Systematic Review of Systematic Reviews:
A critical appraisal of the current evidences in orthodontics

Rosaria Bucci

March, 2015

UNIVERSITA' DEGLI STUDI DI NAPOLI FEDERICO II

SCUOLA DI DOTTORATO IN MEDICINA CLINICA E SPERIMENTALE

Dottorato di Ricerca in Scienze Odontostomatologiche XXVII° Ciclo

Coordinatore: Prof. Sandro Rengo

TESI DI DOTTORATO IN ORTOGNATODONZIA

Systematic Review of Systematic Reviews: A critical appraisal of the current evidences in orthodontics

PROMOTORE
Chiar.mo Prof.
Roberto Martina

CANDIDATA
Dott.ssa Rosaria Bucci

ANNI ACCADEMICI 2012-2015

Candidate:
Dr. Rosaria Bucci

Promotor:
Prof. Roberto Martina

Co-promotors:
Prof. Ambrosina Michelotti
Dr. Vincenzo D'Antò

Department of Neurosciences, Reproductive Sciences and Oral Sciences, School of Orthodontics and Temporomandibular Disorders, University of Naples Federico II, Naples, Italy.

This thesis was the result of a PhD program attended at the Department of Neurosciences, Reproductive Sciences and Oral Sciences, School of Orthodontics and Temporomandibular Disorders, University of Naples Federico II.
The studies included in this thesis are part of articles published or submitted for publication, and are covered by copyrights.

Abstract

Background: Systematic Reviews (SRs) and Meta-analyses (MAs) are the preferred research approaches to manage the growing number of clinical trials, since they bring together the results from multiple studies by applying rigorous methods, and provide summarized and critically appraised evidence on a given topic. Several orthodontic issues suffer from overwhelm of primary literature, and is frequent to find more than one SR on the same topic showing variable quality and scope. The diffusion of systematic approach made difficult to keep track of the number of SRs that are being published. Hence, SRs of SRs were introduced to integrate and to synthesise information from existing SRs.

Objectives: The aims of this doctoral thesis were: (1) to assess the quality of SRs and MAs of debated issues in the orthodontic literature, (2) to collect the results provided by SRs and MAs and to critically appraise the evidence within reviews. More specifically, the current study focused on two issues: the functional orthopaedic treatment of Class II malocclusion and the dentoalveolar effects of palatal expansion.

Materials and Methods: Two separate literature reviews were conducted. In the first review, SRs and MAs focusing on the effects of the functional orthopaedic treatment of Class II malocclusion in growing patients were included. In the second one, SRs and MAs focusing on the dentoalveolar and skeletal effects of palatal expansion were included.

The methodological quality of each included SR or MA was assessed using the AMSTAR (A Measurement Tool to Assess Systematic Reviews). The level of evidence of the primary studies included in each SR or MA was assessed according to hierarchy of the evidence, and

reported with the LRD scoring (Level of Research Design scoring). The main outcomes were summarized and the evidences retrieved were critically appraised applying pre-determined statements. Literature search, study selection, quality assessment and data extraction were independently conducted by two operators, and consensus was reached through discussion.

Results: For the first SR fourteen studies on Class II treatment were included. The quality of SRs/MAs ranged between AMSTAR 2 and 10 (mean score: 6) and three SRs/MAs included only Randomized Clinical Trials (RCTs) (LRD II). The main outcomes analysed were: dentoalveolar effects, maxillary skeletal effects, mandibular effects and soft tissues effects. According to the current evidence provided by SRs and MAs there was still no sufficient evidence on the effects of Class II functional orthopaedic treatment, due to the small number of adequate MAs, the small number of primary studies for each outcome, and the low-medium quality of primary literature. However, we found some evidence of reduction of the overjet with different functional appliances, some evidence of maxillary growth control with Twin-Block and headgear and some evidence of mandibular length increasing with different functional appliances, even though the clinical relevance of the latter results was questionable.

For the second SR twelve studies on dentoalveolar and skeletal effects of palatal expansion were included. The quality of SRs/MAs ranged between AMSTAR 4 and 10 (mean score: 6.8) and two SRs/MAs included only RCTs (LRD II). The main outcomes analysed were short- and long-term effects of rapid and slow maxillary expansion (RME/SME) on dentoalveolar and skeletal structures in the three

directions (sagittal, vertical and transversal). According to our findings there was high level of evidence that RME determined a significant increase of transverse dentoalveolar maxillary dimensions in the short-term; however, long-term maintenance of this result seemed to be strongly influenced by the retention protocol and was supported by moderate evidence. Meanwhile, also SME showed similar results, but supported by moderate to low evidence. Maxillary skeletal expansion was mainly reported in the short-term with RME, but with low evidence.

Conclusions: Due to the variable quality of SRs and MAs in orthodontics clinicians should be aware of the existent tools to assess strength and weakness of these studies, in order to adequately recognize whenever limited information are provided.

SRs and MAs did not provide sufficient evidence to make a final decision about the effectiveness or ineffectiveness of Class II functional orthopaedic treatment, while RME was reported as an effective procedure for the short-term correction of dental transversal discrepancies.

Table of Contents

Overview.

Translating research findings into clinical practice has always been a major challenge for health care providers. One of the main reasons explaining the difficulties in filling the gap between clinical research evidence and dental practice is the large number of single studies relevant to public health that are published every day. In addition, not all the clinicians have adequate skills and expertise to track down the best evidence (Sutherland, 2000).

As the vastness of the body literature on a given topic might be overwhelming for a single reader, Systematic Reviews (SRs) have been proposed as a new type of approach able to provide practitioners with pre-filtered evidence. Hence, this type of literature allows to save time and to minimize the need for appraisal expertise. In SRs explicit processes to collect, synthesise and appraise scientific evidence on a given topic are used, in order to provide evidence-based answers (Cook *et al.*, 1997).

The growing interest in this type of study gave the way to a large production of SRs, frequently on the same topic, presenting variable scope, methods and quality. It also became frequent to find conflicting results from more than one SR on the same issue.

Hence, the readers who were once overwhelmed by the number of primary studies, in the recent years had to deal with the plethora of reviews (Smith *et al.*, 2011).

Overviews of Systematic Reviews (OoSR) have been introduced to summarize the evidence from different SRs across interventions and settings; hence, this type of study is the logical next step to provide an updated and critically appraised "state-of-the-art" on a given topic (Backer *et al.*, 2014).

This doctoral thesis had the objective to put an accent on the importance of applying such "third level of evidence" in the orthodontic literature. In particular, we aimed to summarize and critically appraise the evidence provided by SRs and Meta-Analsyses (MAs) on two debated orthodontic issues: Class II functional orthopaedic treatment and palatal expansion treatment.

In **Chapter I** an overview on the Evidence-Based Medicine and on the methods adopted in this approach are reported. **Chapter II** gives an insight on what SRs and MAs are and how to adequately conduct and report this type of studies. In **Chapter III** instruments for critically appraising the quality of SRs and the quality of the evidence across reviews are described. In **Chapter IV** the innovation and the importance of OoSR in the current evidence-based decision processes are explained. In **Chapter V** the evidence from SRs and MAs as concerned to functional orthopaedic treatment of Class II malocclusion is reported, while in **Chapter VI** similar methodology is dedicated to the dentoalveolar and skeletal effects of palatal expansion. **Chapter VII** includes a synthesis of the results of the methodological analysis, and a summary of the clinical evidences. In **Chapter VIII** supplementary materials are reported.

References

- Baker PR, Costello JT, Dobbins M, B Waters E. The benefits and challenges of conducting an overview of systematic reviews in public health: a focus on physical activity. Public Health (Oxf). 2014; 36: 517-21.

- Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: Synthesis of best evidence for clinical decisions. Ann Intern Med. 1997; 126: 376-80.

- Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol. 2011; 3: 11-15.

- Sutherland SE. The building blocks of evidence-based dentistry. J Can Dent Assoc. 2000; 66: 241-44.

Chapter I.

Concepts of Evidence-Based Medicine

Rapid advances in the medical field have led to a huge increase in the alternative possibilities for the patients' management. The main challenges for physicians have always been to keep up-to-date with the recent innovations, and to figure out which of the suitable treatment option is more likely to provide benefits to a patient.

The evidence-based decision-making is an approach addressed to guarantee the best care possible. It provides a strategy for addressing gaps in the knowledge, for improving the efficiency of integrating evidence into patient care, and for developing treatment plans that are scientifically defensible.

Hence, to accomplish evidence-based decision, new concepts and skills are required to the physicians (Davidoff, 1995).

The following chapter examines the context in which the Evidence-Based Medicine had spread, describes the fundamentals of this approach, and emphasizes the application of this concept to the dental field.

