18	1,529	0,369	0,136	0,805	2,253	4,141	0,000	7,34	26,48	1,74	21,70
Fixed				1,171	0,190	0,036	0,798	1,543	6,162	0,000				
Random				1,171	0,353	0,124	0,480	1,863	3,321	0,001				

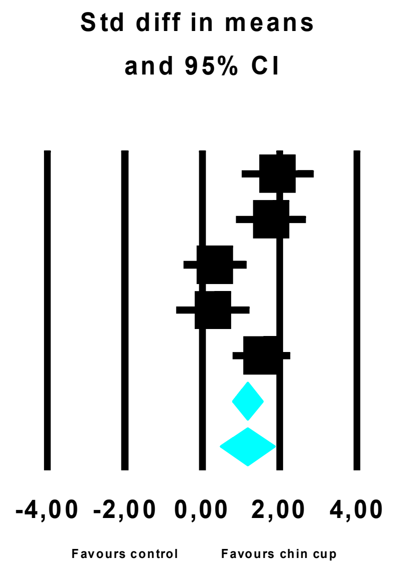

Figure 29. Forest plot of the cephalometric variable SN-ML.

Model	Group by Force level	Study name	Sample size: Control	Sample size: Treated	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight (Fixed)	Weight (Separate tau)	Relative weight (Separate tau)
	heavy	Abdelnaby and Nassar, 2010a	10	20	1,948	0,462	0,213	1,043	2,854	4,219	0,000	4,69	50,96	0,71	50,15
	heavy	Gökalp and Kurt , 2005	7	13	0,273	0,471	0,222	-0,650	1,196	0,580	0,562	4,51	49,04	0,71	49,85
Fixed	heavy				1,127	0,330	0,109	0,481	1,773	3,418	0,001				
Random	heavy				1,113	0,838	0,702	-0,528	2,755	1,329	0,184				
	light	Abdelnaby and Nassar, 2010b	10	20	1,777	0,450	0,203	0,895	2,659	3,947	0,000	4,94	40,22	4,94	40,22
	light	Tuncer et al., 2009	18	20	1,529	0,369	0,136	0,805	2,253	4,141	0,000	7,34	59,78	7,34	59,78
Fixed	light				1,629	0,285	0,081	1,069	2,188	5,705	0,000				
Random	light				1,629	0,285	0,081	1,069	2,188	5,705	0,000				
Fixed	Overall				1,414	0,216	0,047	0,991	1,837	6,551	0,000				
Random	Overall				1,575	0,270	0,073	1,045	2,105	5,829	0,000				

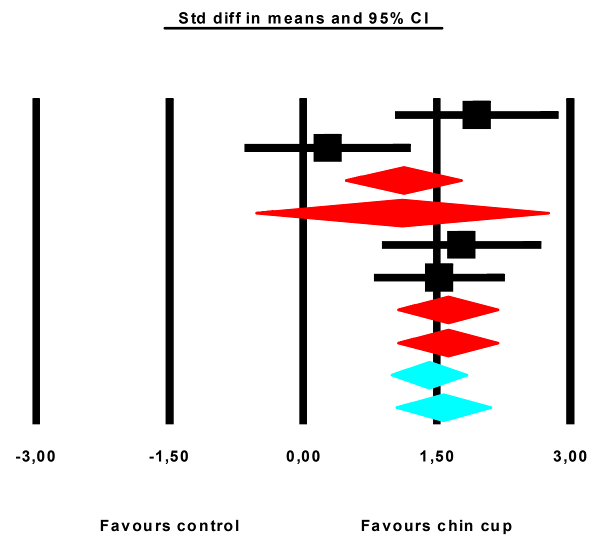

Figure 29a. Subgroup analysis of cephalometric variable SN-ML with regard to the factor force level.

Model	Group by Additional appliance	Study name	Sample size: Treated	Sample size: Control	Statistics for each study: Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed): Weight (Fixed)	Relative weight	Weight (Separate tau): Weight (Separate tau)	Relative weight
	bite plane	Abdelnaby and Nassar, 2010a	20	10	1,948	0,462	0,213	1,043	2,854	4,219	0,000	4,69	27,65	4,69	27,65
	bite plane	Abdelnaby and Nassar, 2010b	20	10	1,777	0,450	0,203	0,895	2,659	3,947	0,000	4,94	29,10	4,94	29,10
	bite plane	Tuncer et al., 2009	20	18	1,529	0,369	0,136	0,805	2,253	4,141	0,000	7,34	43,25	7,34	43,25
Fixed	bite plane				1,717	0,243	0,059	1,241	2,193	7,071	0,000				
Random	bite plane				1,717	0,243	0,059	1,241	2,193	7,071	0,000				
	nothing	Altug et al., 1989	21	9	0,333	0,401	0,161	-0,452	1,119	0,832	0,406	6,23	57,99	6,23	57,99
	nothing	Gökalp and Kurt , 2005	13	7	0,273	0,471	0,222	-0,650	1,196	0,580	0,562	4,51	42,01	4,51	42,01
Fixed	nothing				0,308	0,305	0,093	-0,290	0,906	1,010	0,313				
Random	nothing				0,308	0,305	0,093	-0,290	0,906	1,010	0,313				
Fixed	Overall				1,171	0,190	0,036	0,798	1,543	6,162	0,000				
Random	Overall				1,171	0,190	0,036	0,798	1,543	6,162	0,000				

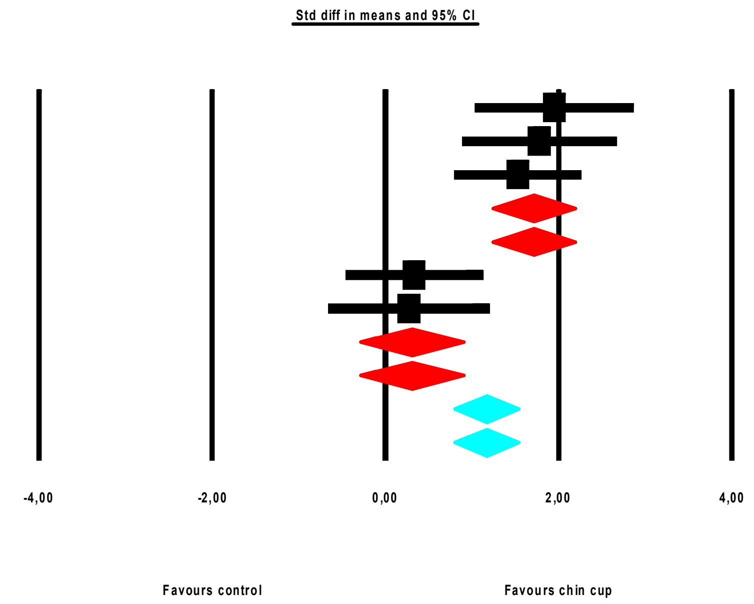

Figure 29b. Subgroup analysis of cephalometric variable SN-ML with regard to the factor additional appliance used.

Gonial angle (^{o})

Statistically significant decrease of the Gonial angle, attributed to chin cup wear, was found following the meta analysis performed on 6 included treated groups (SDM=-0.799, CI=-1.523 to -0.075, *P*-value=0.030) (Table XI) (Figure 30). This result was accompanied, however, by high heterogeneity of the included studies (I^2=78.915%, Q=23.714, *P*-value=0.000) and moderate between studies variance (T^2=0.636) (Table XI). Subgroup analysis of the same variable based on 5 treated groups examining the correlation of the variable with the factor “force level” revealed no particular trend (Table XII) (Figure 30a), while to the same result concluded the subgroup analysis performed, based on 6 treated groups, examining the variable according to the factor “additional appliances used” (Table XIII) (Figure 30b).

Model	Study name	Sample size: Control	Sample size: Trated	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed): Weight (Fixed)	Weight (Fixed): Relative weight	Weight (Random): Weight (Random)	Weight (Random): Relative weight
	Abdelnaby and Nassar, 2010a	10	20	-0,271	0,389	0,151	-1,033	0,491	-0,697	0,486	6,61	18,42	1,27	17,31
	Abdelnaby and Nassar, 2010b	10	20	-0,406	0,391	0,153	-1,172	0,360	-1,039	0,299	6,55	18,23	1,27	17,28
	Altug et al., 1989	9	21	-3,189	0,573	0,328	-4,312	-2,066	-5,566	0,000	3,05	8,49	1,04	14,13
	Barrett et al., 2010a	20	14	-0,071	0,349	0,122	-0,755	0,612	-0,205	0,838	8,23	22,92	1,32	17,99
	Barrett et al., 2010b	20	12	-0,679	0,375	0,141	-1,413	0,056	-1,810	0,070	7,12	19,82	1,29	17,55
	Gökalp and Kurt , 2005	7	13	-0,631	0,479	0,230	-1,570	0,309	-1,316	0,188	4,35	12,12	1,15	15,74
Fixed				-0,622	0,167	0,028	-0,949	-0,295	-3,727	0,000				
Random				-0,799	0,369	0,136	-1,523	-0,075	-2,164	0,030				

Std diff in means and 95% CI

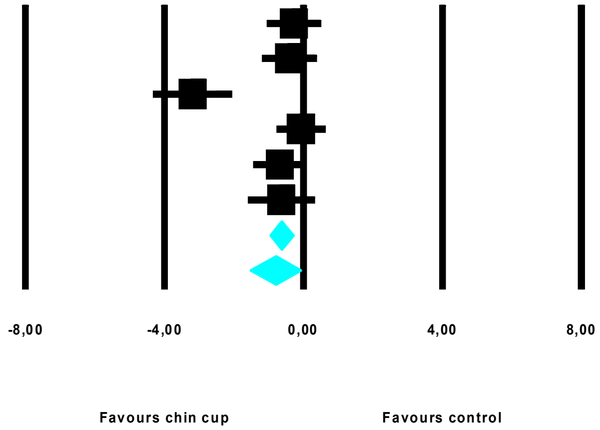

Figure 30. Forest plot of the cephalometric variable Gonial angle.

Model	Group by Force level	Study name	Sample size: Control	Sample size: Trated	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed): Weight (Fixed)	Weight (Fixed): Relative weight	Weight (Separate tau): Weight (Separate tau)	Weight (Separate tau): Relative weight
	heavy	Abdelnaby and Nassar, 2010a	10	20	-0,271	0,389	0,151	-1,033	0,491	-0,697	0,486	6,61	60,30	6,61	60,30
	heavy	Gökalp and Kurt , 2005	7	13	-0,631	0,479	0,230	-1,570	0,309	-1,316	0,188	4,35	39,70	4,35	39,70
Fixed	heavy				-0,414	0,302	0,091	-1,006	0,178	-1,371	0,170				
Random	heavy				-0,414	0,302	0,091	-1,006	0,178	-1,371	0,170				
	light	Abdelnaby and Nassar, 2010b	10	20	-0,406	0,391	0,153	-1,172	0,360	-1,039	0,299	6,55	29,90	6,55	29,90
	light	Barrett et al., 2010a	20	14	-0,071	0,349	0,122	-0,755	0,612	-0,205	0,838	8,23	37,59	8,23	37,59
	light	Barrett et al., 2010b	20	12	-0,679	0,375	0,141	-1,413	0,056	-1,810	0,070	7,12	32,50	7,12	32,50
Fixed	light				-0,369	0,214	0,046	-0,788	0,050	-1,726	0,084				
Random	light				-0,369	0,214	0,046	-0,788	0,050	-1,726	0,084				
Fixed	Overall				-0,384	0,174	0,030	-0,726	-0,042	-2,201	0,028				
Random	Overall				-0,384	0,174	0,030	-0,726	-0,042	-2,201	0,028				

Std diff in means and 95% CI

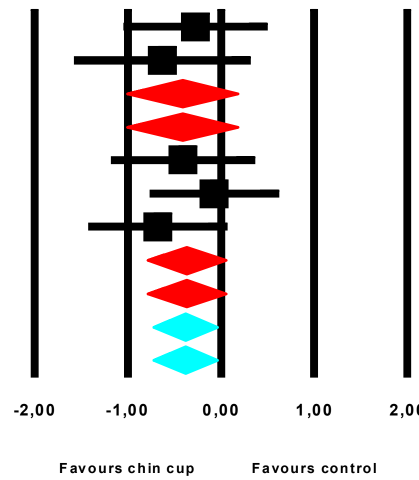

Figure 30a. Subgroup analysis of cephalometric variable Gonial angle with regard to the factor force level.

