The effect of headgear treatment on the development of obstructive sleep apnoea. A systematic review

Vera Studer,* Despina Koletsi,† Anna Iliadi‡ and Theodore Eliades†
Center for Dental Medicine, University of Zurich, Switzerland,* Clinic of Orthodontics and Paediatric Dentistry, Center for Dental Medicine, University of Zurich, Switzerland,† London School of Hygiene and Tropical Medicine, University of London, United Kingdom† and Private Practice limited to Orthodontics, Athens, Greece‡

Aim: To evaluate the effect of the cervical headgear on the development of obstructive sleep apnoea and subsequent alterations of oropharyngeal dimensions.

Materials and method: An electronic database search of published and unpublished literature was performed (MEDLINE, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Clinical Trials.gov and National Research Register). Search terms included obstructive sleep apnoea, sleep disorders, pharyngeal dimensions and headgear. A risk of bias assessment was conducted using the ACROBAT-NRSI tool for non-randomised studies.

Results: Of the 51 articles initially retrieved, only three were eligible for inclusion, while the remainder were retrospective cohort studies presenting serious risk of bias primarily due to undetected confounding factors or selection bias. No quantitative synthesis was possible. One study assessed the potential effect of isolated headgear treatment on apnoeic indices, while two studies described pharyngeal airway dimensions after the use of headgear alone or in combination with an activator appliance. Overall, increased apnoeic indices and the oxygen desaturation index were detected for headgear users. Dimensional changes in the posterior airway space were comparable after headgear or activator use, while combined headgear-activator treatment led to an increase in posterior pharyngeal area when compared with isolated fixed appliance therapy.

Conclusions: Due to methodological inconsistencies and apparent risk of bias of the existing studies, no robust conclusions can be drawn. Prospective controlled or randomised controlled trials are deemed necessary to provide evidence on the effect of headgear treatment on sleep apnoea or pharyngeal airway dimensions.

(Aust Orthod J 2018; 34: 239-249)

Introduction

Obstructive sleep apnoea syndrome (OSAS) is common and aligns within the sleep disordered breathing (SDB) spectrum. It is characterised by repetitive episodes of complete respiratory upper airway obstruction during sleep, which is associated with a reduction in blood oxygen saturation, loud snoring, sleep arousal or awakenings, a cessation of breathing and, in severe cases, cyanosis. Upon awakening, patients typically feel weariend and may describe feelings of disorientation, gogginess, mental dullness and incoordination. The prevalence of OSAS has been estimated to be 4% for men, 2% for women and 1–5.7% in the paediatric population. A predominance of men suffering from OSAS coupled with an increased risk in obese patients has been reported. Predisposing factors in children are nasopharyngeal abnormalities, as well as hypertrophied tonsils and adenoids that narrow the upper airway. Allergies, asthma and an excessive volume of soft tissues in obese children have been recorded as risk factors.
Current task force initiatives on the diagnosis and management of OSAS in 2-to-18-year-old children suggest primary identification and elimination of potential pre-disposing abnormalities, stepwise re-evaluation to detect residual disease and, finally, an evaluation of the need for additional treatment.\textsuperscript{10} Skeletal jaw relationships and anatomical variations within the maxillomandibular complex may reflect a particular pattern of the oropharyngeal apparatus. It has been claimed that individuals with OSAS may demonstrate maxillary and mandibular retrognathism, larger craniofacial angles, reduced upper airway space, a longer and thicker palate and a low hyoid bone position.\textsuperscript{11-12,13,14} However, no causal relationship between these craniofacial characteristics and obstruction of pharyngeal airway has yet been established.\textsuperscript{15} A number of orthodontic appliances have been used to correct a skeletal jaw relationship, of which cervical headgear has been a common option.\textsuperscript{16} Major indications for headgear use are maxillary space deficiencies, a Class II molar and skeletal relationship and to augment anchorage. Cervical headgear use may have a profound effect on the antero-posterior growth of the maxilla, while indirectly affecting mandibular growth.\textsuperscript{17} Whether widening of the maxilla has a favourable effect on sagittal mandibular development is less clear.\textsuperscript{18} Some studies suggest an upward and forward rotation of the mandible,\textsuperscript{19, 20} which could lead to an increase in upper airway space.\textsuperscript{21, 22} Alternative studies indicate that a forward growth restriction of the maxilla through cervical headgear use\textsuperscript{27, 23} will not result in an improvement of upper oropharyngeal airway dimensions or, worse, may result in an aggravation of breathing problems.\textsuperscript{24} While much of the existing literature has paid special attention to the effect of cervical headgear on Class II correction,\textsuperscript{25-26,27,28,29} there is limited evidence with regard to the potential side effects of headgear use, especially related to sleep disorders and associations with upper airway dimensions. Therefore, the aim of the present systematic review was to evaluate the effect of cervical headgear on the potential development of obstructive sleep apnoea and subsequent alterations in the oropharyngeal dimensions in young patients.