I.1.   <u>History</u>

The term "Evidence-Based Medicine" (EBM) firstly appeared in a paper published by Dr. Gordon Henry Guyatt (1991). This term, as later reported by the group led by Guyatt at McMaster University in Canada, described a novel method of teaching medicine and was proposed as a "new paradigm for medical practice" with "the potential for transforming the education and practice of the next generation of physicians" (Evidence-Based Medicine Group, 1992).

Before the concept of EBM was broadly accepted and applied, clinical practice was historically viewed as the "art of medicine". The old clinical

problem-solving model was based on individual experiences or on the use of information gained by consulting authorities. Hence, expert opinions, experience, and authoritarian judgment were the foundation for decision-making. It was assumed that medical education, continuing education, journals, individual clinical experience and interaction with colleagues, would be far enough to perform the best choice for the patient. Moreover, analytical methods and mathematical models were limited to research projects and did not found any application in the clinical field.

The new approach proposed by Guyatt and co-workers brings to completion a long series of previous efforts (Claridge and Fabian, 2005; Roger, 2011; Lim and Ding, 2015). Already by the early 1970s, several flaws in the traditional medical approach started to be underlined, and it was thought that medical decisions made by physicians were far to be appropriate.

Alvan Feinstein (1967) attempting to resolve the uncertainty of medical practice at the bedside, pointed out that most decisions were based on or influenced by clinical judgements and the only background of clinical judgements was clinical experience. To minimize the role of clinical judgment, he proposed a new form of medicine that incorporated principles of basic science. Hence, he thought to combine the epidemiology and the medical research to study clinical populations, in order to overcome the traditional anecdotal medical research.

A milestone of the EBM was published in 1972 by Archie Cochrane. This paper underlined, for the first time, the lack of controlled trials supporting many daily medical practices, and clearly reported the

importance of Randomized Clinical Trials (RCTs) for assessing the effectiveness of interventions.

John Wennberg and coll. (1973) documented wide variations in the clinical choices and practice of physicians, approaching to similar medical issues. According to their findings, clinicians frequently recommended different treatment options for similar patients or conditions; hence, it was impossible to claim that they were all doing the right thing or making the right choice.

David Eddy (1980) introduced the concept of guideline in healthcare, with the intent to develop standards of clinical recommendations. In his paper, for the first time, a health care policy based on scientific evidence rather than on expert opinion was proposed. Few years later, the same author published a seminal paper that described the role of clinical policies and guidelines in medical decision-making, underling their importance in increasing the quality of medical care and reducing its cost (Eddy *et al.*, 1982).

Chassin and co-workers (1987) reported that several procedures performed by physicians were considered inappropriate, even according to the standards of their own experts.

All these considerations arouse the attention of clinicians and researchers on the deficiencies in medical decision-making at both individual-patients and population levels, and paved the way for the introduction of evidence-based methods.

David Sackett and coll. (1996) clarified that EBM is 'the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients'. In the same paper, the authors specified that applying EBM means 'integrating individual clinical

expertise with the best available external clinical evidence from systematic research', and none of the two components alone can be considered enough.

According to Sackett *et al.* (1996) clinical expertise is represented by 'the proficiency and judgment that individual clinicians acquire through clinical experience and clinical practice', while the best available external clinical evidence is 'clinically relevant research … [on the] accuracy and precision of diagnostic tests (including the clinical examination), the power of prognostic markers, and the efficacy and safety of therapeutic, rehabilitative, and preventive regimens'.

Few years later, Sackett and coll. extended the concept of EBM by integrating the role of patients' need and expectations in the decision-making. In particular, it was stated that the EBM 'is the integration of best research evidence with clinical expertise and patient values' (Sackett *et al.*, 2000).

To better understand what EBM is, the authors also clarified what it is not (Sackett *et al.*, 1996). Firstly, EBM is not a "cookbook" approach to practice: it must recognise the complexity of the clinical environment in which is applied and the importance of the individual patient circumstances. Indeed, EBM requires the integration of the best evidence with clinical expertise and patients' preferences and, therefore, it appraises, but never replaces, clinical judgement. Secondly, it is wrong to think that EBM can only be 'conducted from ivory towers or armchair'; although many of the skills for literature searching and critical appraisal have not been taught in the past medical programs, currently most teaching centres are including concepts and skills of evidence-based practice in their curricula. Moreover, there is a great volume of

literature aiming to guide clinicians in the acquisition of the skills needed to use scientific evidence. It has been demonstrated that evidence-based approaches can be acquired by clinicians of varying backgrounds at any stage in their careers (Rosenberg and Donald, 1995). Furthermore, the authors state that EBM is not an "old hat", which everyone already use in their daily practice. Even if for long time clinicians have thought to apply rigorous knowledge to their clinical practice, they were always influenced by personal experiences and opinions; this was frequently due to the inaccessibility of the scientific literature before the electronic revolution.

After the introduction of the evidence-based approach to the scientific evidence, the interest on this topic has grown exponentially, and the number of articles about evidence-based practice rapidly increased. These new concepts brought changes in terms of the methods for gathering the evidence (Randomized Controlled Trials and other well-designed methods), statistical tools for synthesising and analysing the evidence (Systematic Reviews and Meta-Analyses), ways for accessing the literature (electronic databases) and ways for applying the evidence (evidence-based decision-making and practice guidelines) (Davidoff, 1995; Davidoff *et al.*, 1999).

Starting from the late '90, Sackett's definition of EBM (represented in Figure I.1) has been largely cited and adopted as paradigmatic approach to deliver optimal patient care in two difference scenarios: population based policies and individual decision-making.
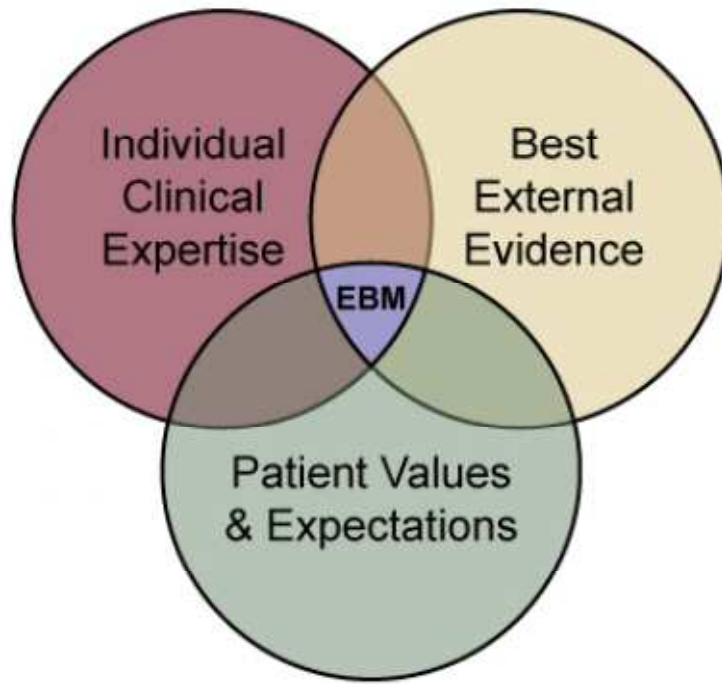
Figure I.1 The Evidence-Based Medicine triad

I.2.   *5-Steps* for EBM

Once defined what EBM was and the importance of this innovative approach, the growing awareness of the need for good evidence led to a rapid spread of this method. Several centres of higher education started teaching these concepts to physicians of undergraduate and postgraduate course. Despite the worldwide spread, the challenge was still how to approach to this method and how to correctly apply it to the current medical practice.