Model	Group by Additional appliances	Study name	Sample size: Control	Sample size: Trated	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Separate tau)	Relative weight
	bite plane	Abdelnaby and Nassar, 2010a	10	20	-0,271	0,389	0,151	-1,033	0,491	-0,697	0,486	6,61	50,25	6,61	50,25
	bite plane	Abdelnaby and Nassar, 2010b	10	20	-0,406	0,391	0,153	-1,172	0,360	-1,039	0,299	6,55	49,75	6,55	49,75
Fixed	bite plane				-0,338	0,276	0,076	-0,879	0,202	-1,227	0,220				
Random	bite plane				-0,338	0,276	0,076	-0,879	0,202	-1,227	0,220				
	nothing	Altug et al., 1989	9	21	-3,189	0,573	0,328	-4,312	-2,066	-5,566	0,000	3,05	19,49	0,41	31,89
	nothing	Barrett et al., 2010a	20	14	-0,071	0,349	0,122	-0,755	0,612	-0,205	0,838	8,23	52,66	0,45	34,87
	nothing	Gökalp and Kurt , 2005	7	13	-0,631	0,479	0,230	-1,570	0,309	-1,316	0,188	4,35	27,85	0,43	33,24
Fixed	nothing				-0,835	0,253	0,064	-1,331	-0,339	-3,301	0,001				
Random	nothing				-1,252	0,879	0,772	-2,974	0,471	-1,424	0,154				
	quadhelix	Barrett et al., 2010b	20	12	-0,679	0,375	0,141	-1,413	0,056	-1,810	0,070	7,12	100,00	7,12	100,00
Fixed	quadhelix				-0,679	0,375	0,141	-1,413	0,056	-1,810	0,070				
Random	quadhelix				-0,679	0,375	0,141	-1,413	0,056	-1,810	0,070				
Fixed	Overall				-0,622	0,167	0,028	-0,949	-0,295	-3,727	0,000				
Random	Overall				-0,505	0,215	0,046	-0,927	-0,083	-2,347	0,019				

Std diff in means and 95% CI

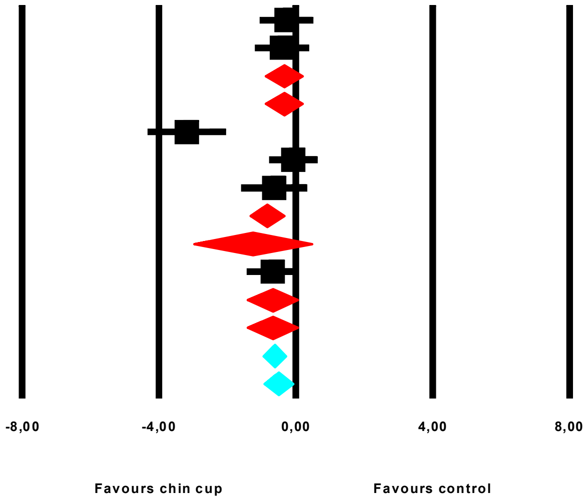

Figure 30b. Subgroup analysis of cephalometric variable Gonial angle with regard to the factor additional appliance used.

N-Me (mm)

Statistically significant increase was detected in the linear variable N-Me due to treatment with chin cup according to the exploratory analysis performed on 3 treated groups (SDM=1.387, CI=0.594 to 2.180, *P*-value=0.001) (Table XI) (Figure 31). Moderate heterogeneity of the included studies (I^2=65.651%, *Q*=5.823, *P*-value=0.054) and small between studies variance (T^2=0.321) were also observed (Table XI). The exploratory subgroup analysis with regard to the modifier "force level" indicated a trend of further increase of N-Me with heavier forces (Table XII) (Figure 31a), while the exploratory analysis of the same variable based on "additional appliances used" was not applicable since in all 3 included treated groups bite plane was present (Table XIII).

Model	Study name	Sample size: Treated	Control	Statistics for each study: Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed): Weight (Fixed)	Relative weight	Weight (Random): Weight (Random)	Relative weight
	Abdelnaby and Nassar, 2010a	20	10	1,710	0,446	0,199	0,836	2,584	3,836	0,000	5,03	26,88	1,92	31,46
	Abdelnaby and Nassar, 2010b	20	10	1,904	0,459	0,210	1,005	2,803	4,150	0,000	4,75	25,39	1,88	30,77
	Tuncer et al., 2009	20	18	0,696	0,335	0,112	0,041	1,352	2,081	0,037	8,93	47,73	2,31	37,77
Fixed				1,275	0,231	0,053	0,822	1,728	5,518	0,000				
Random				1,387	0,405	0,164	0,594	2,180	3,428	0,001				

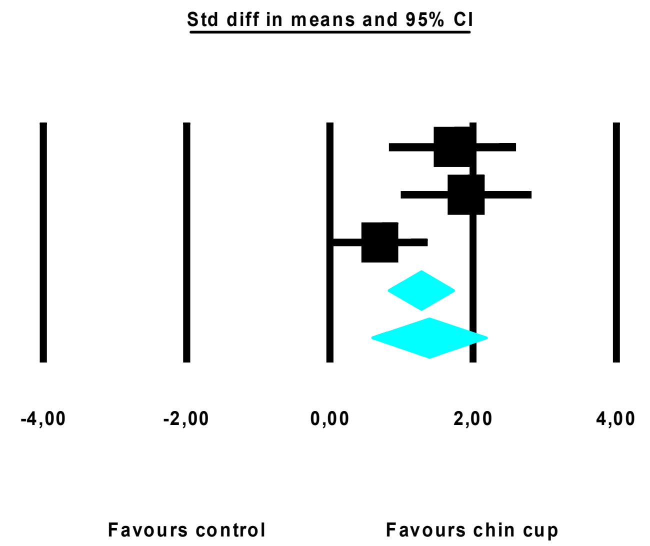

Figure 31. Forest plot of the cephalometric variable N-Me.

Model	Group by Force level	Study name	Sample size: Treated	Control	Statistics for each study: Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed): Weight (Fixed)	Relative weight	Weight (Separate tau): Weight (Separate tau)	Relative weight
	heavy	Abdelnaby and Nassar, 2010a	20	10	1,710	0,446	0,199	0,836	2,584	3,836	0,000	5,03	100,00	5,03	100,00
Fixed	heavy				1,710	0,446	0,199	0,836	2,584	3,836	0,000				
Random	heavy				1,710	0,446	0,199	0,836	2,584	3,836	0,000				
	light	Abdelnaby and Nassar, 2010b	20	10	1,904	0,459	0,210	1,005	2,803	4,150	0,000	4,75	34,73	1,29	46,62
	light	Tuncer et al., 2009	20	18	0,696	0,335	0,112	0,041	1,352	2,081	0,037	8,93	65,27	1,47	53,38
Fixed	light				1,116	0,270	0,073	0,586	1,645	4,127	0,000				
Random	light				1,259	0,602	0,363	0,079	2,440	2,091	0,037				
Fixed	Overall				1,275	0,231	0,053	0,822	1,728	5,518	0,000				
Random	Overall				1,551	0,358	0,128	0,848	2,253	4,327	0,000				

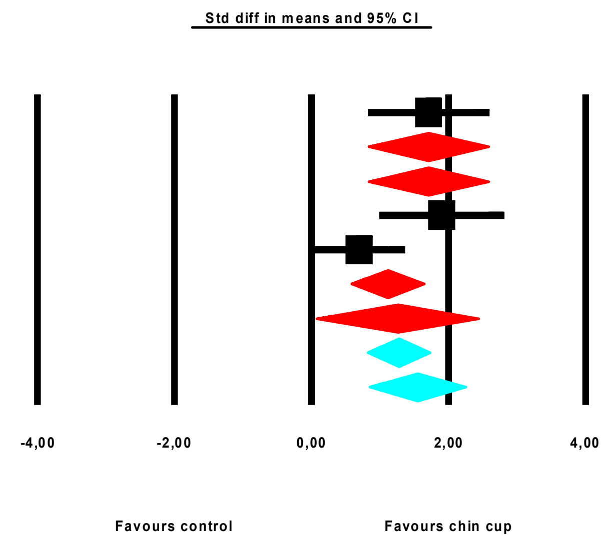

Figure 31a. Subgroup analysis of cephalometric variable N-Me with regard to the factor force level.

UFH (mm)

Exploratory analysis performed on 3 treated groups and examining the linear variable UFH in relation with the chin cup use concluded to no statistically significant results (SDM=0.264, CI=-0.150 to 0.678, *P*-value=0.211) (Table XI) (Figure 32). The heterogeneity of the included studies was low (I^2=9.200%, *Q*=2.203, *P*-value=0.332) and the between studies variance was also very low (T^2=0.012) (Table XI). Exploratory analysis based on "force level" could not be performed because in all 3 included treated groups a light force chin cup was applied (Table XII). From the exploratory analysis of the same variable, examined in relation to the factor "additional appliances used", no significant trends could be observed (Table XIII) (Figure 32b).

Model	Study name	Sample size: Treated	Control	Statistics for each study: Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed): Weight (Fixed)	Relative weight	Weight (Random): Weight (Random)	Relative weight
	Barrett et al., 2010a	14	20	0,000	0,348	0,121	-0,683	0,683	0,000	1,000	8,24	33,32	7,48	33,33
	Barrett et al., 2010b	12	20	0,087	0,365	0,133	-0,629	0,803	0,239	0,811	7,49	30,32	6,86	30,59
	Tuncer et al., 2009	20	18	0,658	0,334	0,111	0,004	1,312	1,973	0,049	8,99	36,37	8,09	36,08
Fixed				0,266	0,201	0,040	-0,128	0,660	1,321	0,186				
Random				0,264	0,211	0,045	-0,150	0,678	1,251	0,211				

Std diff in means and 95% CI

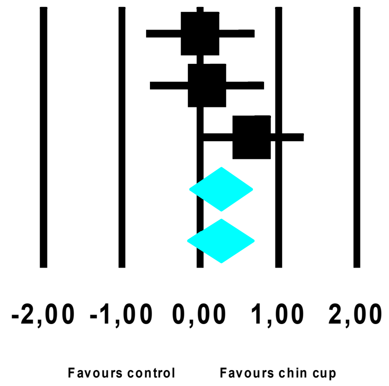

Figure 32. Forest plot of the cephalometric variable UFH.

Model	Group by Additional appliance	Study name	Sample size: Treated	Control	Statistics for each study: Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed): Weight (Fixed)	Relative weight	Weight (Separate tau): Weight (Separate tau)	Relative weight
	bite plane	Tuncer et al., 2009	20	18	0,658	0,334	0,111	0,004	1,312	1,973	0,049	8,99	100,00	8,99	100,00
Fixed	bite plane				0,658	0,334	0,111	0,004	1,312	1,973	0,049				
Random	bite plane				0,658	0,334	0,111	0,004	1,312	1,973	0,049				
	nothing	Barrett et al., 2010a	14	20	0,000	0,348	0,121	-0,683	0,683	0,000	1,000	8,24	100,00	8,24	100,00
Fixed	nothing				0,000	0,348	0,121	-0,683	0,683	0,000	1,000				
Random	nothing				0,000	0,348	0,121	-0,683	0,683	0,000	1,000				
	quadhelix	Barrett et al., 2010b	12	20	0,087	0,365	0,133	-0,629	0,803	0,239	0,811	7,49	100,00	7,49	100,00
Fixed	quadhelix				0,087	0,365	0,133	-0,629	0,803	0,239	0,811				
Random	quadhelix				0,087	0,365	0,133	-0,629	0,803	0,239	0,811				
Fixed	Overall				0,266	0,201	0,040	-0,128	0,660	1,321	0,186				
Random	Overall				0,266	0,201	0,040	-0,128	0,660	1,321	0,186				

Std diff in means and 95% CI

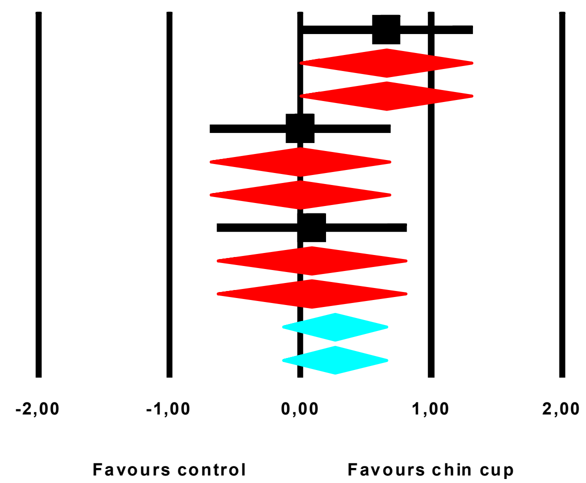

Figure 32b. Subgroup analysis of cephalometric variable UFH with regard to the factor additional appliance used.