**Material and methods**

The Guidelines for Meta-Analyses and Systematic Reviews of Observational Studies (MOOSE)\textsuperscript{30} were followed for the reporting of this systematic review.

**Eligibility criteria**

The following selection criteria were applied for this review:

- **Study design:** Randomised Controlled Trials (RCTs), Controlled Clinical Trials (CCTs) and observational studies including a comparison group (cohort-type, case-control) were considered.
- **Participants:** Children or adolescent patients undergoing orthodontic treatment using headgear.
- **Interventions:** Any type of headgear appliance alone or in combination with other fixed or removable appliances.
- **Comparators:** Appliances other than headgear, or untreated control groups.
- **Outcome measures:** Changes in dimensions related to the pharyngeal airway or apnoea indicators. Both conventional cephalometric measurements and cone beam computed tomography (CBCT) radiography were considered, where relevant. These included but were not confined to: the distance between the soft palate and posterior pharyngeal wall, hyoid bone measurements, apnoeic index, oxygen desaturation, and the number and duration of apnoeic episodes.
- **Exclusion criteria:** Studies involving patients with systematic or other diseases undergoing orthodontic treatment, studies involving adult patients (>18 years of age), and cohort studies without a comparison group.

**Search strategy**

An electronic search within the following databases was undertaken in July, 2016 and updated in December, 2017 for additional potential reports, without language restrictions: Medline via Pubmed, Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials (CENTRAL). Moreover, unpublished literature was searched in ClinicalTrials.gov (www.clinicaltrials.gov) and the National Research Register (www.controlled-trials.com), using the terms «headgear» AND «apnoea». Hand searching of the reference lists of the retrieved full text articles was also conducted. The authors of original studies were contacted for data clarification when needed. The full search strategy employed in Medline via Pubmed is presented in Appendix 1.
An eligibility assessment, data extraction and Risk of Bias (RoB) assessment was implemented independently and, in duplicate, by two reviewers (VS and DK). Disagreements were resolved through discussion and after consultation with a third author (AI).

**Data extraction**

Data extraction was performed on standardised pre-piloted forms by two independently-working reviewers (VS and DK) who were not blinded to author identity nor study origin. The titles and abstracts were examined first, followed by full text screening for the possibility of including articles. Information was obtained from each included study on its design, observation period and methods, participants, interventions, comparators and outcomes.

**Risk of bias within studies**

The risk of bias in individual studies was assessed according to the ACROBAT-NRSI tool as described by the Cochrane Collaboration. In particular, the following domains were considered: (1) bias due to confounding, (2) bias in the selection of participants in the study, (3) bias in the measurement of interventions, (4) bias due to departures from intended interventions, (5) bias due to missing data, (6) bias in the measurement of outcomes, (7) bias in the selection of the reported result. An overall assessment of the risk of bias was made for each included study (critical, serious, moderate, low, or no information). Studies receiving an assessment indicating a critical risk of bias in several domains were considered a critical risk of bias, while those with at least one item were designated to be at a serious risk of bias. Reports with a moderate risk of bias for one or more key domains were considered to be at moderate risk of bias, while those with low risk of bias in all domains were rated a low risk. No information corresponded to domains in which there was no information on which to base a bias judgement.

**Summary measures and data synthesis**

The clinical heterogeneity of the included studies was assessed through the examination of individual trial settings, eligibility criteria, appliances used and data collection methods. Statistical heterogeneity was to be examined through visual inspection of the confidence intervals (CIs) for the estimated treatment effects on forest plots. Also, a chi-square test was to be applied to assess heterogeneity; a $p$-value below the level of 10% ($p < 0.1$) was considered indicative of significant heterogeneity. An $I^2$ test for homogeneity was also to be undertaken to quantify the extent of heterogeneity. Only studies at a moderate or low risk of overall bias were intended to be included in meta-analyses. Random effects meta-analyses were conducted, as they were considered more appropriate to better approximate expected variations in trial settings. Treatment effects were calculated through pooled standardised mean differences (SMD) in cephalometric/CBCT measurement changes along with associated 95% Confidence Intervals (95% CIs) and Prediction Intervals where applicable (at least three trials needed). For binary outcomes, Odds Ratios (OR) were considered.