A *5-steps* model was proposed as the best way to integrate the three factors that concur to the EBM (Sackett, 1997; Akobeng, 2005) (Figure I.2).
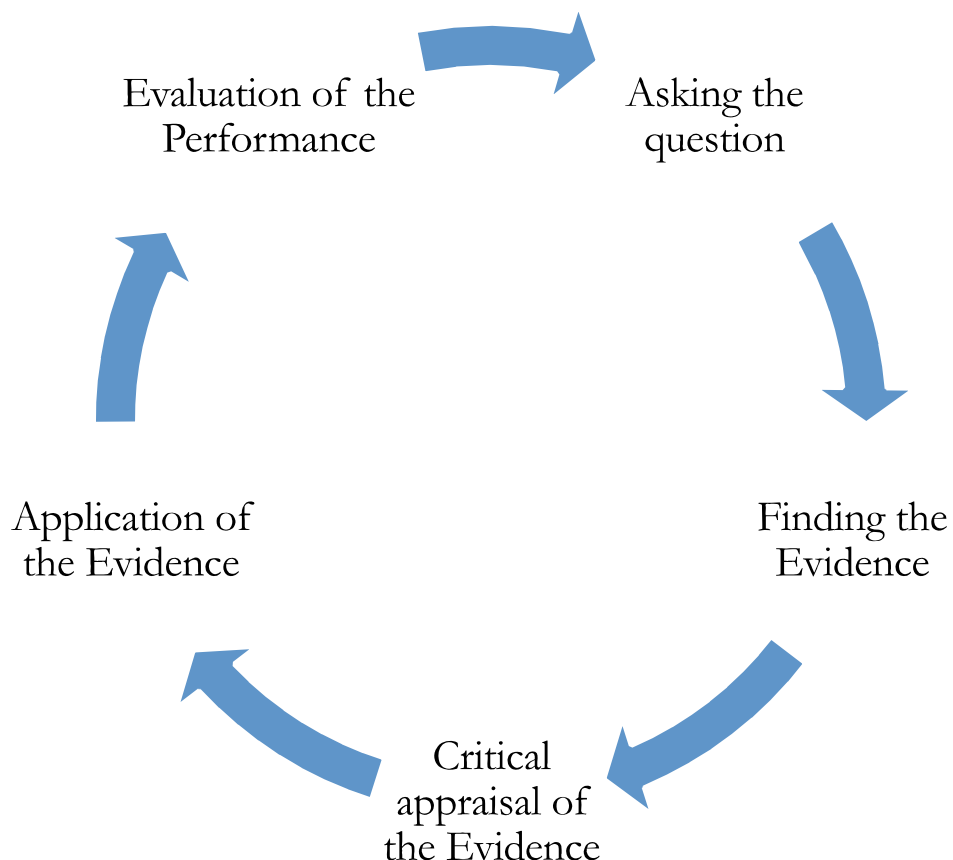
Figure I.2 The *5-steps* model for EBM

The *first step* was 'to convert the information needs into answerable questions'. In particular, it comprises the translation of a clinical problem into structured, clear and well-formulated questions, directly focused on the problem (Carneiro, 1998; Richardson *et al.*, 1995).

A structured way to gain this goal was proposed by Sackett and coll. (2000), by applying the so-called "PICO" (Patient or Problem, Intervention, Comparison, Outcome/s) format.

The *second step* was about 'tracking down, with maximum efficiency, the best evidence with which to answer [the question]'. To seek relevant evidence comprises different phases, from the choice of keywords and sources, to the conduction of the search and the identification of suitable results. To minimize the efforts in this step, trainings' session of database searching seem to improve search performance and quality of evidence retrieved (Rosenberg *et al.*, 1998). However, the instruments providing the easiest approach to the best scientific pre-filtered evidence are Systematic Reviews and Meta-Analyses that will be extensively described in Chapter II.

The *third step* concerned the 'critical appraise [of the] evidence for its validity (closeness to the truth) and usefulness (clinical applicability)'. At this stage, once a potential diagnosis or treatment has been identified, it becomes crucial to decide whether it is adequate for the patient that is about to receive the intervention. Hence, research evidence might be appraised with regard to three main areas: validity, importance and applicability (Rosenberg and Donald, 1995). Different tools for appraising the evidence are available, according to the type of article. For instance, only for non-randomized studies hundreds of quality tools exist and are collected and summarized in a SR (Deeks *et al.*, 2003).

The *forth step* dealt with the 'integration of the appraisal with clinical expertise and application in practice'. At this step, several factors contribute to the decision of putting in practice the evidence found; among those, the costs and the availability of a treatment, the clinical expertise and the patients' values must be taken into account. Patients' values include preferences, concerns and expectations of the subjects, arisen from an appropriate explanation of risks and benefits of the treatment. Hence, being able to communicate to the patient the information collected, in order to gain an informed decision, is a crucial part of this step.

The final stage, *fifth step,* was the 'evaluation [of the] performance'. In this phase, all the previous four stages of the model should be called into question. Hence, a physician should interrogate him/herself on the ability in formulating the right questions, the competence in finding the best evidence, the efficacy in critically appraising the evidence for its validity and potential usefulness, and the capacity of integrating the critical appraisal with clinical expertise and applying the result in clinical practice (Strauss and Sackett, 1998).

I.3.   From EBM to Evidence-Based Dentistry

As for the EBM, the concept of integrating the best available evidences with clinical experience and patients' preferences for making clinical decisions has been applied to the field of dentistry. The American Dental Association (ADA) defined the Evidence-Based Dentistry (EBD) as 'an approach to oral health care that requires the judicious integration of systematic assessments of clinically relevant scientific evidence, relating to the patient's oral and medical condition and history, with the dentist's clinical expertise and the patient's treatment needs and preferences' (ADA policy statement).

This concept is relatively more recent in the dental field than in the general medical subjects (Richards, 1995) and has been widely used in the recent years, sometimes erroneously (Sutherland, 2001). In fact, often the EBD has been employed by selecting the evidence in order to justify some practices, to promote new technologies and products, and to support particular viewpoints, rather than to provide the best treatment.

It has been reported that the translation of the scientific evidence into dental education and clinical practice shows a slow development (Helminen *et al.*, 2002). For a long time, dentists have been used to rely on the knowledge gained from their own experience (both successes and failures) or from colleagues' experiences when making clinical decisions: this explains, to some extent, the slow process of translation of the information from research to dental practice (Bader and Shugars, 1995; Ismail *et al.*, 2004).

McGlone and coll. (2001) aimed to study the complexity of changing practice and to understand which factors influence dentists' ability to

change their clinical practices. This analysis pointed out that the passive dissemination of information through the distribution of educational materials and clinical guidelines or the attendance at didactic meetings had a very little effect on changing practice, while more participative and interactive approaches appeared to have a greater long-term impact on producing sustained changes in dental clinical practice. Hence, it is clear that simply reviewing the latest evidence on dental interventions and then circulating clinical guidelines will have only a very limited effect on most dental practitioners' routine managements.

In addition, some studies highlighted other potential barriers to implement evidence-based methods in dental practice (Richards, 1995; Sutherland, 2000); among those, the huge amount of evidences and the lack of good quality studies, are claimed to be relevant issues. Moreover, dentists must daily deal with new materials and techniques and sophisticated patients' needs and demands, and frequently dental practitioners lack of appropriate skills to face all these information. Last but not least, is not uncommon for dental practitioners to mistrust regarding the use of the evidence, especially by third-party funders and regulatory authorities.

Despite all these issues, there are a number of initiatives and challenges in implementing the evidence-based practice in dentistry.

In 1993 the Cochrane Oral Health Group was founded. This group aimed to produce SRs, which primarily include all RCTs of oral health, and to maintain a Trials Register, which is submitted every month for publication in the Cochrane Central Register of Controlled Trials (CENTRAL) on the Cochrane Library.

In 1995 the Centre for Evidence-based Dentistry was established. This is an independent group whose aim is to promote teaching, learning, practice and evaluation of EBD worldwide.

The Canadian Collaboration on Clinical Practice Guidelines in Dentistry (CCCD) was founded in 1999. This group was born from a Workshop sponsored by The Canadian Dental Association few years before (1997). One of the main aims of the workshop was to begin to develop a collaborative approach to implement guidelines in the dental field. The CCCD's plan embraces the principles of evidence-based practice and involves administrative and methodological support for the development of sound guideline in dentistry.

Currently, "the Journal of Evidence-Based Dental Practice" and "Evidence-Based dentistry" are two scientific journals entirely dedicated to bride the gap between research and dental practice, by providing accessible evidence in the oral and dental branches of medicine.

References

- ADA policy statement on evidence-based dentistry. Available at: "www.ada.org/prof/resources/positions/statements/evidencebased.asp".
- Akobeng AK. Principles of evidence based medicine. Arch Dis Child. 2005; 90: 837-40.
- Bader JD, Shugars DA. Variation in dentists' clinical decisions. J Public Health Dent. 1995; 55: 181-8.
- Carneiro AV. The correct formulation of clinical questions for the practice of evidence based medicine. Acta Med Port 1998; 11:745-8.
- Chassin MR, Kosecoff J, Solomon DH, Brook RH. How coronary angiography is used. Clinical determinants of appropriateness. JAMA. 1987; 258: 2543-7.
- Claridge JA, Fabian TC. History and development of evidence-based medicine. World J Surg. 2005; 29: 547-53.
- Cochrane AL. Effectiveness and Efficiency: Random Reflections on Health Services. Nuffield Provincial Hospitals Trust. 1972
- Davidoff F, Case K, Fried PW. Evidence-based medicine: why all the fuss? Ann Intern Med. 1995 1; 122: 727.
- Davidoff F. In the teeth of the evidence: the curious case of evidence-based medicine. Mt Sinai J Med. 1999; 66:75-83.
- Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG; International Stroke Trial Collaborative Group; European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. Health Technol Assess. 2003; 7: iii-x: 1-173.
- Eddy DM. Guidelines for the Cancer-related Checkup: Recommendations and Rationale. CA: A Cancer Journal for Clinicians. 1980; 30: 193-240.
- Eddy DM. Clinical Policies and the Quality of Clinical Practice. New England Journal of Medicine. 1982; 307: 343-47.
- Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA. 1992; 268: 2420-5.
- Feinstein AR. Clinical Judgment. Williams and Wilkins, Baltimore, 1967.
- Guyatt GH. Evidence-based medicine. ACP J Club. 1991; 114: A–16.
- Helminen SE, Vehkalahti M, Murtomaa H. Dentists' perception of their treatment practices versus documented evidence. Int Dent J 2002. 52: 71-4.