LAFH (mm)

No statistically significant results could be found from the exploratory analysis of the linear variable LAFH in correlation with the chin cup use, which was performed on 3 treated groups (SDM=0.267, CI=-0.127 to 0.661, *P*-value=0.184) (Table XI) (Figure 33). The included studies presented very low heterogeneity (I^2=0%, Q=1.944, *P*-value=0.378) whereas the between studies variance was zero (T^2=0) (Table XI). In all 3 included treated groups a light force chin cup was applied and, therefore, the relevant exploratory subgroup analysis was not applicable (Table XII). The exploratory analysis of the same variable based on “additional appliances used”, also performed in 3 treated groups did not indicate any trends (Table XIII) (Figure 33b).

Model	Study name	Sample size: Treated	Sample size: Control	Std diff in means	Standard error	Variance	Upper limit	Lower limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Random)	Relative weight
	Barrett et al., 2010a	14	20	0,037	0,348	0,121	0,720	-0,646	0,106	0,916	8,23	33,27	8,23	33,27
	Barrett et al., 2010b	12	20	0,075	0,365	0,133	0,791	-0,641	0,204	0,838	7,50	30,29	7,50	30,29
	Tuncer et al., 2009	20	18	0,636	0,333	0,111	1,289	-0,016	1,911	0,056	9,02	36,44	9,02	36,44
Fixed				0,267	0,201	0,040	0,661	-0,127	1,328	0,184				
Random				0,267	0,201	0,040	0,661	-0,127	1,328	0,184				

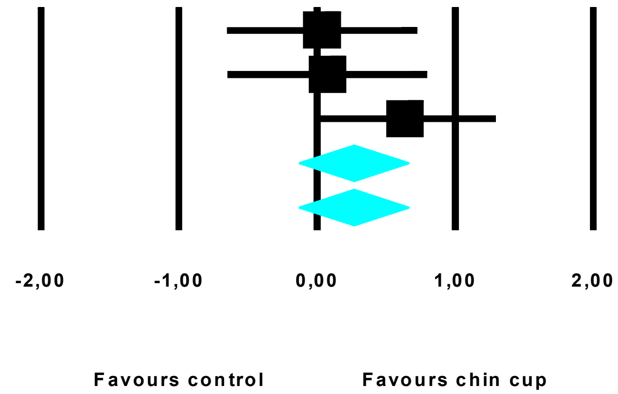

Figure 33. Forest plot of the cephalometric variable LAFH.

Model	Group by Additiona appliance	Study name	Sample size: Control	Sample size: Treated	Std diff in means	Standard error	Variance	Upper limit	Lower limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Separate tau)	Relative weight
	bite plane	Tuncer et al., 2009	18	20	0,636	0,333	0,111	1,289	-0,016	1,911	0,056	9,02	100,00	9,02	100,00
Fixed	bite plane				0,636	0,333	0,111	1,289	-0,016	1,911	0,056				
Random	bite plane				0,636	0,333	0,111	1,289	-0,016	1,911	0,056				
	nothing	Barrett et al., 2010a	20	14	0,037	0,348	0,121	0,720	-0,646	0,106	0,916	8,23	100,00	8,23	100,00
Fixed	nothing				0,037	0,348	0,121	0,720	-0,646	0,106	0,916				
Random	nothing				0,037	0,348	0,121	0,720	-0,646	0,106	0,916				
	quadhelix	Barrett et al., 2010b	20	12	0,075	0,365	0,133	0,791	-0,641	0,204	0,838	7,50	100,00	7,50	100,00
Fixed	quadhelix				0,075	0,365	0,133	0,791	-0,641	0,204	0,838				
Random	quadhelix				0,075	0,365	0,133	0,791	-0,641	0,204	0,838				
Fixed	Overall				0,267	0,201	0,040	0,661	-0,127	1,328	0,184				
Random	Overall				0,267	0,201	0,040	0,661	-0,127	1,328	0,184				

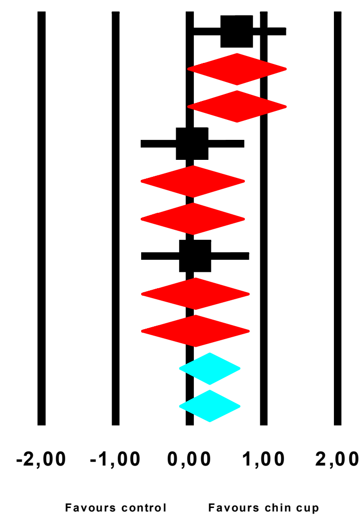

Figure 33b. Subgroup analysis of cephalometric variable LAFH with regard to the factor additional appliance used.

Co-Go (mm)

The results for Co-Go were based on an exploratory analysis of 3 included treated groups and revealed no statistical significance of the slight decrease of this variable (SDM=-0.220, CI=-0.657 to 0.217, *P*-value=0.324) (Table XI) (Figure 34). The accompanying heterogeneity of the included studies was very low (I^2=0%, *Q*=0.369, *P*-value=0.832) and there was no between studies variance (T^2=0) (Table XI). Exploratory subgroup analysis was performed in all 3 included treated groups with regard to the factor "force level" and to the factor "additional appliances used" did not reveal any trend of significance [(Table XII) (Figure 34a) and (Table XIII) (Figure 34b) respectively].

Model	Study name	Sample size		Statistics for each study							Weight (Fixed)		Weight (Random)	
		Treated	Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Random)	Relative weight
	Barrett et al., 2010a	14	20	-0,127	0,349	0,122	-0,811	0,556	-0,365	0,715	8,22	40,83	8,22	40,83
	Barrett et al., 2010b	12	20	-0,173	0,366	0,134	-0,890	0,544	-0,472	0,637	7,47	37,13	7,47	37,13
	Gökalp and Kurt , 2005	13	7	-0,472	0,475	0,225	-1,402	0,459	-0,994	0,320	4,44	22,04	4,44	22,04
Fixed				-0,220	0,223	0,050	-0,657	0,217	-0,987	0,324				
Random				-0,220	0,223	0,050	-0,657	0,217	-0,987	0,324				

Std diff in means and 95% CI

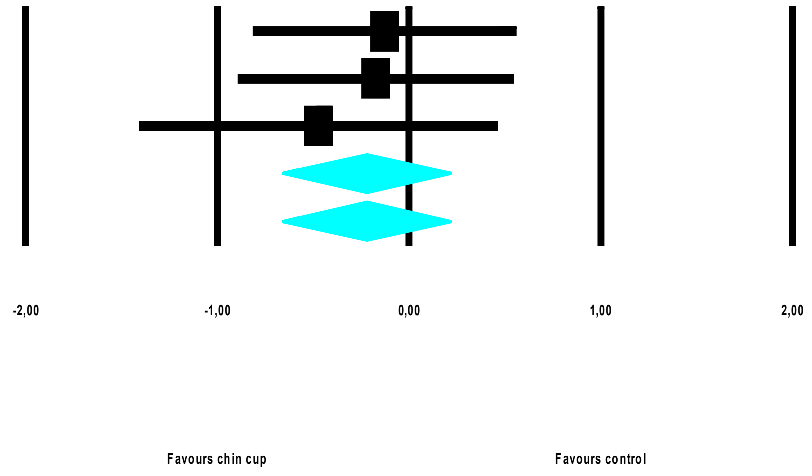

Figure 34. Forest plot of the cephalometric variable Co-Go.

Model	Group by Force level	Study name	Sample size		Statistics for each study							Weight (Fixed)		Weight (Separate tau)	
			Treated	Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Separate tau)	Relative weight
	heavy	Gökalp and Kurt , 2005	13	7	-0,472	0,475	0,225	-1,402	0,459	-0,994	0,320	4,44	100,00	4,44	100,00
Fixed	heavy				-0,472	0,475	0,225	-1,402	0,459	-0,994	0,320				
Random	heavy				-0,472	0,475	0,225	-1,402	0,459	-0,994	0,320				
	light	Barrett et al., 2010a	14	20	-0,127	0,349	0,122	-0,811	0,556	-0,365	0,715	8,22	52,37	8,22	52,37
	light	Barrett et al., 2010b	12	20	-0,173	0,366	0,134	-0,890	0,544	-0,472	0,637	7,47	47,63	7,47	47,63
Fixed	light				-0,149	0,252	0,064	-0,644	0,346	-0,590	0,555				
Random	light				-0,149	0,252	0,064	-0,644	0,346	-0,590	0,555				
Fixed	Overall				-0,220	0,223	0,050	-0,657	0,217	-0,987	0,324				
Random	Overall				-0,220	0,223	0,050	-0,657	0,217	-0,987	0,324				

Std diff in means and 95% CI

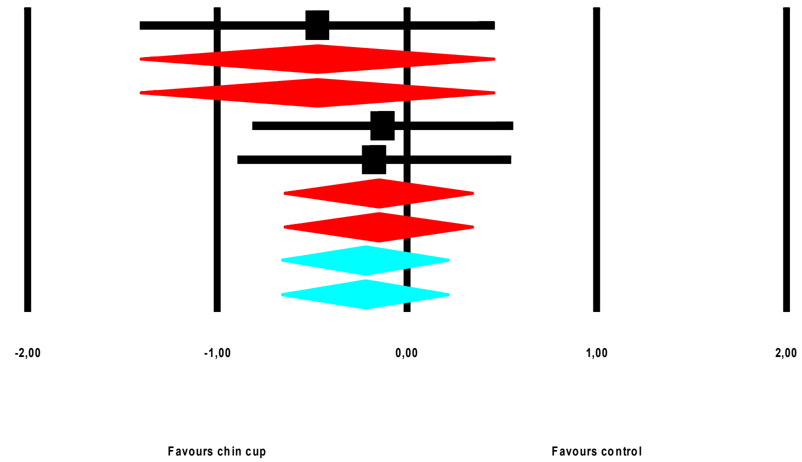

Figure 34a. Subgroup analysis of cephalometric variable Co-Go with regard to the factor force le

Model	Group by Additional appliance	Study name	Sample size		Statistics for each study							Weight (Fixed)		Weight (Separate tau)	
			Treated	Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Separate tau)	Relative weight
	nothing	Barrett et al., 2010a	14	20	-0,127	0,349	0,122	-0,811	0,556	-0,365	0,715	8,22	64,94	8,22	64,94
	nothing	Gökalp and Kurt , 2005	13	7	-0,472	0,475	0,225	-1,402	0,459	-0,994	0,320	4,44	35,06	4,44	35,06
Fixed	nothing				-0,248	0,281	0,079	-0,799	0,303	-0,882	0,378				
Random	nothing				-0,248	0,281	0,079	-0,799	0,303	-0,882	0,378				
	quadhelix	Barrett et al., 2010b	12	20	-0,173	0,366	0,134	-0,890	0,544	-0,472	0,637	7,47	100,00	7,47	100,00
Fixed	quadhelix				-0,173	0,366	0,134	-0,890	0,544	-0,472	0,637				
Random	quadhelix				-0,173	0,366	0,134	-0,890	0,544	-0,472	0,637				
Fixed	Overall				-0,220	0,223	0,050	-0,657	0,217	-0,987	0,324				
Random	Overall				-0,220	0,223	0,050	-0,657	0,217	-0,987	0,324				

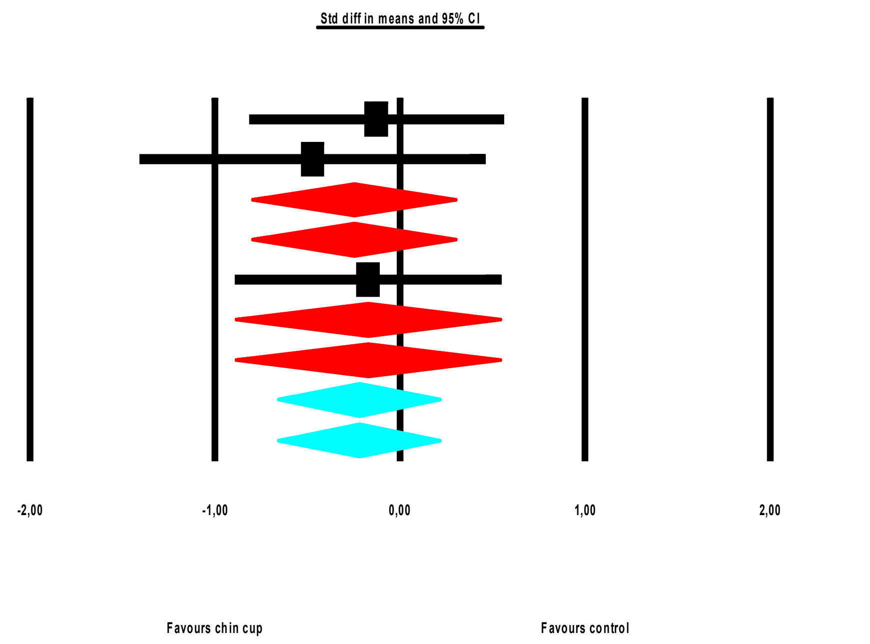

Figure 34b. Subgroup analysis of cephalometric variable Co-Go with regard to the factor additional appliance used.