### Appendix 1

**MEDLINE search (via Pubmed)**

- **Limits:** ‘Humans’, no language restriction applied
- **Publication date:** no restriction
- **Search Builder:** ‘All Fields’

Two consecutive searches combined with “AND” Boolean operator, using “OR” between MeSH terms or keywords:

1. headgear
2. cervical headgear
3. headgear appliance
4. extraoral traction
5. 1 OR 2 OR 3 OR 4
6. apnea
7. obstructive sleep apnea
8. apnoea
9. obstructive sleep apnoea
10. sleep disorder
11. pharyngeal airway
12. pharyngeal dimension
13. oropharyngeal airway
14. oropharyngeal dimension
15. nasopharyngeal airway
16. nasopharyngeal dimension
17. 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16
18. 6 AND 17
Risk of bias across studies

If more than 10 studies were included in the meta-analysis, publication bias was to be explored through standard funnel plots.32

Additional analyses

Sensitivity analyses were predetermined to explore and isolate the effect of studies with a moderate risk of bias on the overall treatment effect if both low and moderate risk of bias studies were included.

Results

Study selection

The initial search yielded 51 studies. A total of six studies24,33,34,35,36,37 were left for full text evaluation and potential inclusion in the review. Three of these studies were rejected after full text screening, due to an absence of comparator groups or irrelevant outcomes in relation to the formulated question of the present review. Finally, three studies24,33,34 were included in the qualitative synthesis.

Table I. Characteristics of included studies

| Origin | Godt (2011) | Department of Orthodontics, Eberhard-Karls-University in Tübingen, Germany |
|--------|-------------|---------------------------------------------------------------------------|
| Design | Retrospective cohort study |
| Observation Period / Methods | Single investigator evaluated cephalograms at baseline and end of treatment Overall treatment duration: 5.5–6.4 years |
| Participants | Class I & II patients from a private orthodontic office without history of sleep disorders. Group 1: 209 (average age at baseline: 11.24 years) Group 2: 50 (average age at baseline: 9.27 years) Group 3: 49 (average age at baseline: 10.38 years) |
| Interventions | Group 1: headgear + multibracket appliance Group 2: activator + multibracket appliance Group 3: bite-jumping appliance + headgear + multibracket appliance |
| Outcomes | A. Short term after appliance therapy (first treatment phase) B. Long term after full orthodontic therapy (overall) 1. PAS-NL (posterior airway space – nasal line) 2. PAS-OcclPl (posterior airway space – occlusion plane) 3. PAS-Uvula (posterior airway space – uvula) 4. PAS-ML (posterior airway space – mandibular line) 5. Hyoid (Hyoid – mandibular base) |

| Hanggi (2008) | |
| Origin | Clinic for Orthodontics and Pediatric Dentistry; Institute of Social and Preventive Medicine; Biostatistics Unit; University of Zurich, Switzerland |
| Design | Retrospective cohort study |
| Observation Period / Methods | Evaluation of cephalograms before treatment (T1), at the end of active therapy (T2) and long-term follow-up (T3) Duration: Active treatment: ca. 4 years Post-treatment: ca. 7.5 years |
| Participants | 64 children (32 male, 32 female) without Class III, extractions, space closure, malocclusion, deep/open bite, orthognathic surgery, rapid maxillary expansion, remaining growth. Study group: 32 Control group: 32 Baseline: Between 9 and 14 years of age |
| Interventions | Study group: Class II therapy with activator-HG for at least 9 month + fixed appliance therapy Control group: minor orthodontic treatment (without headgear, activator or Class II elastics) |
THE EFFECT OF HEADGEAR USE ON OBSTRUCTIVE SLEEP APNOEA

5. p (mm): smallest distance between soft palate & posterior pharyngeal wall
6. w (mm): largest distance between p & t
7. t (mm): smallest distance between tongue base & posterior pharyngeal wall
8. length (mm): distance between most cranial point of pharynx and a.
9. area (mm²): area between p & a (a: parallel line to FH through most anterior inferior point of C4)

Pirilä-Parkkinen (1999)

Origin
Institute of Dentistry, University of Oulu, Finland; Department of Otolaryngology, Oulu University Hospital; Health Centre of Oulu; Karolinska Institute, Faculty of Odontology, Stockholm, Sweden

Design
Retrospective cohort study

Observation
A polygraphic sleep evaluation measured the OSAS tendency. All participants sleep one night under laboratory conditions, those with HG spending first half of night with HG and after without.