- Ismail AI, Bader JD; ADA Council on Scientific Affairs and Division of Science; Journal of the American Dental Association. Evidence-based dentistry in clinical practice. JADA. 200; 135: 78-83.
- Lim WS, Ding YY. Evidence-balanced Medicine: "Real" Evidence-based Medicine in the Elderly. Ann Acad Med Singapore. 2015; 44: 1-5.
- McGlone P, Watt R, Sheiham A. Evidence-based dentistry: an overview of the challenges in changing professional practice. Br Dent J. 2001 23: 636-9.
- Richards D, Lawrence A. Evidence-based dentistry. Br Dent J 1995; 179: 270-3.
- Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club. 1995, 123: A12-3.
- Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem solving. BMJ 1995; 310: 1122–6.
- Rosenberg WM, Deeks J, Lusher A, Snowball R, Dooley G, Sackett D. Improving searching skills and evidence retrieval. J R Coll Physicians Lond. 1998; 32: 557-63.
- Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. B.M.J. 1996; 312: 71-72.
- Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. 2nd ed. Edinburgh & New York: Churchill Livingstone, 2000.
- Sackett DL. Evidence-based medicine. Semin Perinatol. 1997; 21: 3-5.
- Straus SE, Sackett DL. Using research findings in clinical practice. BMJ 1998; 317: 339-42.
- Sur RL, Dahm P. History of evidence-based medicine. Indian J Urol. 2011; 27: 487-489.
- Sutherland SE. Evidence-based dentistry: Part I. Getting started. J Can Dent Assoc. 2001; 67: 204-6.
- Sutherland SE. The building blocks of evidence-based dentistry. J Can Dent Assoc. 2000; 66: 241-4.
- Wennberg J.E. and Gittelsohn A. Small Area Variations in Health Care Delivery. Science. 1973; 182: 1102-8.

Chapter II.

Literature Reviews

With the increase of clinical studies and journal publications, keeping current with relevant research is nearly impossible. Over two million papers are published every year in about twenty thousand biomedical journals (Mulrow, 1994). It has been estimated that one should read approximately 17 papers per day, every day, to be totally aware of the advances in the field (Haynes, 1992). Above all, most busy doctors lack the time or skills to collect and evaluate relevant evidence (Davidoff *et al.*, 1995).

In this context, with the increasing demand of transforming the best research evidence into clinical practice, the need to summarize in one paper the existent literature on a given topic takes place.

In the following chapter Systematic Reviews (SRs) and Narrative Reviews (NRs) will be presented and compared. In addition, an overview of the processes of SR writing will be provided. Finally, the milestones in the diffusion of systematic approach will be described.

II.1. <u>Narrative vs. Systematic Reviews</u>

Scientific literature reviews are studies in which different methods are adopted to collect the literature on a given topic and to summarize primary research findings into a form that provides a trustworthy overview of the current knowledge (Cipriani and Geddes, 2003). This objective can be achieved by applying two different approaches: Narrative Reviews (NRs) and Systematic Reviews (SRs).

For decades authors who were considered experts or well-known figures in the correspondent area, have collected and summarized the published literature on certain topics. This traditional approach represents what is

called *Narrative Review* (or more generically, literature review). The synthesis performed in a NR is free from any clear and objective method to collect and interpret the information. In addition, the synthesis is frequently influenced by personal theories, beliefs and expectations that make this kind of study comparable to a chapter book. Moreover, these studies normally focus on a subset of primary literature chosen on availability or by independent author selection (Uman, 2011). Hence, NRs are useful to obtain a broad perspective on a topic, but are very prone to biases, and the results can often be confusing.

On the contrary, a *Systematic Review* is a rigorous study that attempts to collect all the empirical evidence that fits pre-specified eligibility criteria, in order to answer a clearly formulated question. This approach requires systematic and explicit methods to identify, select and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review (Moher *et al.*, 2009; Higgins and Green, 2011). In this case, the methods are clearly stated and most likely to be reproducible. Minimizing biases in the literature synthesis means providing more reliable findings from which conclusions can be drawn and decisions can be made (Antman *et al.*, 1992, Oxman and Guyatt, 1993).

Differences between NRs and SRs are summarized in Table II.1.

	Systematic Review	**Narrative Review**
Question	Specific and Focused	Broad
Protocol	A priori and peer-reviews	No protocol
Aim	Clear	Usually not specified
Selection criteria (inclusion and exclusion)	A priori and clearly stated	Often not stated
Source	Comprehensive, systematic and clearly stated search strategies	No explicit strategies
Study Selection	Clear and explicit	Often not reported
Evaluation	Rigorous critical appraising	Variable
Extraction of information	Clear and explicit	Not explicit
Data synthesis	Qualitative or Quantitative	Qualitative

Table II.1 Main differences between Systematic Reviews and Narrative Reviews

Due to the rigorous approach to the literature, well-conducted SRs progressively became the "gold standard" for judging whether a treatment does more good than harm (Sackett *et al.*, 1996).

For this reason, although there is no single, universally accepted hierarchy of evidence, there is broad agreement on placing SRs on the top of the research evidence pyramid (Figure II.1).
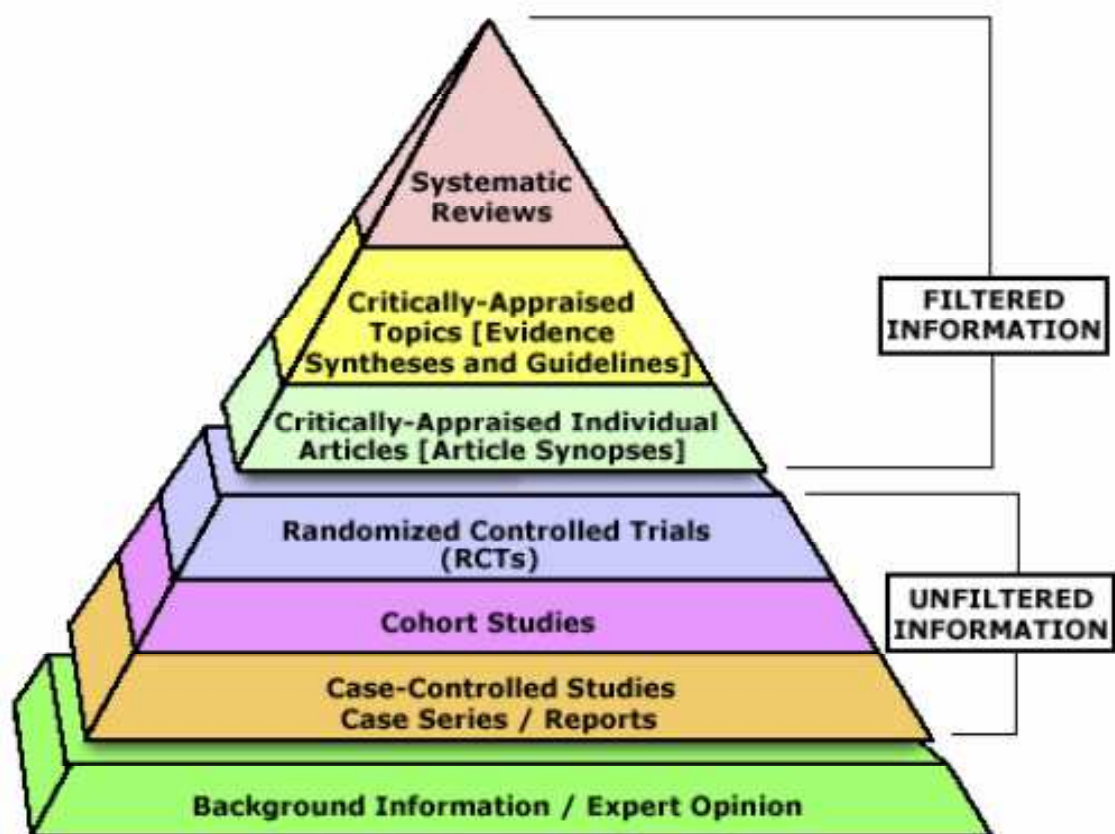
Figure II. 1 The evidence pyramid

II.2.   <u>Methodology of SRs</u>

As previously stated, the main strength of SRs lies in the transparent approach and in the methodological rigor adopted to find, collect and appraise all the relevant literature on a given topic.