5.6.3. Dentoalveolar variables

Overjet (mm)

Three treated groups participated in the exploratory analysis for the assessment of Overjet alterations in relation to the chin cup use. Results indicated statistically significant increase of Overjet (SDM=2.623, CI=1.061 to 4.185, P-value=0.001) (Table XI) (Figure 35). Yet, the heterogeneity observed in the included studies was high (I^2=84.912%, Q=13.256, P-value=0.001), as well as the between studies variance (T^2=1.548) (Table XI). Subgroup exploratory analysis with regard to the modifier “force level” included all three treated groups and revealed a tendency of Overjet to increase further with heavier forces (Table XII) (Figure 35a). Exploratory subgroup analysis with regard to the factor “additional appliances used”, also comprised of 3 treated groups, but did not indicate any specific trend (Table XIII) (Figure 35b).

Model	Study name	Sample size: Treated	Sample size: Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight (Fixed)	Weight (Random)	Relative weight (Random)
	Barrett et al., 2010a	14	20	1,423	0,389	0,151	0,661	2,185	3,660	0,000	6,61	51,83	0,59	37,38
	Barrett et al., 2010b	12	20	2,068	0,447	0,200	1,191	2,945	4,623	0,000	5,00	39,15	0,57	36,34
	Gökalp and Kurt , 2005	13	7	5,097	0,932	0,869	3,269	6,924	5,467	0,000	1,15	9,02	0,41	26,28
Fixed				2,007	0,280	0,078	1,458	2,556	7,169	0,000				
Random				2,623	0,797	0,635	1,061	4,185	3,291	0,001				

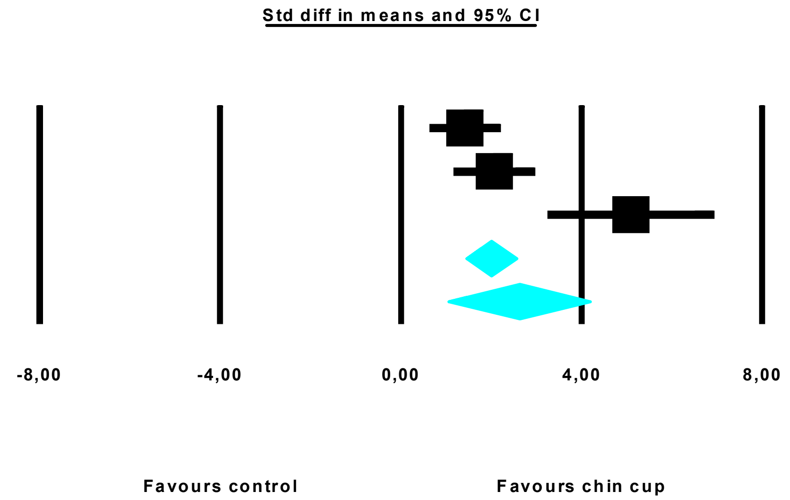

Figure 35. Forest plot of the cephalometric variable Overjet.

Model	Group by Force level	Study name	Sample size: Treated	Sample size: Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Weight (Separate tau)	Relative weight	Relative weight
	heavy	Gökalp and Kurt , 2005	13	7	5,097	0,932	0,869	3,269	6,924	5,467	0,000	1,15	1,15	100,00	100,00
Fixed	heavy				5,097	0,932	0,869	3,269	6,924	5,467	0,000				
Random	heavy				5,097	0,932	0,869	3,269	6,924	5,467	0,000				
	light	Barrett et al., 2010a	14	20	1,423	0,389	0,151	0,661	2,185	3,660	0,000	6,61	5,45	55,88	56,97
	light	Barrett et al., 2010b	12	20	2,068	0,447	0,200	1,191	2,945	4,623	0,000	5,00	4,30	44,12	43,03
Fixed	light				1,701	0,293	0,086	1,125	2,276	5,795	0,000				
Random	light				1,708	0,320	0,103	1,080	2,335	5,332	0,000				
Fixed	Overall				2,007	0,280	0,078	1,458	2,556	7,169	0,000				
Random	Overall				2,065	0,303	0,092	1,472	2,659	6,819	0,000				

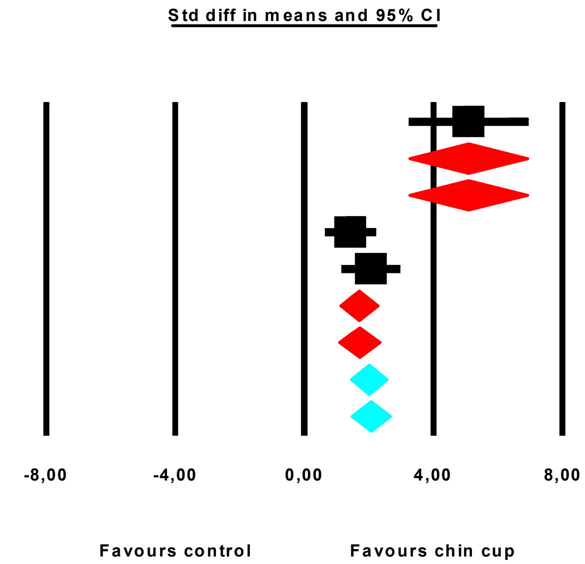

Figure 35a. Subgroup analysis of cephalometric variable Overjet with regard to the factor force level.

Model	Group by Additional appliance	Study name	Sample size: Treated	Sample size: Control	Statistics for each study: Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Separate tau): Weight (Fixed)	Weight (Separate tau)	Relative weight	Relative weight
	nothing	Barrett et al., 2010a	14	20	1,423	0,389	0,151	0,661	2,185	3,660	0,000	6,61	0,16	52,66	85,18
	nothing	Gökalp and Kurt , 2005	13	7	5,097	0,932	0,869	3,269	6,924	5,467	0,000	1,15	0,14	47,34	14,82
Fixed	nothing				1,967	0,359	0,129	1,264	2,671	5,482	0,000				
Random	nothing				3,162	1,834	3,364	-0,433	6,757	1,724	0,085				
	quadhelix	Barrett et al., 2010b	12	20	2,068	0,447	0,200	1,191	2,945	4,623	0,000	5,00	5,00	100,00	100,00
Fixed	quadhelix				2,068	0,447	0,200	1,191	2,945	4,623	0,000				
Random	quadhelix				2,068	0,447	0,200	1,191	2,945	4,623	0,000				
Fixed	Overall				2,007	0,280	0,078	1,458	2,556	7,169	0,000				
Random	Overall				2,130	0,435	0,189	1,278	2,982	4,900	0,000				

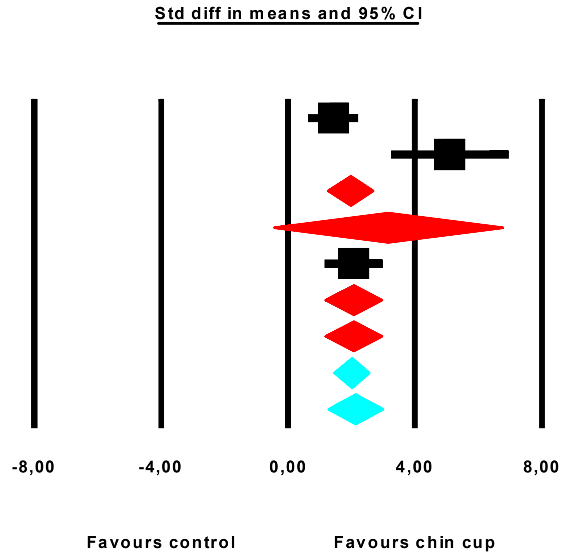

Figure 35b. Subgroup analysis of cephalometric variable Overjet with regard to the factor additional appliance used.

Overbite (mm)

No statistical significance of Overbite following chin cup use was found, as derived from the exploratory analysis performed on 3 treated groups (SDM=-0.239, CI=-0.737 to 0.285; *P*-value=0.345) (Table XI) (Figure 36). This result was accompanied by low heterogeneity (I^2=20.552%, *Q*=2.517, *P*-value=0.284) and even lower between studies variance (T^2=0.040) (Table XI). Neither force magnitude, nor presence or absence of additional appliances seemed to have any impact of significance on Overbite according to the relevant exploratory subgroup analysis which was performed in all 3 treated groups [(Table XII) (Figure 36a) and (Table XIII) (Figure 36b) respectively].

Model	Study name	Sample size		Statistics for each study							Weight (Fixed)		Weight (Random)	
		Treated	Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Random)	Relative weight
	Barrett et al., 2010a	14	20	-0,421	0,352	0,124	-1,111	0,269	-1,196	0,232	8,06	40,58	6,09	39,21
	Barrett et al., 2010b	12	20	0,215	0,366	0,134	-0,503	0,932	0,586	0,558	7,46	37,55	5,74	36,96
	Gökalp and Kurt , 2005	13	7	-0,645	0,480	0,230	-1,585	0,295	-1,345	0,179	4,34	21,87	3,70	23,82
Fixed				-0,231	0,224	0,050	-0,671	0,208	-1,031	0,302				
Random				-0,239	0,254	0,064	-0,737	0,258	-0,943	0,345				

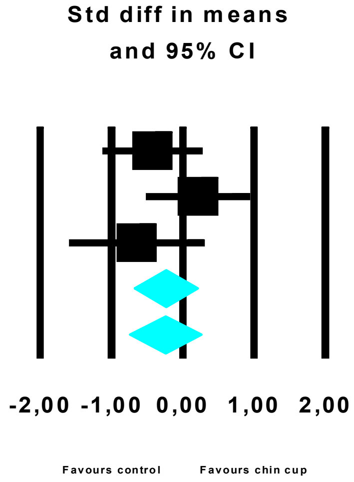

Figure 36. Forest plot of the cephalometric variable Overbite.

Model	Group by Force level	Study name	Sample size		Statistics for each study							Weight (Fixed)		Weight (Separate tau)	
			Treated	Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Separate tau)	Relative weight
	heavy	Gökalp and Kurt , 2005	13	7	-0,645	0,480	0,230	-1,585	0,295	-1,345	0,179	4,34	100,00	4,34	100,00
Fixed	heavy				-0,645	0,480	0,230	-1,585	0,295	-1,345	0,179				
Random	heavy				-0,645	0,480	0,230	-1,585	0,295	-1,345	0,179				
	light	Barrett et al., 2010a	14	20	-0,421	0,352	0,124	-1,111	0,269	-1,196	0,232	8,06	51,94	5,08	51,24
	light	Barrett et al., 2010b	12	20	0,215	0,366	0,134	-0,503	0,932	0,586	0,558	7,46	48,06	4,83	48,76
Fixed	light				-0,116	0,254	0,064	-0,613	0,382	-0,455	0,649				
Random	light				-0,111	0,318	0,101	-0,734	0,512	-0,350	0,727				
Fixed	Overall				-0,231	0,224	0,050	-0,671	0,208	-1,031	0,302				
Random	Overall				-0,274	0,265	0,070	-0,793	0,245	-1,034	0,301				

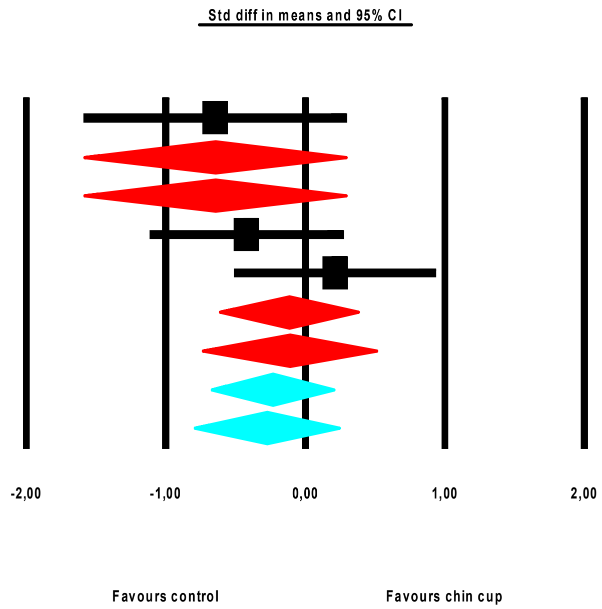

Figure 36a. Subgroup analysis of cephalometric variable Overbite with regard to the factor force level.