Period / Methods
Cephalograms prior to therapy and corresponding control group were analysed.

Participants
30 children (12 male, 18 female; 8.2 ± 1.61 years)
Group 1: 10 children with OSAS-Symptoms while using headgear
Group 2: 10 healthy children
Group 3: 10 children with OSAS

Interventions
Group 1: headgear while first half of the night
Group 2: no orthodontic treatment
Group 3: no orthodontic treatment

Outcomes
1. Apnea index
2. Number of apnea and hypopnea periods
3. Number of obstructive apnea
4. Number of central apnea periods
5. Number of mixed apnea periods
6. Total apnea time (%)
7. Time per apnea /hypopnea episode (s)
8. ODI4/h (oxygen desaturation index; ≥4% decrease from control level in silent rest)
9. ODI4/night (oxygen desaturation index; ≥4% decrease from control level in silent rest)
10. ODI10/night (oxygen desaturation index; ≥10% decrease from control level in silent rest)
11. Total sleeping time (h)
12. Time spent sleeping supine (%)
13. SaO2 90-10% (difference between 90 and 10 percentile oxygen saturation values for whole night)
14. Mean value for oxygen saturation (%)

Study characteristics
The three selected studies were controlled, all published in English and classified as retrospective cohort studies.

Godt et al.33 and Hänggi et al.34 reported on pharyngeal dimensions after headgear treatment in Class I or Class II patients. Godt et al.33 also investigated whether pharyngeal narrowness was expected in conjunction with differential vertical growth patterns. Hänggi et al.34 compared the physiological changes in the pharyngeal area of healthy individuals after the use of a combined type of headgear and activator treatment. The total sample size used in both studies was 372 patients, with an age range of 9.3 to 11.2 years at the beginning of the study. Conventional cephalometric radiographs were assessed and compared at baseline and at the end of the active phase of treatment in both studies to record alterations in the pharyngeal airway.

Pirilä-Parkkinen et al.24 reported a mixed population of 30 children who presented with apnoea, and who were either healthy or headgear users with an average initial age of 8.2 years. The proportions of breathing abnormalities, apnoea indices and oxygen saturation were recorded under laboratory conditions (Table I).
Table II. Risk of bias assessment (ACROBAT-NRSI overview).

| Bias                     | Godt (2011) | Hänggi (2008) | Pirilä-Parkkinen (1999) |
|--------------------------|-------------|---------------|-------------------------|
| Confounding              | Serious     | Moderate      | Serious                 |
| Selection of participants| Serious     | Serious       | Serious                 |
| Measurement of intervention| Serious   | Serious       | Serious                 |
| Departures from intended interventions | Low      | Low           | Low                     |
| Missing data             | Low         | Low           | Low                     |
| Measurement of outcomes  | Moderate    | Moderate      | Low                     |
| Selection of Reported result | Low         | Low           | Low                     |
| Overall                  | Serious     | Serious       | Serious                 |

**Risk of bias within studies**

Based on the assessment of ACROBAT-NRSI, the three studies acquired a rating ‘serious risk of bias’ in the overall risk of bias judgement. To achieve a selection of participants unrelated to intervention was difficult, since headgear is normally used in Class II patients and, therefore, children with such malocclusions were selected for these studies. Confounding was another parameter that yielded a serious risk of bias, as all the studies were retrospective. In addition, a number of baseline factors could possibly affect the association between headgear use and airway dimensions or breathing conditions (i.e., cephalometric/CBCT measurements that differed between the study groups), while none were taken into account during statistical analysis. Parameters associated with missing data and selection of reported outcomes/results were less prone to bias, as complete patient records were recorded for all follow-up time-points and all reported outcomes had been pre-specified in the article methodologies (Table II; Table III).

**Results of individual studies**

In an assessment of nasopharyngeal dimensions, the use of headgear was found to present reduced posterior airway space at all levels from the nasopharynx to the mandibular base at the initial headgear phase of treatment; however, this finding was also reported for other occlusion-modifying, first phase appliances. These changes were not related to different vertical patterns of growth. The combined use of an activator-headgear appliance revealed increases in pharyngeal airway parameters related to area, length and the smallest distance between the tongue and the posterior pharyngeal wall after the active phase of treatment. In the post-treatment period, changes in nasopharyngeal dimensions were established and were of the same amount as in the untreated control group.