Hence, high-quality SRs take a lot of time and efforts: it has been estimated by the Centre for Reviews and Dissemination (http://www.york.ac.uk/inst/crd/index.htm) that a team will take, on average, 9-24 months to complete a SR.

Applying a correct approach represents the cornerstone to minimize biases, and allows readers to decide for themselves whether the information reported are appropriate or not.

Correct methods to conduct a SR, even though not presented in a uniform way, have been extensively studied and, overall, most of the publications on this topic report the same main steps (Khalid *et al.*, 2003; Russel *et al.*, 2009; Uman, 2011).

Well-conducted SRs should start from a clear objective. Once the topic has been focused, and before undertaking the review, it is necessary to check if there are already existing or on-going reviews, and whether a new review is justified.

If there is the need in literature of a new SR, a clear and valid clinical question should be identified. An adequately *a-priori* formulated review question defines several methodological factors of the review, as the type of study to be analysed, the inclusion and exclusion criteria to be applied and the data to be extracted. Khalid and coll. (2003) clearly schematized how to structure an unambiguous and focused question for a SR (Fig. II.2) in terms of population, interventions, comparators and outcomes

(PICO). In this phase, also the choice of the type of studies (study design) to be included in the SR might play an important role in determining the reliability of the findings and the validity of the estimates of effect resulting from the review.
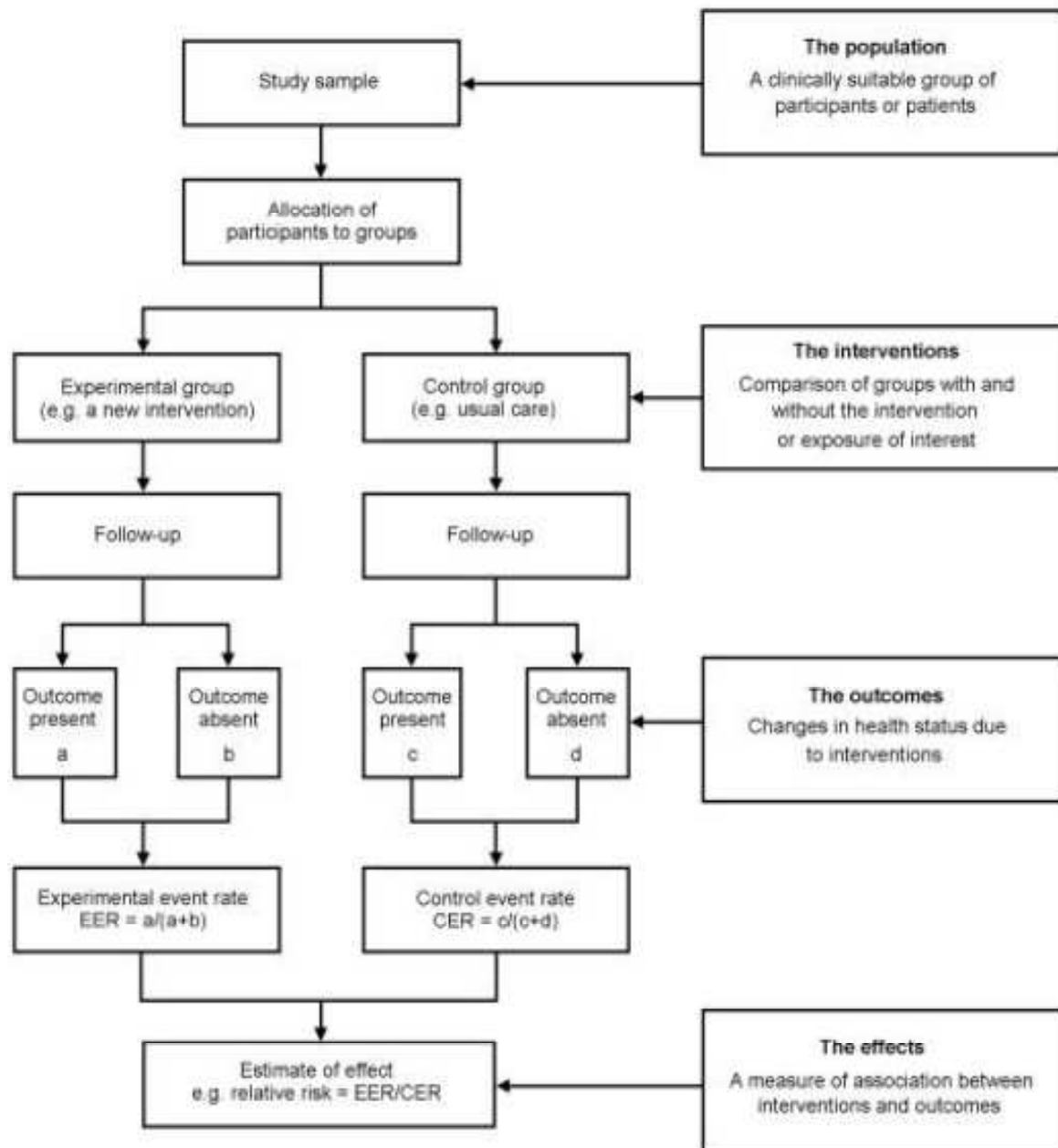
Figure II. 2 Structured questions for Systematic Reviews as proposed by Khan *et al.*, 2003. (Copyright © 2003, The Royal Society of Medicine)

Once the review questions have been set, the following step is to identify the relevant literature using different sources. Electronic databases are

the preferred source, but still checking the article reference lists, hand-searching relevant journals, and personal communicating with experts in the field are essential methods for tracking down additional published and unpublished studies. The search strategies should be built upon the PICO criteria previously identified. At this stage, the collaboration with a librarian can be extremely helpful in terms of developing and running electronic searches (Uman, 2011).

After screening all the retrieved literature, relevant studies that seem to meet inclusion criteria should be further analysed and reviewed in full; the final inclusion/exclusion of the papers follows the full-text reading.

From the included primary literature relevant data should be extracted; commonly, the process to extract the information and the data to be extracted are identified before conduct the study, and again these issues reflect the aim the review. Moreover, at this stage, a critical appraisal of the included papers should be performed by assessing the quality of the individual studies. The quality of a study has been defined as 'the confidence that the trial (study) design, conduct, and analysis has minimized or avoided biases in its treatment (exposure) comparison' (Moher *et al.*, 1995).

There are several checklists, scales and tools developed to assess the quality of the studies (Moja *et al.*, 2005); however, this process is not as easy as it may seems, and a lack of consensus on the best practice exists. Multiple factors can contribute to determine the quality of a study, and it is still challenging to sum up into a single instrument all the factors relevant for a specific branch.

The identified evidence should be combined and synthesized, as far as possible. The aggregation of the results can be performed in two main

ways: qualitative or quantitative synthesis. While the qualitative synthesis is a narrative description of the results found, the quantitative synthesis is performed with statistical methods to combine similar results (MA). Strengths and weaknesses of the meta-analytic approach will be discussed below (Chapter II.3).

Despite the way the data are combined, the results should be interpreted in the light of the quality assessment, in order to provide reliable conclusion and recommendation graded on the strengths and weaknesses of the evidence.

Standardized recommendations to properly conduct a SR have been proposed by the Cochrane Collaboration (see. also Chapter II.3.1) in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins and Green, 2001). Similar instructions are reported in the textbook diffused by the University of York, Centre of Review and Dissemination *"Systematic Reviews: CRD's guidance for undertaking systematic reviews in health care"* (2009).

II.3. <u>Meta-Analysis</u>

A Meta-Analysis (MA) is a review paper in which statistical combination of the results of two or more studies is performed. The term Meta-Analysis was coined by Grass and co-worker, who first introduced this approach in psychological research. The authors state that MA is 'the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding research literature' (1976).

The statistical aggregation of single data aims to achieve higher statistical power for the measure of interest and improves the precision of the estimated effect, so providing more powerful conclusions. (Walker, 2008; Higgins and Green, 2011).

The main answers a MA can provide are: establish whether there is evidence of an effect, estimate the size of the effect and the uncertainty surrounding that size, and investigate whether the effect is consistent across studies. In addition, meta-analytic comparisons can point out controversies between conflict studies, generating new hypotheses for future research.

Whenever the outcomes of several studies show more variation than the variation that is expected because of the sampling of different numbers of research participants, the results should be interpreted with caution and reasons for the heterogeneity should be explored (Uman, 2011).