Model	Group by Additional appliance	Study name	Sample size: Treated	Sample size: Control	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Relative weight	Weight (Separate tau)	Relative weight
	nothing	Barrett et al., 2010a	14	20	-0,421	0,352	0,124	-1,111	0,269	-1,196	0,232	8,06	64,98	8,06	64,98
	nothing	Gökalp and Kurt , 2005	13	7	-0,645	0,480	0,230	-1,585	0,295	-1,345	0,179	4,34	35,02	4,34	35,02
Fixed	nothing				-0,500	0,284	0,081	-1,056	0,057	-1,759	0,079				
Random	nothing				-0,500	0,284	0,081	-1,056	0,057	-1,759	0,079				
	quadhelix	Barrett et al., 2010b	12	20	0,215	0,366	0,134	-0,503	0,932	0,586	0,558	7,46	100,00	7,46	100,00
Fixed	quadhelix				0,215	0,366	0,134	-0,503	0,932	0,586	0,558				
Random	quadhelix				0,215	0,366	0,134	-0,503	0,932	0,586	0,558				
Fixed	Overall				-0,231	0,224	0,050	-0,671	0,208	-1,031	0,302				
Random	Overall				-0,231	0,224	0,050	-0,671	0,208	-1,031	0,302				

Std diff in means and 95% CI

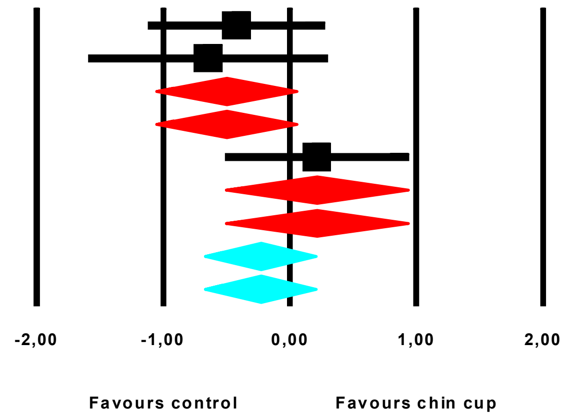

Figure 36b. Subgroup analysis of cephalometric variable Overbite with regard to the factor additional appliance used.

6. DISCUSSION

6.1. Strengths and weaknesses

In the current meta-analysis a thorough electronic and hand search of the literature was performed and 5 studies were finally included (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Barrett et al., 2010; Gökalp and Kurt, 2005; Tuncer et al., 2009), among which none was an RCT while from the included CCTs none was assessed as very good in quality according to the above-mentioned publication bias tests. Three of them were appraised as of medium quality (Abdelnaby and Nassar, 2010; Barrett et al., 2010; Tuncer et al., 2009) and two of them as of low (Altuğ et al., 1989; Gökalp and Kurt, 2005), while the overall estimate was low. The sample sizes were adequate in 3 of the included studies (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Tuncer et al., 2009) and inadequate in the rest 2 of them (Barrett et al., 2010; Gökalp and Kurt, 2005). From the included studies 7 treated groups were derived, however, two studies contributed with two treatment groups each (Abdelnaby and Nassar, 2010; Barrett et al., 2010) whereas the rest 3 studies contributed with just one treated group each and, therefore, this could comprise an additional source of publication bias.

When examining the characteristics of each study, it was observed that not all treated groups were identical in the amount of force exercised, additional appliances used, duration of treatment, suggested hours of chin cup wear and the chronological age of patients. Different protocols in the clinical procedures might have led in different results, however.

For the lateral cephalometric radiographs of the included studies possibly different magnifications might have been used, since they had been taken at different times in different environments from different equipment. This possible source of bias affects mainly the linear variables, namely Wits appraisal, Co-Gn, N-Me, UFH, LAFH, Co-Go, Overjet and Overbite.

Not all studies and not all independent treated groups participated in the statistical processing of many of the variables examined. For the variables Co-Gn, N-Me, UFH, LAFH, Co-Go, Overjet and Overbite, where 4 or less treated groups were included, an exploratory analysis was conducted instead of a formal meta-analysis. The small number of included treated groups significantly attenuated the validity of the corresponding results and imposed precaution in their interpretation since they are considered as secondary in comparison with those derived from a meta analysis. High levels of heterogeneity, evident in most of the variables examined, questions in addition the validity and reliability of the results. Regarding subgroup analysis, the number of treated groups per subgroup was really small, ranging between 1 and 3 and this deduced even more strength of the relevant outcomes, where trends only, if any, could be observed.

The existence of more RCTs and more detailed and properly designed and structured cohort studies regarding chin cup for Class III malocclusion could have significantly contributed in the investigation of more treatment effects from the chin cup appliance and in the extraction of more sound results.

To balance the reported weaknesses and to increase the strength of the current evaluation, a strict inclusion-exclusion protocol was applied according to which specific restrictions in the potentially included study designs, the participants' characteristics and the principal outcome measures were imposed. To reduce publication bias, no language restrictions were applied and translation was arranged for one included

Turkish paper (Altuğ, 1989). Also, no restrictions were applied concerning publication year and status of the original studies, whereas "Grey literature" was not excluded as well. Selection of studies, data extraction and management and assessment of risk of bias of included studies (quality analysis) were independently performed by two reviewers and a third one was involved where in doubt to remove subjectivity from the processes. Authors were contacted when no precise information was given. With regard to the statistical computations, the pooled estimate Standard Difference in Means was used instead of Mean Difference, to alleviate discrepancies derived from possibly different magnification factors of the original lateral cephalometric radiographs, and from non standardized and non calibrated measurements as well. Finally, in current meta-analysis the random effects model was used instead of the fixed effect to take into consideration the heterogeneity of the studies included and to counterbalance the dominance of the studies with big sample sizes over those with small ones.

6.2. Discussion on the variables examined

SNA does not seem to de statistically significantly affected by the chin cup treatment according to the results of the current meta analysis in which all 7 treated groups participated. However, the observed heterogeneity of the included studies, being between medium and high, and the medium between studies variance imposes precaution in the interpretation of this result but also justifies the use of the random effects model since variation of the true effect size among the studies can not be attributed purely to sampling error. This outcome is in accordance with 3 of the included studies (Abdelnaby and Nassar, 2010; Barrett et al., 2010; Gökalp and Kurt, 2005), while it opposes to the fourth one (Tuncer et al., 2009) which supports the significant increase of SNA with the chin cup, and to the fifth one (Altuğ et al., 1989), which supports on the contrary a significant decrease of SNA.

SNB presented statistically significant reduction following the use of the chin cup as revealed from the meta-analysis performed on all 7 treated groups from all five included studies (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Barrett et al., 2010; Gökalp and Kurt, 2005; Tuncer et al., 2009). Both the heterogeneity and the between studies variance were high, which means that the true effect size definitively varies among the studies and this is definitively not purely due to sampling error. The high heterogeneity observed in the included studies questions the validity of the results, as the statistical significance observed might have not been real, but also justifies the use of the random effects model. According to McNamara (2005) the reduction of SNB could be explained by the force chin cup applies directly to the mandible trying to restrict its horizontal growth converting it to some extend into vertical growth. Deguchi et al., (1999) though, expressed the opinion that despite the clockwise rotation of the mandible, the simultaneous closure of the gonial angle and the upward and forward rotation of the corpus of the mandible tend to bring B point anteriorly. As a result, there is a reduction in the absolute value of SNB, however, the bigger the corpus of the mandible, in other words the more accentuated the Class III, the lesser this reduction is (Deguchi et al., 1999).

Statistically significant increase in ANB is the result of the corresponding meta-analysis, in which all 7 treated groups participated. All 5 included studies (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Barrett et al., 2010; Gökalp and Kurt, 2005; Tuncer et al., 2009) are in agreement with this result but in the Altuğ et al. (1989) study a much further increase in this variable than the rest ones was found. Perhaps this is in connection with the individualized protocol applied in that study. The observed heterogeneity and the between studies variance are high once again and one should consider the reservations they impose and the weakness of the significance of the result they entail. Besides, the use of the random effects model was deemed as more appropriate. Graber (1977) expressed the opinion that when ANB increases, while it does not indicate where exactly the changes are affected, it does mirror the resolution of the skeletal malocclusion that is seen clinically. According to Deguchi et al. (1999) this increase elaborates an evident improvement of the facial profile. Therefore, the clinical effectiveness of the chin cup, by increasing ANB, can be positively appraised.

For the meta-analysis of the variable Wits appraisal only three studies (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Barrett et al., 2010) were quantitatively assessed forming 5 independent groups. Statistically significant increase of this variable was found, and this is in accordance with all 3 included studies. Once again, in the study of Altuğ et al. (1989) a much grater increase than the rest ones was found. The accompanying heterogeneity and the between studies variance were very high and together with the small number of included treated groups, suggest considerable precaution in the interpretation of the results and the statistical significance found. Besides, the random effects method was deemed as obligatory since true effect size definitively varies among the studies and this is definitively not purely due to sampling error.

No statistically significant results were found for the linear mandibular variable Co-Gn. This result was based on the statistical elaboration of the data of 4 independent treatment groups derived from 3 studies (Barrett et al., 2010; Gökalp and Kurt, 2005; Tuncer et al., 2009). Thus, exploratory analysis was performed instead of the formal meta-analysis and its outcome is attenuated and regarded as secondary. In addition, this variable, being a linear measurement, confines the additional potential bias of different magnification factors in the original radiographs. The non statistically significant decrease of this variable following chin cup use, as derived from the aforementioned analysis, is also supported by two of the included studies (Barrett et al., 2010; Gökalp and Kurt, 2005) while the third study (Tuncer et al., 2009) opposed to them by supporting a significant increase. The observed heterogeneity of the included studies was low and accordingly low was the between studies variance.

The angular variable SN-ML presented a statistically significant increase following chin cup use according to the results of the current meta-analysis, where quantitative assessment of 5 treated groups from 4 studies (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Gökalp and Kurt, 2005; Tuncer et al., 2009) was performed. All 4 included studies agree with this result which is indicative of the increase in the anterior face height that takes place with the chin cup use, probably due to the backwards and downwards rotation of the mandible (Graber, 1977). However, the moderate to high accompanying heterogeneity and the moderate between studies variance, apart from justifying the use of the random effects model, also suggest caution with regard to the reliability of this finding.

By pooling data from 4 studies (Abdelnaby and Nassar, 2010; Altuğ et al., 1989; Barrett et al., 2010; Gökalp and Kurt, 2005), which formed 6 independent treated groups, meta-analysis was performed and revealed statistically significant reduction of the Gonial angle. Although this is in accordance with all 4 participated studies, the Altuğ et al. (1989) study supported a much further reduction than the rest ones standing as an outlier in the relevant forest plot (Figure 30). The included studies presented high heterogeneity and moderate between studies variance and call for precaution and critical appraisal when interpreting the corresponding results. Deguchi (1998) attempted in the past to approach etiologically the closure of the gonial angle due to chin cup wear by considering the direction of the force applied which passes through the occipital area and the glenoid fossa either via or underneath the condyle and seems to work as fulcrum around which the ramus of the mandible tends to rotate (Deguchi, 1998). Therefore, according to him, apart from the condyle remodeling to a more forward direction (Deguchi et al., 1999) the gonial area has also some freedom to remodeling and becomes less obtuse. The same was also reported by other authors more recently (Basciftci et al., 2008; Chang et al., 2005).