Related to breathing pattern and apnoea symptoms, the headgear group presented an increased oxygen desaturation index (ODI$^{10}$) when compared with the sample of apnoea children. In addition, other apnoea related parameters such as apnoea indices, obstructive, central and mixed apnoea periods, and total apnoea time were elevated during sleeping with a headgear appliance. However, headgear patients were those previously reported as having presented at least one apnoea symptom and were selected based on this characteristic.

**Discussion**

Cervical headgear is a common orthodontic appliance used for managing space problems, the correction of Class II skeletal and molar relationship and anchorage reinforcement. The effect of cervical headgear on the maxilla and/or the mandible has been reviewed by previous studies. Nevertheless, only a small number of publications have evaluated the potential side effects of cervical headgear on sleep disorders and pharyngeal airway space. Therefore, this systematic review was designed to provide clear evidence related to the development of obstructive sleep apnoea and subsequent alterations in the naso- and oropharyngeal dimensions in young patients during the use of headgear as part of orthodontic treatment.

Only three studies were related to the research question and were included in this systematic review, which is indicative of the scarcity of publications.
Risk of bias assessment (cohort-type studies)

1. Bias due to confounding

1.1. Is confounding of the effect of intervention unlikely in this study?
   Godt (2011) 1.1. No
   Hänggi (2008) 1.1. Yes
   Pirilä-Parkinen (1999) 1.1. No

If Y or PY to 1.1.

1.2. Were participants analysed according to their initial intervention group throughout follow up?
   If N or PN to 1.2.

1.3. Were intervention discontinuations or switches unlikely to be related to factors that are prognostic for the outcome?
   If Y or PY to 1.2., or Y or PY to 1.3.

1.4. Did the authors use an appropriate analysis method that adjusted for all the critically important confounding domains?
   If Y or PY to 1.4.

1.5. Were confounding domains that were adjusted for measured validly and reliably by the variables available in this study?
   If Y or PY to 1.5.

1.6. Did the authors avoid adjusting for post-intervention variables?
   If N or PN to 1.2. and 1.3.

2. Bias in selection of participants into the study

2.1. Was selection into the study unrelated to intervention or unrelated to outcome?
   2.1. No
   (Unrelated to outcome, but subjects have been selected due to Class II malocclusion)

2.2. Do start of follow-up and start of intervention coincide for most subjects?
   If N or PN to 2.1. or 2.2.

2.3. Were adjustment techniques used that are likely to correct for the presence of selection biases?
   2.3. No

Risk of bias judgement

Table III. The ACROBAT-NRSI tool: risk of bias based on signalling questions: for each study/cohort type studies.

|                      | Godt (2011) | Hänggi (2008) | Pirilä-Parkinen (1999) |
|----------------------|-------------|---------------|------------------------|
| Risk of bias assessment (cohort-type studies) | | | |
| 1. Bias due to confounding | | | |
| 1.1. Is confounding of the effect of intervention unlikely in this study? | 1.1. No | 1.1. Yes | 1.1. No |
| If Y or PY to 1.1. | | | |
| 1.2. Were participants analysed according to their initial intervention group throughout follow up? | 1.2. - | 1.2. Yes (No treatment switches known) | 1.2. - |
| If N or PN to 1.2. | | | |
| 1.3. Were intervention discontinuations or switches unlikely to be related to factors that are prognostic for the outcome? | 1.3. - | 1.3. - | 1.3. - |
| If Y or PY to 1.2., or Y or PY to 1.3. | | | |
| 1.4. Did the authors use an appropriate analysis method that adjusted for all the critically important confounding domains? | 1.4. - | 1.4. Probably Yes (adjusted multiple linear regression) | 1.4. No |
| If Y or PY to 1.4. | | | |
| 1.5. Were confounding domains that were adjusted for measured validly and reliably by the variables available in this study? | 1.5. - | 1.5. Probably Yes (Cephalograms were evaluated by an investigator: Subjective measure, measurement error analysis implemented) | 1.5. - |
| 1.6. Did the authors avoid adjusting for post-intervention variables? | 1.6. No / No information | 1.6. Yes | 1.6. No / No information |
| If N or PN to 1.2. and 1.3. | | | |
| 1.7. Did the authors use an appropriate analysis method that adjusted for all the critically important confounding domains and for time-varying confounding? | 1.7. - | 1.7. - | 1.7. - |
| If Y or PY to 1.7. | | | |
| 1.8. Were confounding domains that were adjusted for measured validly and reliably by the variables available in this study? | 1.8. - | 1.8. - | 1.8. - |
| Risk of bias judgement | Serious risk of bias | Moderate risk of bias | Serious risk of bias |
### 3. Bias in measurement of interventions