Frequently, the results of the MA are reported with graphical representation (e.g. forest plot – Figure II.3).
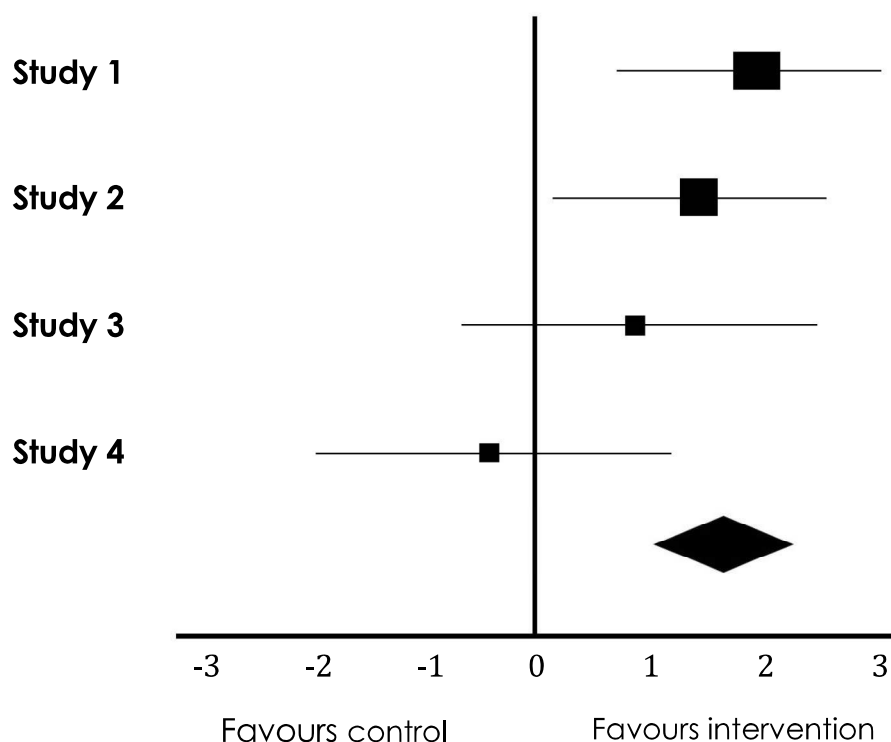
Figure II. 1 Graphical representation of a meta-analysis. Example of a forest plot of four studies assising the effect of a given intervention as compared to a control group.

In the common format of a forest plot, each trial is represented by a line whose borders are the expression of the 95% of Confidence Interval (CI). For each line the square represents the estimate of effect reported in the study; the size of the square depends on the relative weight of each individual study in the MA. This weighting is usually related to the sample sizes of the individual studies, although it can also include other factors, such as the study quality. The vertical central line divides the directions of the effect: in our example (Figure II.3) the left side of the graph (< zero) represents the side favouring control, while the right side (> zero) represents the side favouring treatment. Obviously, the interpretation of the two sides of the graph depends on the measurement evaluated (means, differences, odds ratios, risk ratios).

The diamond reported in the lower line shows the overall result of the MA. Ideally, as reported in Figure II.3, a diamond entirely falling into the right side of the graph shows that the intervention is better than control. The more the diamond is close to the central line, the less the difference between the two groups is significant.

The forest plot can provide also a visual representation of the heterogeneity of results between studies, due to the presence or absence of overlap of the CIs of the trials. Nevertheless, the inconsistency between studies can be statistically calculated with different approaches (chi-squared test or I^2).

If the primary studies are too different in terms of methodology (comparators) and outcome, a common risk for MA is to 'combine apples with oranges' (Higgins and Green, 2011) providing meaningless results.

II.4. <u>Strengths and weaknesses of SRs and MAs</u>

Due to the overwhelming volume of medical literature it is obvious for the readers to prefer summaries of information instead of full publications. This makes a well-conducted SR invaluable for both researchers and practitioners.

The main advantages for practitioners are: to save considerable time and to rely on someone else's expertise in finding the evidence. Hence, they are provided with access to trustworthy pre-filtered evidence, less likely to be biased as compared to a single study. Above all, frequently reviewers' may have obtained information from the authors of the primary studies, which were not available in the original reports; therefore, the evidence retrieved from SR might be even more accurate than that available in the published literature (Garg *et al.*, 2008).

Moreover, mathematically combining data from well-conducted primary studies into a MA means combining the samples of the individual studies to provide a more precise estimate of the underlying "true effect" than any individual study.

In addition, SRs are able to provide a picture on what has been studied in literature, but they also document knowledge gaps and report whenever particular issues are provided with very low quality primary studies. Hence, the identification of these gaps is a useful guidance for future primary studies. In fact, funding agencies increasingly require indication deriving from an existent SR prior to initiating a new research study.

Even if SRs can be a great resource for the above-mentioned reasons, they still have several limits. Firstly, simply stating that a manuscript is a SR or a MA does not guarantee that the review was conducted with rigorous methods (Yusuf, 1997). Furthermore, common problems for this type of studies are: the overload of publications, the scarce update, the variability in reporting and quality standards, and the presence of some biases often not well controlled (in particular publication bias).

The overload of published SRs is a recent issue that seems not to show an inversion in trend (Bastian *et al.*, 2010). The proposed method to face this problem is to conduct 'third-level' research studies to summarize the evidence from SRs and MAs; this topic will be further explained in Chapter IV. Nevertheless, more attention should be put by the reviewers' to explore online registers (i.e. PROSPERO – International prospective register of systematic reviews; Booth *et al.*, 2012) to evaluate the presence of running review protocol and to prevent redundancy of effort.

SRs commonly suffer from publication bias, which is the tendency to submit or accept more positive findings than negative or unfavourable results (Dickersin, 1990). Moreover, there is a significant number of trials and studies that remains unpublished. Hence, to provide really unbiased results, reviews' authors should carefully track down the "grey literature" and measure the risk of publication bias by means of specific statistical approaches (Deeks *et al.*, 2005).

Due to the on-going proliferation of the number of trials published every day, it is likely for a SR to rapidly become "out of date" (Shojania *et al.*, 2007).

Finally, a great variability in the methods adopted and in the quality of reporting among reviews has been found (Moher *et al.*, 2007). Efforts to some extent have been done to overcome the latter issues (absence of update and variability of methods), and they will be discussed in the following paragraph.

II.5. Advancements in systematic approach

II.5.1. Organizations involved in SRs

a. *The Cochrane Collaboration*

In 1993 the British National Health Service founded the Cochrane Centre (www.cochrane.org), at Oxford (UK), named in honour of Archie Cochrane. The collaboration aimed to organize medical research information in a systematic way, in order to facilitate the choices of health professionals and patients, according to the principles of EBM.

This group was developed from the ideas of research update and accessibility of relevant resources, firstly proposed by Archie Cochrane (1972).

In one of his papers, Cochrane (1979) stated: 'it is surely a great criticism of our profession that we have not organised a critical summary, by specialty or subspecialty, adapted periodically, of all relevant randomised controlled trials'. In 1987, he referred to a SR of randomised controlled trials (RCTs) of care during pregnancy and childbirth as 'a real milestone in the history of randomised trials and in the evaluation of care', and suggested to adopt this method for future research.

Nowadays, the Cochrane Collaboration is an independent, non-profit, non-governmental organization consisting of a group of more than 31,000 volunteers in more than 120 countries. The core work of the collaboration is done by the Collaborative Review Groups, which are formed by individuals who have a common interest in health care problems and busily work to prepare, maintain, update and constantly promote the accessibility to SRs in the health care.

It 2011 it was estimated that more than 4500 full Cochrane reviews and 2000 protocols were available (Allen and Richmond, 2011). Up to 2015 only the Cochrane Oral Health Group (COHG) had produced more than thousand SRs, half of which are strictly related to dentistry and oral health.

The Cochrane SRs are constantly updated by searching for new potentially relevant studies and underling the differences between the updated review and the original one. Caring about the update of the review if one of the major tasks of the Cochrane groups; as a matter of fact, the number of Cochrane SRs updated is definitely higher than that of non-Cochrane SRs published in peer-reviewed papers (Jadad *et al.*, 1998).

The literature produced by the Cochrane Collaboration is collected on a dedicated online database (*The Cochrane Library;* www.cochranelibrary.com).