Two studies (Abdelnaby and Nassar, 2010; Tuncer et al., 2009) contributed, by providing 3 treated groups in the exploratory analysis performed for the variable N-Me. The results revealed statistical significant increase, yet, there are plenty of possible biases that deduce strength and validity of this result. Not only is the number of included studies very small but it is also the linearity of the measurement that makes it more susceptible to the possible different magnification parameter and additionally a moderate heterogeneity of the included studies accompanied by a small between studies variance further attenuated the significance of the derived outcome. However, both the included studies (Abdelnaby and Nassar, 2010; Tuncer et al., 2009) support the increase of N-Me.

Although N-Me seemed to present statistically significant increase following chin cup treatment, neither UFH nor LAFH presented any statistical significance in their results. This of course is not in logical conclusion and possibly harbors a kind of bias. The results for both UFH and LAFH derived from 2 studies (Barrett et al., 2010; Tuncer et al., 2009) which formed 3 treated groups and, therefore, exploratory analysis was performed with all the subsequent consequences in the validity of the results. In addition, they are all linear measurements, which further deduce credit and gravity from the outcome. However, the observed heterogeneity and between studies variance was minimal for UFH and zero for LAFH. For UFH two treated groups (Barrett et al., 2010b; Tuncer et al., 2009) supported its increase while the third treated group (Barrett et al., 2010a) did not show any difference at all. For LAFH all 3 treated groups (Barrett et al., 2010a; Barrett et al., 2010b; Tuncer et al., 2009) agreed with its augmentation. Due to all these possible limitations and possible sources of bias, one would better be reserved in the interpretation of the results and consider that there is not sufficient proof to examine properly if statistical significance can be attributed to the observed increase of UFH and LAFH.

Exploratory analysis was also performed for Co-Go, in which 2 included studies (Barrett et al., 2010; Gökalp and Kurt, 2005) formed 3 treated groups. Although a slight reduction in the Co-Go was found, which was supported by all 3 participating treated groups, no statistical significance has been identified. Exploratory analysis of linear measurements, as explained earlier, is submitted to serious contestation of its validity and one would obviously require more data to conclude to a more reliable result. Observed heterogeneity and between studies variance was zero (Figure 34).

Statistically significant increase of Overjet was the outcome of the exploratory analysis performed on 2 studies (Barrett et al., 2010; Gökalp and Kurt, 2005) which formed 3 treated groups. All 3 treated groups agree with this increase but one of them (Gökalp and Kurt, 2005) supports a much further increase than the rest two (Figure 35). Heterogeneity of the included studies and between studies variance were both high. Therefore, although statistical significance was found, exploratory analysis of a linear measurement with high heterogeneity imposes serious precautions in the validity of this result.

Overbite is the last variable examined, based like Overjet in 2 studies (Barrett et al., 2010; Gökalp and Kurt, 2005) which formed 3 treated groups. According to the corresponding exploratory analysis there is no statistical significance in the slight reduction of Overbite observed. From the included treated groups, two (Barrett et al., 2010a; Gökalp and Kurt, 2005) supported the reduction of Overbite while the third (Barrett et al., 2010b) its increase. It seems that Quad-helix impedes the Overbite reduction, given that this is the only difference between the two Barrett et al. subgroups (Barrett et al., 2010a; Barrett et al., 2010b). Observed heterogeneity and between studies variance was low. One possible explanation of the Overbite reduction could be the clockwise rotation of the mandible due to chin cup use which “opens” the bite (Thilander, 1965). Increase of the SN-ML angle seems also to contribute to the Overbite reduction (Chang, 2005).

Subgroup analyses performed for all variables with both modifiers (i.e. “force level” and “additional appliance used”) including very few studies per subgroup, so only trends and tendencies could be observed. A trend of the chin cup to be generally more effective when heavy forces were applied was evident since there was a tendency towards further reduction of the SNB angle and simultaneously further increase of the variables ANB, Wits appraisal, N-Me and Overjet. The concomitant use of dissocclusion bite planes also seemed to enhance the clinical effectiveness of the chin cup with the further increase of the variables ANB and Wits appraisal, but at the same time a grater increase in the vertical dimensions through the variable SN-ML seemed to have taken place. No similar subgroup analysis has been found in the literature to enable comparison of these results.

Liu et al. (2011) also performed a systematic review on chin cup as mentioned in the introduction. The following six variables were examined in their study: SNB, ANB, Gonial angle, Co-Gn, Go-Me and Ar-Go. For the variables SNB, ANB and Co-Gn their results coincided with the present meta-analysis. The variables Go-Me and Ar-Go were not examined in the current study but the variable Co-Go was investigated instead; the general outcome was that there are no significant alterations in the dimensions of the mandible, which coincides with the results of Liu et al. (2011). Regarding the gonial angle, they observed a statistically significant increase, which is in contrast with the present study.

However, the study of Liu et al. (2011) based on four original studies (Alhaija and Richardson, 1999; Arman et al., 2004; Deguchi and McNamara, 1999 and Lin et al., 2007) presents the following flows: (a) In the Arman et al. (2004) study the control sample is not contemporaneous with the treated sample, not even comprised from the same group of patients including Class III or Class I patients, although the treated sample consists solely of Class III patients. The matching is referred only to the skeletal and chronological age of the participants. Also in this study 3 different treated subgroups are discerned, one in which only chin cup was used, one where the chin cup was accompanied by a bite plate and a third one where reverse headgear was used. However, no similar subgroups can be discerned in the meta-analysis of Liu et al. (2011). (b) In all the quantitatively assessed studies (Alhaija and Richardson, 1999; Deguchi and McNamara, 1999; Lin et al., 2007) except from the Arman et al. (2004) study an additional maxillary appliance is used for protracting or flaring the incisors concomitantly with the chin cup. Maxillary protraction appliances were excluded from the present study, as they significantly affect the results of the variables SNA and ANB. Therefore, their presence deprives the author from the ability to critically appraise whether observed changes should be attributed to the chin cup or to the additional appliances used and, if both involved, to what extend to each one of them. Gathering in the same meta-analysis primary studies where only chin cup was applied and others where chin cup and maxillary protraction appliances were used, seems like “mixing studies that measure apples with those that measure oranges” (Jones, 1992). (e) The authors used as point estimate the Mean Difference (MD) of the included studies, even for the linear measurements, where probably the magnification factor of the lateral cephalometric radiographs of the original studies might have differed.

In contrast to the study of Liu et al. (2011), in the present meta-analysis although more strict inclusion criteria were applied, more studies were included, which in turn provided more subgroups. In addition 13 variables were examined in total by using the pooled estimate Standard Difference in Means (SDM) that compares the difference of two means after having standardized the difference.

7. CONCLUSIONS

The use of the occipital pull chin cup is a common treatment approach for the correction of Class III malocclusion in the early age. Thirteen variables were examined in total either with meta-analysis or with exploratory analysis, depending on the number of studies included. These variables were divided into 3 categories: skeletal sagittal (SNA, SNB, ANB, Wits appraisal and Co-Gn), skeletal vertical (SN-ML, Gonial angle, N-Me, UFH, LAFH and Co-Go) and dentoalveolar (Overbite and Overjet). Statistically significant alterations were found both for skeletal and dentoalveolar variables. More specifically, the variables ANB, Wits appraisal, SN-ML, N-Me and Overjet were increased significantly whereas the variables SNB and Gonial angle decreased significantly. For the variables SNA UFH, LAFH, Co-Gn, Co-Go and Overbite no statistically significant alterations were found. According to the results of the subgroup analyses heavier forces and the simultaneous use of dissocclusion bite planes seem to enhance the clinical effectiveness of the chin cup. All these alterations are indicative of a clockwise, backwards and downwards rotation of the mandible with simultaneous reformation of its shape in the gonial area and corresponding increase of the overjet.

The limited number of included studies though, the high heterogeneity observed in most of the variables and the linear manner of many of them, suggest some precaution in the interpretation of the results of this meta-analysis. The existence of more comparable studies with proper design, sample size, appliance use and measurements examined could have been of importance in the extraction of more reliable results in a broader scale.

8. ACKNOWLEDGEMENTS

I owe my deepest gratitude to my supervisors, Associate Professor Moschos A. Papadolpoulos and Associate Professor Ioulia Marathiotou Ioannidou for devoting me their time and energy to familiarize me with the essence of meta-analysis and for very carefully and pedagogically, with patience and encouragement, leading me towards the proper formation of this thesis.

I would also like to express, in addition to my grate appreciation, my admiration to the former Director of the Orthodontic Department, Professor Athanasios E. Athanasiou for his academic rigorism and for guiding us (the team of the post-graduate students), being always by our side, provisioning for our significant or insignificant problems.

I additionally feel obliged to my husband, Mr. Stelios Mavromichalis for his invaluable help in managing all software systems necessary for the fulfillment of the requirements of this thesis.

I further need to express my thanks to the Elsevier company and to Dr. Jan Glover, Dr. Dave Izzo, Dr. Karen Odato and Dr. Lei Wang for granting me permission to reuse photos, originally published in their journals, books or websites.

I finally would like to express my thanks to Dr. Gökalp for providing me with the *P*-values of the variables examined in his study included in this thesis and Dr. Chang for sending me his articles relevant to the subject of the current thesis.

9. FUNDING

This study has been performed in the context of the thesis during the postgraduate studies of the author Maria Chatzoudi and there has been no funding from any source.

10. REFERENCES

Abdelnaby YL, Nassar EA. Chin cup effects using two different force magnitudes in the management of Class III malocclusions. Angle Orthod 2010;80:957-62.

Abu Alhaija ES, Richardson A. Long-term effect of the chincap on hard and soft tissues. Eur J Orthod 1999;21:291-8.

Ainsworth NJ. The incidence of dental disease in the children. In Medical Research Counsil: Reports of the Committee for the investigation of dental disease. Special Report Series 1925;No 97.

Alarcón JA, Bastir M, Rosas A, Molero J. Chincup treatment modifies the mandibular shape in children with prognathism. Am J Orthod Dentofacial Orthop 2011;140:38-43.

Alexander KN. Genetic and phenotypic evaluation of the class III dentofacial deformity: Comparisons of three populations [Dissertation]. North Carolina (PA): University of North Carolina at Chapel Hill, 2007.

Allwright WC, Burndred WH. A survey of handicapping dentofacial among Chinese in Hong Kong. Int Dent J 1964;14:505-19.

Altemus LA. Frequency of the incidence of malocclusion in American Negro children aged 12-16. Angle Orthod 1959;29:189-200.

Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001;134:663-694.

Altuğ Z, Erdem D, Rübendüz M. Investigation of the functional treatment effects of the skeletal and dental Class III anomalies on the skeletal region. Ankara Univ Hekim Fak Der 1989;16:447-52.

Anello C, Fleiss JL. Exploratory or analytic meta-analysis: should we distinguish between them? J Clin Epidemiol 1995;48:109-18.

Angle EH. Orthodontia. Dent Cosmos 1920:11.

Antczak AA, Tang J, Chalmers TC. Quality assessment of randomized control trials in dental research. I. Methods. J Periodontal Res 1896;21:305-14.

Arman A, Toygar TU, Abuhijleh E. Profile changes associated with different orthopedic treatment approaches in Class III malocclusions. Angle Orthod 2004;74:733-40.

Baccetti T, Franchi L, Schulz SO, McNamara JA. Treatment timing for an orthopedic approach to patients with increased vertical dimension. Am J Orthod Dentofacial Orthop 2008;133:58-64.

Barrett AA, Baccetti T, McNamara JA. Treatment effects of the light-force chincup. Am J Orthod Dentofacial Orthop 2010;138:468-76.

Basciftci FA, Korkmaz HH, Uşümez S, Eraslan O. Biomechanical evaluation of chincup treatment with various force vectors. Am J Orthod Dentofacial Orthop 2008;134:773-81.

Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994;50:1088-101.

Borenstein M, Hedges LV, Rothstein H. Fixed-effect versus random-effects models. In: Borenstein M, Hedges LV, Higgins JPT, Rothstein H, eds. Introduction to Meta-Analysis. Hoboken, NJ: Wiley, 2009.

Buschang H, Sankey W, English P. Early treatment of hyperdivergent open-bite malocclusions. Semin Orthod 2002;8:130-40.

Chang HP, Lin HC, Liu PH, Chang CH. Geometric morphometric assessment of treatment effects of maxillary protraction combined with chin cup appliance on the maxillofacial complex. J Oral Rehabil 2005;32:720-8.

Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In Egger M, Davey Smtih G, Altman DG. Systematic reviews in health care: meta-analysis in context. London: BMJ books, 2001:285-312.