| 3.1 Is intervention status well defined? | 3.2 Was information on intervention status recorded at the time of intervention? | 3.3 Was information on intervention status unaffected by knowledge of the outcome or risk of the outcome? |
|----------------------------------------|---------------------------------------------------------------------|--------------------------------------------------|
| 3.1. Yes (Group 1/2: 6 month HG (14h per day) / activator, then brackets; Group 3: First BJA, then HG) | 3.2. Yes | 3.3. Probably No |
| Risk of bias judgement | Serious risk of bias | Serious risk of bias |

### 4. Bias due to departures from intended interventions

| 4.1 Were the critical co-interventions balanced across intervention groups? | 4.2 Were numbers of switches to other interventions low? | 4.3 Was implementation failure minor? |
|-------------------------------------------------------------------------|------------------------------------------------------|-----------------------------------|
| 4.1. Yes (There are no co-interventions) | 4.2. Yes (There are no switches in this study) | 4.3. Probably Yes (Adherence of study participants: 4.4) |
| Risk of bias judgement | Low risk of bias | Low risk of bias |

### 5. Bias due to missing data

| 5.1 Are outcome data reasonably complete? | 5.2 Was intervention status reasonably complete for those in whom it was sought? | 5.3 Are data reasonably complete for other variables in the analysis? |
|----------------------------------------|---------------------------------------------------------------------|---------------------|
| 5.1. Yes | 5.2. Yes | 5.3. Yes |
| Risk of bias judgement | Low risk of bias | Low risk of bias |

| 5.4 Are the proportion of participants and reasons for missing data similar across interventions? | 5.5 Were appropriate statistical methods used to account for missing data? |
|------------------------------------------------------------------------------------------|--------------------------|
| 5.4. - | 5.5. - |
| Risk of bias judgement | Low risk of bias | Low risk of bias |
related to possible side-effects of extra-oral traction and the use of a headgear in growing patients. Of note, only one study was designed to specifically address whether headgear use aggravates obstructive sleep apnoea under laboratory conditions. Obstructive sleep apnoea is multifactorial in aetiology and its progression may result in a potentially life-threatening condition if not properly identified and addressed early in life. Therefore, the identification and isolation of potential triggering factors would be a significant step towards prevention of the disease.

Alterations in pharyngeal dimensions have been described after the use of growth modification appliances, including headgear treatment in children, as an additional and potential triggering factor in the development or aggravation of obstructive sleep apnoea. Alternatively, when evaluating children with confirmed OSA during early ages, the natural history of the disease should be considered, as a number of cases may resolve during adolescence.

Common practice when evaluating treatment outcomes in relation to oropharyngeal dimensions has been the assessment and interpretation of treatment effect though within group comparisons. Invariably, this involves a comparison against a baseline and over time, separately for each experimental group. However, the interpretation of the findings following such a comparison should be treated with caution as there have been associations with flawed inferences, increased likelihood of false positive errors or confounding of the outcome due to natural improvement over time.

---

| 6. Bias in measurement of outcomes | 6.1. Yes (Cephalograms were evaluated by one investigator) | 6.1. Yes (Cephalograms were evaluated by first author) | 6.1. Yes |
|-----------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------|
| 6.2. Were outcome assessors unaware of the intervention received by study participants? | 6.2. Probably No / No information | 6.2. Probably No / No information | 6.2. Yes (Blinded doctor) |
| 6.3. Were the methods of outcome assessment comparable across intervention groups? | 6.3. Yes (Same investigator evaluated all cephalograms) | 6.3. Yes (Same investigator evaluated all cephalograms) | 6.3. Yes (Same investigator evaluated all PGs) |
| 6.4. Were any systematic errors in measurement of the outcome unrelated to intervention received? | 6.4. Probably Yes | 6.4. Probably Yes | 6.4. Probably Yes |
| Risk of bias judgement | Moderate risk of bias | Moderate risk of bias | Low risk of bias |