One of the main contributions that the Cochrane Collaboration provided in order to implement the production of SRs is the distribution of the "*Cochrane Handbook for Systematic Reviews of Interventions*" (Higgins and Green, 2011). This document describes in detail the steps for the preparation and the update Cochrane SRs on the effects of healthcare interventions and is constantly updated.

b. *The Campbell Collaboration*

The Campbell Collaboration (www.campbellcollaboration.org) was established in 2000, after a preliminary meeting in London held in 1999. Eighty people took part to the meeting and many of them were active members of the Cochrane Collaboration. This group was born with the

intent to promote evidence-based policy in the education, criminal justice, social policy and social care. In particular is dedicated to prepare and maintain SRs on the effects of social, behavioural, and educational interventions. The collaboration gives its name to the physiologist Dr. Donald T. Campbell, who firstly advocated to apply the scientific evidences to the fields of social sciences and human well-being.

c. . *The Eppi-Centre*

The work of the EPPI-Centre (Evidence for Policy and Practice Information and Co-ordinating Centre) (http://eppi.ioe.ac.uk) concerns with development and training of systematic methods and tools. It is part of the Social Science Research Unit at the Institute of Education (University of London) and is funded by a number of research councils, government departments and charities and national and international partners. More specifically, they provide tools and methods that serve as help and guidance to those who are interested in conducting a SR, and they work on the diffusion of the reviews on an online library. The EPPI-Centre is actively involved also in teaching and training the practice of systematic research and promotes the collaboration among research groups. One of the main contributions of this group is to underline which topics most urgently need to be reviewed, and to analyse and disseminate the results of reviews.

d. *Centre for Reviews and Dissemination (CDR) – University of York*

The CDR (http://www.york.ac.uk/) was established in 1994 and is funded by the UK National Institute for Health Research (NIHR). This Centre is exclusively dedicated to the evidence synthesis in the health

issues and provides to the users four different databases: (1) **DARE** (Database of Abstracts of Reviews of Effects), that contains over 9,000 SRs on health interventions, health services and social determinants of health; (2) **NHS EED** (NHS Economic Evaluation Database), that focuses on the economic evaluations of health intervention (i.e. cost-benefit, cost-utility and cost-effectiveness analyses); (3) **HTA** (Health technology assessment Database), that contains over 10,000 summaries of completed and in-progress health technology assessments, which are typically only available from individual funding agencies; (4) **PROSPERO** (International prospective register of systematic reviews) (Booth *et al.,* 2012), launched in 2011, that collects all the registered protocols of the on-going SRs. The latter tool aims to reduce duplication of SRs and to increase transparency of systematic review methods.

II.5.2.   Standards of the Reporting

One of the key point of a SR is the clear statement of the methods adopted at all the stages of the review process, in order to ensure unbiased and reproducible methodology. In addition, the results should be reported well enough to allow the assessment of the level of bias in the underlying evidence. Hence, without a detailed reporting it becomes impossible to assess strengths and weaknesses of the answers provided by a SR.

In the late '80s Sacks and coll. (1987) evaluated the quality of reporting of 83 MAs, using a scoring method that considered 23 items in six main areas: study design, combinability, control of bias, statistical analysis, sensitivity analysis, and application of results. The findings of this

research showed that reporting was generally poor (only in the 28% of the included MAs all the areas were addressed) and pointed out an urgent need for improved methods in literature searching, quality evaluation of trials, and synthesis of the results. In the same year, Mulrow (1987) evaluated wheatear 50 review articles published in four leading medical journals accomplished eight explicit scientific criteria and he found that none of them met all criteria.

Different checklists have been proposed to help the authors to adequately describe the steps of the review process, but they never reached a wide consensus. A first effort of developing methodological guidelines for SRs was made by Cook and coll. (1995).

The tools described below are the result of international conferences and workshops of experts in the field, and are widely accepted and adopted by the scientific community.

a. *QUORUM – Appendix II.A*

The QUORUM (Quality of Reporting of Meta-analyses) (Moher *et al.*, 1999) statements were developed in 1996 from an international group of 30 clinical epidemiologists, clinicians, statisticians, editors, and researchers who conducted or were interested in MAs. The participants were asked to identify items that they thought should be included in a checklist of standards that would be useful for investigators, editors, and peer-reviewers. The conference resulted into a tool, in table format, which described the preferred way to present abstract, introduction, methods, results, and discussion of a MA. In particular, the tool was composed by headings and subheadings and encouraged the authors to provide the readers with information on searches, selection, validity

assessment, data abstraction, study characteristics, quantitative data synthesis, and trial flow. Along with the checklist, a flow diagram, in figure format, was proposed to provide clear information on the study selection process (number of RCTs identified, included, and excluded and the reasons for excluding them). This tool was specifically developed for MA of RCTs, but was reported to be useful also for reporting of SR (without statistical aggregation) of RCTs.

b. *MOOSE – Appendix II.B*

Observational studies are defined as "an etiologic or effectiveness studies using data from an existing database, a cross-sectional study, a case series, a case-control design, a design with historical controls, or a cohort design" (Stroup *et al.* 2000). When data from higher level of evidence as RCTs are not available, the readers should rely on data from lower level of evidence, as the observational studies. In addition, for several conditions, such as the presence of a risk factor, observational studies are the only valid scientific source. Since the number of SRs and MAs of observational studies was in constant increase, Stroup and coll. (2000) proposed a tool specifically dedicated to this kind of study.

The MOOSE (Meta-analysis of Observational Studies in Epidemiology) (Stroup *et al.*, 2000) checklist is the result of a workshop of 27 experts, held in Atlanta in 1997. These guidelines focus on the reporting of background, search strategy, methods, results, discussion and conclusions.

c. *PRISMA – Appendix II.C*

The PRISMA (Preferred Reporting Items for Systematic Reviews and

Meta-Analyses) (Moher *et al.*, 2009) guidelines have been developed from a revision and expansion of the QUORUM checklist, during a meeting held in Ottawa, Canada, in 2005. Twenty-nine participants, including review authors, methodologists, clinicians, medical editors and consumers took part to the workshop. The authors stated that the change of the name from QUOROM to PRISMA was due to the desire to encompass both SRs and MAs. The final PRISMA Statement consists of a 27-item checklist that is mainly focused on randomized trials, but can also be used for reporting SRs of other types of research. Main implementations from QUORUM to PRISMA checklist are related to: the reporting of a review protocol and the acknowledgement of the iterative process of the review (item 5, 11, 16, 23), the assessment of the difference between study-level risk of bias and outcome-level risk of bias (item 12), and the reporting of different type of biases (item 15). Moreover, also the flow diagram reporting the different phases of the review process showed several changes, such as the number of references retrieved from the search after duplicates were removed.

After its publication the PRISMA Statement aimed to replace the QUOROM Statement for those journals that have endorsed QUOROM. Currently, more the 150 journals in the health sciences endorsed the PRISMA Statement for the reporting of SRs and MAs published in their papers (www.prisma-statement.org/endorsers.htm).

References

- Allen C, Richmond K. The Cochrane Collaboration: international activity within Cochrane Review Groups in the first decade of the twenty-first century. J Evid Based Med. 2011; 4: 2-7.

- Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: Treatments for myocardial infarction. JAMA. 1992; 268: 240-8.

- Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med. 2010 21; 7: e1000326.

- Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, Stewart L. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Systematic Reviews 2012; 1: 2.

- Centre for Reviews and Dissemination. University of York. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare. 3rd edition 2009.

- Cipriani A, Geddes J. Comparison of systematic and narrative reviews: the example of the atypical antipsychotics. Epidemiol Psichiatr Soc. 2003; 12: 146-53.

- Cochrane AL. 1931-1971: a critical review, with particular reference to the medical profession. Medicines for the year 2000. London: Office of Health Economics, 1979; 1-11.

- Cochrane AL. Effectiveness and Efficiency. Random Reflections on Health Services. London: Nuffield Provincial Hospitals Trust. 1972.

- Cochrane AL. Foreword. In: Chalmers I, Enkin M, Keirse MJNC. Effective care in pregnancy and childbirth. Oxford: Oxford University Press. 1989.

- Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. J Clin Epidemiol. 1995; 48: 167-71.

- Davidoff F, Haynes B, Sackett D, Smith R. Evidence based medicine. BMJ. 1995; 29: 310: 1085-6.

- Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol. 2005; 58: 882-93.

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

- Garg AX, Hackam D, Tonelli M. Systematic review and meta-analysis: when one study is just not enough. Clin J Am Soc Nephrol. 2008; 3: 253-60.

- Glass GV. Primary, secondary and meta-analysis of research. Educational Researcher. 1976; 10: 3-8.

- Haynes RB Where's the meat in clinical journals? ACP Journal Club 1993; 119: A23–4.

- Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

- Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, Moher D. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. JAMA. 1998; 280: 278-80.

- Khan KS, Kunz R, Kleijnen J, Antes G. Five steps to conducting a systematic review. J R Soc Med. 2003; 96: 118-21.

- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet. 1999; 354: 1896-900.

- Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials. 1995; 16: 62-73.

- Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009; 6: e1000097.

- Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. PLoS Med. 2007; 27; 4: e78.