Deguchi T. Force distribution of the temporomandibular joint and temporal bone surface subjected to the head-chincup force. Am J Orthod Dentofacial Orthop 1998;114:277-82.

Deguchi T, Kuroda T, Hunt NP, Graber TM. Long-term application of chincup force alters the morphology of the dolichofacial Class III mandible. Am J Orthod Dentofacial Orthop 1999;116:610-5.

Deguchi T, Kuroda T, Minoshima Y, Graber TM. Craniofacial features of patients with Class III abnormalities: growth-related changes and effects of short-term and long-term chincup therapy. Am J Orthod Dentofacial Orthop 2002;121:84-92.

Deguchi T, McNamara JA. Craniofacial adaptations induced by chincup therapy in Class III patients. Am J Orthod Dentofacial Orthop 1999;115:175-82.

Delaire J. Confection du masque orthopedique. Rev Stomatol Paris 1971;72:579-84.

DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177-88.

Dickersin K. The existence of publication bias and risk factors for its occurrence. J Am Med Assoc 1990;263:1385-9.

Dickersin K, Berlin JA. Meta-analysis: state-of-the-science. Epidemiol Rev 1992;14:159–76.

Downs HS, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomized studies of health care interventions. J Epid Community Health 1998;52:377-84.

Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomized controlled trials published in Engilsh and German. Lancet 1997c;350:326-9.

Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. BMJ 1997;315:629-34.

Enrich RE, Brodie AG, Blayney JR. Prevalence of Class I, Class II and Class III malocclusions (Angle) in an urban population: an epidemiological study. J Dent Res 1964;44:947-1014.

Forrest JL, Miller SA. Evidence-Based Decision Making: A Translational Guide for Dental Professionals. Philadelphia: Lippincott, Williams & Wilkins, 2008.

Fränkel R. Decrowding during eruption under the screening influence of vestibular shields. Am J Orthod 1974;65:372-406.

Fränkel R, Fränkel C. Orofacial orthopedics with the function regulator. Munich: S. Karger, 1989.

Garner LD, Butt MH. Malocclusion in black America and Nyeri Kenyans. Angle Orthod 1985;55:139-46.

Gehlbach SH. Interpreting the medical literature. New York: McGraw-Hill, 1993.

Glass G. Primary, secondary and meta-analysis of research. Educ Res 1976;5:3-8.

Gökalp H, Kurt G. Magnetic resonance imaging of the condylar growth pattern and disk position after chin cup therapy: a preliminary study. Angle Orthod 2005;75:568-75.

Goose DH, Thompson DG, Winter FC. Malocclusion of school children of the West Midlands (England). Br Dent J 1957;102:174-8.

Graber W. Chin cup therapy for mandibular prognathism. Am J Orthod 1977;72:23-41.

Green B, Hall J. Quantitative Methods for Research Synthesis. Beverly Hills, CA: SagePubns, 1986.

Greenhalgh T, Taylor R. Papers that go beyond numbers (qualitative research) BMJ 1997;315:740-3.

Harris JE, Kowalski CJ, Watnick SS. Genetic factors in the shape of the craniofacial complex. Angle Orthod 1973;43:109-11.

Harvold EP. A study of pressure and tension systems in bone modelling and mandibular positioning. Dent Clin North Am 1975;19:473-84.

Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration 2011; Available from: www.cochrane-handbook.org, Aug 2011.

Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.

Hill IN, Blayney JR, Wolf W. The Evanston Dental Caries Study. XIX. Prevalence of malocclusion of children in a fluoridated and control area. J Dent Res 1959;38:782-94.

Horowitz HS, Doyle BS. Occlusal relations in children born and reared in an optimally fluoridated community. Angle Orthod 1970;40:104-11.

Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods 2006;11:193-206.

Humphreys HF, Leighton BC. A survey of anteroposterior abnormalities of the jaws in children between the ages of two and five and a half years of age. Br Dent J 1950;88:3-15.

Huque MF. Experiences with meta-analysis in NDA submissions. Proceedings of the Biopharmaceutical Section of the American Statistical Association 1988;2:28-33.

Ishii H, Morita S, Takeuchi Y, Nakamura S. Treatment effect of combined maxillary protraction and chincap appliance in severe skeletal Class III cases. Am J Orthod Dentofacial Orthop 1987;92:304-12.

Isiekwe MC. Malocclusion in Laso, Nigeria. Community Dent Res 1983;1:59-62.

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996;17:1-12.

Janket SJ, Wightman A, Baird AE, Van Dyke TE, Jones JA. Does periodontal treatment improve glycemic control in diabetic patients? A meta-analysis of intervention studies. J Dent Res 2005;84:1154-9.

Jones DR. Meta-analysis of observational epidemiological studies: a review. J R Soc Med 1992;35:37-53.

Kalavritinos M, Papadopoulos MA, Nasiopoulos A. Dental arch and cephalometric changes following treatment of Class III malocclusion by means of the Functional Regulator (FR-3) appliance. J Orofac Orthop 2005;66:135-47.

Kerr WJ, Ten Have TR. Changes in soft tissue profile during the treatment of Class III malocclusion. Br J Orthod 1987;14:243-9.

Kolokitha OE, Kaklamanos EG, Papadopoulos MA. Prevalence of nickel hypersensitivity in orthodontic patients: a meta-analysis. Am J Orthod Dentofacial Orthop 2008;134:722.e1-12;disc 722-3.

Langlade M. Diagnostic orthodontique. Paris: Maloine, 1981:704-28.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis J, Clarke M, Devereaux P, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. J Clin Epidemiol 2009;62:1-34.

Light RJ, Pillemer DB. Summing up: the science of reviewing research. Cambridge, MA: Harvard University Press, 1984.

Lin HC, Chang HP, Chang HF. Treatment effects of occipitomental anchorage appliance of maxillary protraction combined with chincup traction in children with Class III malocclusion. J Formos Med Assoc 2007;106:380-91.

Lin JJ. Prevalence of malocclusion in Chinese children age 9-15. Clin Dent 1985;5:57-65.

Linder-Aronson S. Adenoids-their efforts on mode of breathing and nasal airflow and their relationship to characteristics of the facial skeleton and the dentition. Acta Otolaryngol 1970;265:1.

Liu ZP, Li CJ, Hu HK, Chen JW, Li F, Zou SJ. Efficacy of short-term chincup therapy for mandibular growth retardation in Class III malocclusion. Angle Orthod 2011;81:162-8.

Lu YC, Tanne K, Hirano Y, Sakuda M. Craniofacial morphology of adolescent mandibular prognathism. Angle Orthod 1993;63:277-82.

Macdonald KE, Kapust AJ, Turley PK. Cephalometric changes after the correction of Class III malocclusion with maxillary expansion/facemask therapy. Am J Orhtod Dentofacial Orthop 1999;116:13-24.

Magnusson TE. An epidemiologic study of occlusal anomalies in relation to development of the dentition in Icelandic children. Community Dent Oral Epidemiol 1976;4:121-8.

Massler M, Fränkel JM. Prevalence of malocclusion in children aged 14-18 years. Am J Orthod 1951;37:751-68.

McGuigan DG. The Hapsburgs. London: WH Allen, 1966.

McNamara JA. Treatment of Patients in the Mixed Dentition. In Graber TM, Vanarsdall RL, Vig KWL. Orthodontics: Current Principles and

Techniques. China: CV Mosby, 2005:543-77. Figure 3 from page 549, figure 4 from page 548, Copyright Elsevier.

McNamara JA Jr, Huge SA. The functional regulator (FR-3) of Fränkel. Am J Orthod 1985;88:409-24.

Mitchell L. An introduction to orthodontics. UK: Oxford University Press, 1996: 114-21.

Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, Pham B, Klassen TP. Assessing the quality of reports of randomized trials: implication for the conduct of meta-analyses. Health Technol Assess 1999;3:i-iv,1-98.

Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomized trials. Br Med J 2010;340:c869.

Moyers RE. Skeletal contributions to occlusal development. In McNamara JA. The biology of occlusal development. Ann Arbor: University of Michingan, 1977:122-35.

Newman GV. Prevalence of malocclusion in children 6-14 years of age and treatment in preventable cases. J Am Dent Assoc 1956;52:566-75.

Ngan P. Treatment of Class III Malocclusion in the Primary and Mixed Dentitions. In Bishara ES, Textbook of Orthodontics. US: Saunders, 2001:375-414.

Ngan P, Hägg U, Yiu C, Merwin D, Wei SH. Soft tissue and dentoskeletal profile changes associated with maxillary expansion and protraction headgear treatment. Am J Orthod Dentofacial Orthop 1996;109:38-49. Figure 1 from page 39, Copyright Elsevier.

Oppenheim A. A possibility for physiologic orthodontic movement. Am J Orthod Oral Surg 1944;30:345-6.

Otuyemi OD, Abidoye RO. Malocclusion in 12-year-old suburban and rural Nigerian children. Community Dent Health 1993;10:375-80.

Oxford Centre of Evidence Based Medicine. Table of Evidence Glossary. Available from http://www.cebm.net/?o=1116

Papadopoulos MA. Interview - Part I. Orthodontic Forum - J Polish Orthodontic Society 2007;5:137-43.

Papadopoulos MA. Meta-analysis in evidence-based orthodontics. Orthod Craniofac Res 2003;6:12-26.

Papageorgiou SN, Zogakis IP, Papadopoulos MA. Failure rates and associated risk factors of orthodontic miniscrew implants: a meta-analysis. Am J Orthod Dentofacial Orthop 2012;142:577-95.e7.

Pearson LE. Vertical control in treatment of patients having backward-rotational growth tendencies. Angle Orthod 1978;48:132-40.

Pearson LE. Vertical control in fully-banded orthodontic treatment. Angle Orthod 1986;56:205-24.

Pedrin F, Almeida MR, Almeida RR, Almeida-Pedrin RR, Torres F. A prospective study of the treatment effects of a removable appliance with palatal crib combined with high-pull chincup therapy in anterior open-bite patients. Am J Orthod Dentofacial Orthop 2006;129:418-23.

Petit HP. Syndromes prognathiques: schemas de traitement "global" autour de masks faciaux. Rev Orthop Dento Faciale 1982;16:381-411.

Pigott TD. Advances in Meta-Analisis. Chicago: Springer, 2012:49-51.

Pildal J, Chan AW, Hróbjartsson A, Forfang E, Altman DG, Gøtzche PC. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. BMJ 2005;330:1049.

Porta M. A dictionary of epidemiology. New York: Oxford University Press, 2008.

Proffit WR, Fields HW. Contemporary Orthodontics. US: CV Mosby, 2000. Figure 6 from page 272, Copyright Elsevier.

Rakosi T. Treatment of Class III malocclusions. In Graber T, Rakosi T, Petrovic A, Dentofacial Orthopedics with Functional Appliances. US: CV Mosby, 1997:461-80.

Ritucci R, Nanda R. The effect of chin cup therapy on the growth and development of the cranial base and midface. Am J Orthod Dentofacial Orthop 1986;90:475-83.

Rakosi T, Schilli W. Class III anomalies: a coordinated approach to skeletal, dental and soft tissue problems. J Oral Surg 1981;39:860-70.

Straus SE, Richardson WS, Glasziou P, Haynes RB, eds. Evidence-based medicine: how to practice and teach EBM. Philadelphia: Elsevier, 2005.

Sugawara J. Clinical practice guidelines for developing Class III malocclusion. In Nanda R, Biomechanics and esthetic strategies in clinical orthodontics. US: Saunders, 2005:211-63. Figure 2 from page 213, Copyright Elsevier.

Sugawara J, Mitani H. Facial growth of skeletal Class III malocclusion and the effects, limitations, and long-term dentofacial adaptations to chincap therapy. Semin Orthod 1997;3:244-54.

Thilander B. Chin-cap treatment for Angle Class III malocclusion. Rep Congr Eur Orthod Soc 1965;41:311-27.

Thilander B, Myrberg N. The prevalence of malocclusion in Swedish school children. Scan J Dent Res 1973;81:12-20.

Torres F, Almeida RR, de Almeida MR, Almeida-Pedrin RR, Pedrin F, Henriques JF. Anterior open bite treated with a palatal crib and high-pull chin cup therapy. A prospective randomized study. Eur J Orthod 2006;28:610-7.

Tortop T, Keykubat A, Yuksel S. Facemask therapy with and without expansion. Am J Orthod Dentofacial Orthop 2007;132:467-74.