---

| 7. Bias in selection of the reported result | 7.1. Probably Yes | 7.1. Probably Yes | 7.1. Probably Yes |
|-----------------------------------------------|--|--|--|
| Is the reported effect estimate unlikely to be selected, on the basis of the results, from … | 7.2. Probably Yes | 7.2. Probably Yes | 7.2. Probably Yes |
| 7.1. … multiple outcome measurements within the outcome domain? | 7.3. Probably Yes | 7.3. Probably Yes | 7.3. Probably Yes |
| 7.2. … multiple analyses of the intervention-outcome relationship? | Low risk of bias (unlikely to be manipulated) | Low risk of bias (unlikely to be manipulated) | Low risk of bias (unlikely to be manipulated) |
| Risk of bias judgement | Low risk of bias (unlikely to be manipulated) | Low risk of bias (unlikely to be manipulated) | Low risk of bias (unlikely to be manipulated) |

---

| 8. Overall bias | Serious risk of bias | Serious risk of bias | Serious risk of bias |
|------------------|----------------------|----------------------|----------------------|
| Risk of bias judgement | Serious risk of bias | Serious risk of bias | Serious risk of bias |
The apparently contradictory findings with regard to naso- and oro-pharyngeal airway dimensions in two of the included studies,\textsuperscript{33,34} notwithstanding the potential risk of bias and methodological shortcomings of each, are most likely related to the difference in the biological mechanisms activated by the different treatment procedures followed. The use of a combination of a headgear-activator appliance in one study\textsuperscript{34} might have been an influential factor inducing a more anterior repositioning of the lower jaw, resulting in an increase or a non-decrease of the pharyngeal dimensions. This might potentially counteract the initial effect of a headgear-only use and a reduced posterior airway space.

The ACROBAT-NRSI tool\textsuperscript{31} was used in the present review to identify the risk of bias involved in the included studies. Sterne et al., in 2014\textsuperscript{31} indicated that this tool used a number of signalling questions specifically designed to assess potential sources of bias that may arise in non-randomised studies, mainly represented by: issues on the selection of participants, confounding, bias in the measurement of the outcomes or selection of the reported results. Improvements have followed in the newest ROBINS-I tool\textsuperscript{41} designed and validated in 2016 and practically based on the same signalling questions. The Cochrane Collaboration has proposed the use of these tools over previous scales or scoring systems when evaluating non-randomised studies.\textsuperscript{42} The orthodontic literature is still new in the use of these tools, while the vast majority of systematic reviews have addressed the methodological quality of observational studies based on customised non-validated scales or tools. Modifications of the Newcastle-Ottawa scale\textsuperscript{43} have been described; however, these may consist of questions/items non-specific to the design of individual studies or lack relevant guidance on how to be used, resulting in varying interpretations considered by different investigators.\textsuperscript{44}

The review is not free of limitations, which are mainly dependent on the structure/design of the available studies. Only three studies, retrospective in nature, were deemed eligible for inclusion and all presented a serious risk of bias. Study design, confounding factors as well as the recruitment of participants and methods of outcome measurements/analyses were the primary determinants of bias detection in the included studies. No quantitative synthesis was possible in view of the different settings, populations and outcomes assessed and subsequently evaluation of publication bias was not possible, although pre-specified.

**Conclusion**

Based on the appraised literature, the evidence is not sufficiently solid to determine the effect of the cervical headgear appliance on the development of obstructive sleep apnoea and subsequent alterations of the naso- and oro-pharyngeal dimensions. Further prospective studies or randomised controlled trials are necessary to fill knowledge gaps and eliminate biases within the available evidence. The elimination of confounding factors and appropriate sample matching at the design level and at the analysis stage are important prerequisites when considering evidence from observational research. Only then can clear recommendations for future practice be identified to enable informed and optimal clinical decision making.