Tschill P, Bacon W, Sonko A. Malocclusion in the deciduous dentition of Caucasian children. Eur J Orthod 1997;19:361-7.

Tuncer BB, Kaygisiz E, Tuncer C, Yüksel S. Pharyngeal airway dimensions after chin cup treatment in Class III malocclusion subjects. J Oral RehabIl 2009;36:110-7.

Turpin DL. Early Class III treatment, unpublished thesis presented at 81st session. Am Assoc Orthod San Francisco, 1981.

Ulgen M, Firatli S. The efects of Frankel's function regulator on the Class III malocclusion. Am J Orthod 1994;105:561-7.

Victor N. The challenge of meta-analysis: discussion. Indications and contra-indications for meta-analysis. J Clin Epidemiol 1995;48:5-8.

Wahabuddin S. Determinants of successful chincup therapy in skeletal Class III malocclusions. Am J Orthod Dentofacial Orthop 2005;127:2-3.

Williams S, Anderson CE. The morphology of the potential Class III skeletal pattern in the growing child. Am J Orhtod 1986;89:301-11.

Williams S, Bundgaard P. Extraoral orthodontic appliances. Aarhus: The Royal Dental College, 1985:20.

Yüksel S, Uçem TT, Keykubat A. Early and late facemask therapy. Eur J Orthod 2001;23:559-68.

Appendix Table 1. Articles excluded on basis of full-text from this study and reason for exclusion.

Nr	*Study**	*Reason for exclusion*
1	Abu Alhaija ES, Richardson A. Long-term effect of the chincap on hard and soft tissues. Eur J Orthod 1999;21:291-8.	Study with maxillary protraction appliance and simultaneous use of chin cup.
2	Alarcón JA, Bastir M, Rosas A, Molero J. Chincup treatment modifies the mandibular shape in children with prognathism. Am J Orthod Dentofacial Orthop 2011;140:38-43.	Geometric and morphometric assessment without cephalometric values.
3	Allen RA, Connolly IH, Richardson A. Early treatment of Class III incisor relationship using the chincap appliance. Eur J Orthod 1993;15:371-6.	Study with maxillary protraction appliance and simultaneous use of chin cup.
4	Arat ZM, Akçam MO, Gökalp H. Long-term effects of chin-cap therapy on the temporomandibular joints. Eur J Orthod 2003;25:471-5.	Inadequate data: not given cephalometric values post treatment for control and for treated group.
5	Arman A, Toygar TU, Abuhijleh E. Profile changes associated with different orthopedic treatment approaches in Class III malocclusions. Angle Orthod 2004;74:733-40.	Study without contemporaneous and matching in Angle Class control sample.
6	Arun T, Erverdi N. A cephalometric comparison of mandibular headgear and chin-cap appliances in orthodontic and orthopaedic view points. J Marmara Univ Dent Fac 1994;2:392-8.	Case series without a control.
7	Asensi Cros C: Analysis de prognatismos mandibulares tratados ortodoncicamente: Estudio mediante cefalometria [Dissertation]. Valencia (PA): Univ. de Valencia (Spain), 1987.	Uncontrolled retrospective clinical trial.
8	Baccetti T, Franchi L, Schulz SO, McNamara JA Jr.Treatment timing for an orthopedic approach to patients with increased vertical dimension. Am J Orthod Dentofacial Orthop 2008;133:58-64.	Study on Class I malocclusion.
9	Basdra EK, Stellzig A, Komposch G. Dentofacial changes in patients with Class III malocclusions treated by a combination of activator and chin-cup appliances. Aust Orthod J 1997;14:225-8.	Study without contemporaneous and matching in Angle Class control sample.
10	Bawazeer OA: Post-treatment stability of class III treatment using RPE and heavy force chin cup protraction [Dissertation]. Illinois (PA): Univ of Illinois at Chicago, 2005.	Uncontrolled retrospective clinical trial.
11	Burns NR: Class III camouflage treatment: A retrospective study [Dissertation]. West Virginia (PA): Univ. of West Virgina, 2008.	Study without contemporaneous and matching in Angle Class control sample.

12	De Alba A, Chaconas J, Caputo A, Emison W. Stress distribution under high pull extraoral chin cup traction. A photoelastic study. Angle Orthod 1982;52:69-78.	Technique description.
13	De Alba A, Chaconas J, Caputo A.Orthopedic effect of the extraoral chin cup appliance on the mandible. American Journal of Orthodontics 1976;69:29-41.	Technique description.
14	Deguchi T, Kitsugi A. Stability of changes associated with chin cup treatment. Angle Orthod 1996;66:139-45.	Results given after full orthodontic treatment including chin cup and fixed appliances.
15	Deguchi T, Kuroda T, Hunt NP, Graber TM. Long-term application of chincup force alters the morphology of the dolichofacial Class III mandible. Am J Orthod Dentofacial Orthop 1999;116:610-5.	Results given after full orthodontic treatment including chin cup and fixed appliances.
16	Deguchi T, Kuroda T, Minoshima Y, Graber TM. Craniofacial features of patients with Class III abnormalities: growth-related changes and effects of short-term and long-term chin-cup therapy. Am J Orthod Dentofacial Orthop 2002;121:84-92.	Results given after full orthodontic treatment including chin cup and fixed appliances.
17	Deguchi T, McNamara JA. Craniofacial adaptations induced by chincup therapy in Class III patients. Am J Orthod Dentofacial Orthop 1999;115:175-82.	Study with maxillary protraction appliance and simultaneous use of chin cup.
18	Deguchi T, Uematsu S, Kawahara Y, Mimura H. Clinical evaluation of temporomandibular joint disorders (TMD) in patients treated with chin cup. Angle Orthod 1998;68:91-4.	Uncontrolled retrospective clinical trial.
19	Gavakos K, Witt E. The functional status of orthodontically treated prognathic patients. Eur J Orthod 1991;13:124-8.	Uncontrolled retrospective clinical trial.
20	Gökalp H, Arat M, Erden I. The changes in temporomandibular joint disc position and configuration in early orthognathic treatment: a magnetic resonance imaging evaluation. Eur J Orthod 2000;22:217-24.	Study without matching in Angle Class control sample.
21	Graber W. Hyoid changes following orthopedic treatment of mandibular prognathism. Angle Orthod 1978;48:33-8.	Uncontrolled retrospective clinical trial.
22	Graber W. Chin cup therapy for mandibular prognathism. Am J Orthod 1977;72:23-41.	Uncontrolled prospective clinical trial.
23	Harrison E, Shaw C, Worthington V, Bickley S, Scholey J, O'Brien D. Orthodontic treatment for prominent lower front teeth in children [Protocol]. Cochrane Database of Systematic Reviews 2002, Issue 1.	Protocol of clinical procedures.

24	Heymann GC: Temporary anchorage devices for maxillary protraction: Three-dimensional analysis of treatment outcomes [Dissertation]. North Carolina (PA): The Univ. of North Carolina at Chapel Hill, 2008.	Dissertation on maxillary protraction appliance and simultaneous use of chin cup.
25	Lentini-Oliveira D et al. Orthodontic and orthopaedic treatment for anterior open bite in children. Cochrane Database Syst Rev 2000;18:CD005515.	Systematic review.
26	Lin HC, Chang HP, Chang HF. Treatment effects of occipitomental anchorage appliance of maxillary protraction combined with chincup traction in children with Class III malocclusion. J Formos Med Assoc 2007;106:380-91.	Study with maxillary protraction appliance and simultaneous use of chin cup.
27	Liu ZP, Li CJ, Hu HK, Chen JW, Li F, Zou SJ. Efficacy of short-term chincup therapy for mandibular growth retardation in Class III malocclusion.Angle Orthod 2011;81:162-8.	Meta-analysis.
28	Lu YC, Tanne K, Hirano Y, Sakuda M. Craniofacial morphology of adolescent mandibular prognathism. Angle Orthod 1993;63:277-82.	Study with inadequate data: not given cephalometric values pre and post treatment for control or for treated group., just changes given schematically.
29	Luther F, Layton S, McDonald F. Orthodontics for treating temporomandibular joint (TMJ) disorders. Cochrane Database of Systematic Reviews 2010, Issue 7.	Systematic review.
30	Matsumoto M, Kanaya M, Itoh T, Kawagoe H, Kamogashira K, Ichikawa K, Himeno R, Hirose T, Kawai S, Yanabu M, et al. Statistical investigation on orthodontic patients who received treatment at Fukuoka Dental College during 15 years. Fukuoka Shika Daigaku Gakkai Zasshi 1989;16:485-94.	Investigation not relevant to the study.
31	Mimura H, Deguchi T. Morphologic adaptation of temporomandibular joint after chincup therapy. Am J Orthod Dentofacial Orthop 1996;110:541-6.	Results given after full orthodontic treatment including chin cup and fixed appliances.
32	Murata S, Iwata R, Hayakawa S, Fuwa Y, Goto S, Suzuki N. Long term results of skeletal profile changes occurring from chin cap therapy of Japanese female skeletal Class III cases. Aichi Gakuin Daigaku Shigakkai Shi 1990;28:573-9.	Uncontrolled retrospective clinical trial.
33	Oda H, Ueda A, Miyagawa Y, Hongou H, Kuroki K, Suzuki M, Suzuki Y. Individual growth study of the effects of chin cap force to the mandible. Nippon Kyosei Shika Gakkai Zasshi 1989 Jun;48:355-61.	Uncontrolled retrospective clinical trial.

34	Pedrin F, Almeida MR, Almeida RR, Almeida-Pedrin RR, Torres F. A prospective study of the treatment effects of a removable appliance with palatal crib combined with high-pull chin-cup therapy in anterior open-bite patients. Am J Orthod Dentofacial Orthop 2006;129:418-23.	Study on Class I malocclusion.
35	Ritucci R, Nanda R. The effect of chin cup therapy on the growth and development of the cranial base and midface. Am J Orthod Dentofacial Orthop 1986;90:475-83.	Study with inadequate data: not given cephalometric values post treatment for control and for treated group.
36	Sakamoto T, Iwase I, Uka A, Nakamura S. A roentgenocephalometric study of skeletal changes during and after chin cup treatment. Am J Orthod 1984;85:341-50.	Study without matching in age control sample.
37	Sugawara J, Asano T, Endo N, Mitani H. Long-term effects of chincap therapy on skeletal profile in mandibular prognathism. Am J Orthod Dentofacial Orthop 1990;98:127-33.	Uncontrolled prospective clinical trial.
38	Tahmina K, Tanaka E, Tanne K. Craniofacial morphology in orthodontically treated patients of class III malocclusion with stable and unstable treatment outcomes. Am J Orthod Dentofacial Orthop 2000;117:681-90.	Uncontrolled prospective clinical trial.
39	Toffol LD, Pavoni C, Baccetti T, Franchi L, Cozza P. Orthopedic treatment outcomes in Class III malocclusion. A systematic review. Angle Orthod 2008;78:561-73.	Systematic review.
40	Torres F, Almeida RR, de Almeida MR, Almeida-Pedrin RR, Pedrin F, Henriques JF. Anterior open bite treated with a palatal crib and high-pull chin cup therapy. A prospective randomized study. Eur J Orthod 2006;28:610-7.	Study on Class I malocclusion.
41	Wendell PD, Nanda R, Sakamoto T, Nakamura S. The effects of chin cup therapy on the mandible: a longitudinal study. Am J Orthod 1985;87:265-74.	Retrospective control clinical trial with non consecutive selection of patients. Inadequate data: not given cephalometric values post treatment for control and for treated group.
42	Thilander B. Chin cup treatment for Angle Class III maloclusion: a longitudinal study. Trans Eur Orthod Soc 1965;41:311-27.	Uncontrolled retrospective clinical trial.
43	Thilander B.Treatment of Angle Class III malocclusion with chin cup. Trans Eur Orthod Soc 1963;39:384-98.	Uncontrolled retrospective clinical trial.

Appendix Table 2. The PRISMA statement checklist.

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	9-15
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	16-55
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	57-59
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	61
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	62-69
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	61-69
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	61-65
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	66-69

Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	70
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	71-72
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	70-71
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	71-72
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	71-75

Page 1 of 2

Section/topic	**#**	**Checklist item**	**Reported on page #**
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	73-74
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	75
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	77-79, 198-201
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	80-91
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	92-106
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	106-168

Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	108-168
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	108-168
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	110-168
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	171-179
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	169-171
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	181
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	185

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

Page 2 of 2