**Conflict of interest**

Nothing to declare

**Funding**

None

**Corresponding author**

Theodore Eliades
Plattenstrasse 11, CH-8032
Zurich
Email: theodore.eliades@zzm.uzh.ch

**References**

1. Obstructive Sleep Apnea Syndrome (780.53-0). In: The International Classification of Sleep Disorders (PDF). Westchester, Illinois: American Academy of Sleep Medicine, 2001;52-8,199.
2. Lee W, Nagubadi S, Kryger MH, Mokhlesi B. Epidemiology of obstructive sleep apnea: a population-based perspective. Expert Rev Respir Med 2008;2:349-64.
3. Bixler EO, Vgontzas AN, Lin HM, Liao D, Calhoun S, Vela-Bueno A et al. Sleep disordered breathing in children in a general population sample: prevalence and risk factors. Sleep 2009;32:731-6.
4. Li AM, So HK, Au CT, Ho C, Lau J, Ng SK et al. Epidemiology of obstructive sleep apnoea syndrome in Chinese children: a two-phase community study. Thorax 2010;65:991-7.
5. O’Brian LM, Holbrook CR, Mervis CB, Klaus CJ, Bruner JL, Raffield TJ et al. Sleep and neurobehavioral characteristics of 5- to 7-year-old children with parentally reported symptoms of attention-deficit/hyperactivity disorder. Pediatrics 2003;111:554-63.
6. Brunetti L, Tesse R, Miniello VL, Colella I, Delvecchio M, Logrillo
THE EFFECT OF HEADGEAR USE ON OBSTRUCTIVE SLEEP APNEA

1. Kulnis R, Nelson S, Strehl K, Hans M. Cephalometric assessment of snoring and nonsnoring children. Chest 2000;118:596-603.

2. Pirrós-Parkkinen K, Löppönen H, Nieminen P, Tolonen U, Pirttiniemi P. Cephalometric evaluation of children with nocturnal sleep-disordered breathing. Eur J Orthod 2010;32:662-71.

3. Wang T, Yang Z, Yang F, Zhang M, Zhao J, Chen J et al. A three dimensional study of upper airway in adult skeletal Class II patients with different vertical growth patterns. PLoS One 2014;9:e95544.

4. Johal A, Patel SI, Bartagel JM. The relationship between craniofacial anatomy and obstructive sleep apnea: a case-controlled study. J Sleep Res 2007;16:319-26.

5. Katyal V, Pamula Y, Martin AJ, Daynes CN, Kennedy JD, Sampson WJ. Craniofacial and upper airway morphology in pediatric sleep-disordered breathing: Systematic review and meta-analysis. Am J Orthod Dentofacial Orthop 2013;143:20-30.

6. Melsen B, Dalstra M. Distal molar movement with Kloehn headgear: is it stable? Am J Orthod Dentofacial Orthop 2003;123:374-8.

7. Siqueira DF, de Almeira RR, Janson G, Brandão AG, Coelho Filho CM. Dentofacial analysis of skeletal Class II malocclusions. Am J Orthod Dentofacial Orthop 2007;131:447.

8. Keeling SD, Wheeler TT, King GJ, Garvan CW, Cohen DA, Cabassa S et al. Anteroposterior skeletal and dental changes after early Class II treatment with bionators and headgear. Am J Orthod Dentofacial Orthop 1998;113:40-50.

9. Kirjavainen M, Kirjavainen T, Hurmenrinta K, Haavikko K. Orthopedic cervical headgear with an expanded inner bow in class II correction. Angle Orthod 2000;70:317-25.

10. Haralabakis NB, Halazonetis DJ, Sifakakis I. Activator versus cervical headgear: superimpositional cephalometric comparison. Am J Orthod Dentofacial Orthop 2003;123:296-305.

11. Jena AK, Singh SP, Ureja AK. Sagittal mandibular development effects on the dimensions of the awake pharyngeal airway passage. Angle Orthod 2010;80:1061-7.

12. Ozbek MM, Memikoglu TU, Gögen H, Lowe AA, Baspinar E. Oropharyngeal airway dimensions and functional-orthopedic treatment in skeletal Class II cases. Angle Orthod 1998;68:327-36.

13. Tüllroh JF, Philips C, Koch G, Profti WR. The effect of early intervention on skeletal pattern in Class II malocclusion: a randomized clinical trial. Am J Orthod Dentofacial Orthop 1997;111:391-400.

14. Pirrós-Parkkinen KI, Pirttiniemi P, Nieminen P, Löppönen H, Tolonen U, Uotila R et al. Cervical headgear therapy as a factor in obstructive sleep apnea syndrome. Pediatr Dent 1999;21:39-45.

15. Thiruvenkatachari B, Harrison JE, Worthington HV, O’Brien KD. Orthodontic treatment for prominent upper front teeth (Class II malocclusion) in children. Cochrane Database Syst Rev 2013;11:CD003452.

16. Harrison JE, O’Brien KD, Worthington HV. Orthodontic treatment for prominent upper front teeth in children. Cochrane Database Syst Rev 2007;3:CD003452.