- Moja LP, Telaro E, D'Amico R, Moschetti I, Coe L, Liberati A. Assessment of methodological quality of primary studies by systematic

reviews: results of the metaquality cross sectional study. BMJ. 2005; 330: 1053.

• Mulrow CD. The medical review article: state of the science. Annals of Internal Medicine. 1987; 106: 485-8.

• Oxman AD, Guyatt GH. The science of reviewing research. Annals of the New York Academy of Sciences. 1993; 703: 125-33.

• Russell R, Chung M, Balk EM, Atkinson S, Giovannucci EL, Ip S, Mayne ST, Raman G, Ross C, Trikalinos T, West KP, Lau J. Issues and Challenges in Conducting Systematic Reviews to Support Development of Nutrient Reference Values: Workshop Summary Nutrition Research Series, Vol. 2 Technical Reviews, No. 17.2. Rockville (MD): Agency for Healthcare Research and Quality (US). 2009.

• Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996; 312: 71-2.

• Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med. 1987; 316: 450-5.

• Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med. 2007; 147: 224-33.

• Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000; 283: 2008-12.

• Uman LS. Systematic Reviews and Meta-Analyses. J Can Acad Child Adolesc Psychiatry. 2011; 20: 57-9.

• Walker E, Hernandez AV, Kattan MW. Meta-analysis: Its strengths and limitations. Cleve Clin J Med. 2008; 75: 431-9.

• Yusuf S. Meta-analysis of randomized trials: looking back and looking ahead. Control Clin Trials. 1997; 18: 594-601.

Appendix II.A: Quality of Reporting of Meta-Analysis (QUORUM) checklist (Moher *et al.*, 1999).

Heading	Subheading	Descriptor	Reported (Y/N)	Page n.
Title		Identify the report as a meta-analysis of RCTs		
Abstract		Use a structured format		
		Describe:		
	Objective	The clinical question explicitly		
	Data sources	The databases (ie, list) and other information sources		
	Review methods	The selection criteria (ie, population, intervention, outcome, and study design); methods for validity assessment, data abstraction, and study characteristics, and quantitative data synthesis in sufficient detail to permit replication		
	Results	Characteristics of the RCTs included and excluded; qualitative and quantitative findings (ie, point estimates and confidence intervals); and subgroup analyses		
	Conclusion	The main results		
Introduction		*Describe:*		
		The explicit clinical problem, biological rationale for the intervention, and rationale for review		
Methods	**Searching**	The information sources, in detail (eg, databases, registers, personal files, expert informants, agencies, hand-searching), and any restrictions (years		

		considered, publication status, language of publication)		
	Study selection	The inclusion and exclusion criteria (defining population, intervention, principal outcomes, and study design)		
	Validity assessment	The criteria and process used (eg, masked conditions, quality assessment, and their findings)		
	Data abstraction	The process or processes used (eg, completed independently, in duplicate)		
	Study characteristics	The type of study design, participants' characteristics, details of intervention, outcome definitions, and how clinical heterogeneity was assessed		
	Quantitative data synthesis	The principal measures of effect (eg, relative risk), method of combining results (statistical testing and confidence intervals), handling of missing data; how statistical heterogeneity was assessed; a rationale for any a-priori sensitivity and subgroup analyses; and any assessment of publication bias		
Results	**Trial Flow**	Provide a meta-analysis profile summarizing trial flow (figure)		
	Study Characteristics	Present descriptive data for each trial (eg, age, sample size, intervention, dose, duration, follow-up period) (table)		

	Quantitative Data Synthesis	Report agreement on the selection and validity assessment; present simple summary results (for each treatment group in each trial, for each primary outcome); present data needed to calculate effect sizes and confidence intervals in intention-to-treat analyses (eg 2x2 tables of counts, means and SDs, proportions)		
Discussion		Summarize key findings; discuss clinical inferences based on internal and external validity; interpret the results in light of the totality of available evidence; describe potential biases in the review process (eg, publication bias); and suggest a future research agenda		

Appendix II.B: Meta-analysis of Observational Studies in Epidemiology (MOOSE) guidelines (Stroup et al., 2000).

Reporting background should include
Problem definition
Hypothesis statement
Description
Type of exposure or intervention used
Type of study designs used
Study population
Reporting of search strategy should include
Qualifications of searches (e.g. librarians and investigators)
Search strategy, including time period included in the synthesis and keywords
Effort to include all available studies, including contact with authors
Databases and registries searched
Search software used, name and version, including special features
Use of hand searching (e.g. reference lists of obtained articles)
List of citations located and those excluded including justification
Method of addressing articles published in languages other than English
Method of handling abstracts and unpublished studies
Description of any contact with authors
Reporting methods should include
Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested
Rationale for the selection and coding of data (eg, sound clinical principles or convenience)
Documentation of how data were classified and coded (eg, multiple raters, blinding, and interrater reliability)
Assessment of confounding (eg, comparability of cases and controls in studies where appropriate)
Assessment of study quality, including blinding of quality assessors; stratification or regression on possible predictors of study results
Assessment of heterogeneity
Description of statistical methods (eg, complete description of fixed or random effects models, justification of whether the chosen models account for predictors of study results, dose-response models, or cumulative meta-analysis) in sufficient detail to be replicated

Provision of appropriate tables and graphics
Reporting of results should include
Graphic summarizing individual study estimates and overall estimate
Table giving descriptive information for each study included
Results of sensitivity testing (eg, subgroup analysis)
Indication of statistical uncertainty of findings
Reporting of discussion should include
Quantitative assessment of bias (eg, publication bias)
Justification for exclusion (eg, exclusion of non–English-language citations)
Assessment of quality of included studies
Reporting of conclusions should include
Consideration of alternative explanations for observed results
Generalization of the conclusions (ie, appropriate for the data presented and within the domain of the literature review)
Guidelines for future research
Disclosure of funding source

Appendix II.C: Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Moher et al., 2009).

Section/ topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	
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.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	
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.	
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.	
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.	
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	
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).	

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.	
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	
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.	
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	
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.	
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).	
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
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.	
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.	
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).	
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.	
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	

Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	
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).	
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).	
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	
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.	

Chapter III.

Quality assessment

III.1. <u>Quality assessment tools for SRs and MAs</u>

As for all the other research studies, the quality of a SR or MA can be variable. Indeed, it depends on what was done, on the results found, and on the way it is reported (Moher *et al.*, 2009). The fact that a review article is published in a peer-reviewed journal is not a guarantee of scientific methodological quality.

In this context, it is necessary to distinguish between methodological quality and reporting quality. A reporting quality checklist provides a help to the reviews authors to correctly state what was actually done in the review process, but includes items irrelevant the methodological quality. On the other hand, the methodological quality deals with the extent to which scientific review methods were appropriately used to limit biases. Consequently, a SR can have good quality of reporting but still low methodological quality; on the other hand, when the quality of reporting is poor, it is difficult to judge the methodological quality (Pieper *et al.*, 2014).

It has been proved that SRs focusing on similar issues but with variable methodological quality may lead to different conclusions (Moher *et al.*, 2002).

Hence, in the same way that readers must be prepared to assess the quality of primary articles, it becomes crucial to be able to critically appraise the methodological quality of SRs before applying the results into clinical practice.

Much has been written with the intent to establish which are the key items to evaluate the methodology of a review (Mulrow, 1987; Sacks *et*

al., 1987; Oxaman and Guyatt, 1988; Oxaman and Guyatt, 1993), but few efforts were initially done to provide valid instruments. Moher and coll. clearly stated in their paper that the PRISMA checklist 'is not a quality assessment instrument to gauge the quality of a systematic review' (Moher *et al.*, 2002).

The tools described below, the OQAQ (Overview Quality Assessment Questionnaire) and the AMSTAR (A Measurement Tool to Assess Systematic Reviews), are two valid instruments proposed with the purpose to assess the methodological quality of SRs. In addition, they are currently the most widely used instruments, as they are clear and easy to use (Pieper *et al.*, 2014).

a. *OQAQ – Appendix III.A*

Oxman and co-workers (1991) identified 10 questions addressing several tasks of a SR: problem formulation, study identification, study selection, validation of studies, data extraction, data synthesis, and inference.

The identified criteria demonstrated good reliability by showing excellent agreement among judges with different experience (experts in research methodology, clinicians with research training and research assistants).

In a later paper, Oxman and Guyatt (1991) formally structured these questions into a scientific instrument, named OQAQ (Overview Quality Assessment Questionnaire), and tested it for validation. The OQAQ resulted to be a valid instrument to measure the quality of SR. Hence, the authors suggested this tool to readers and editors of clinical journals in order to identify scientifically sound reviews and thus to judge the confidence that should be placed in their conclusions.

According to this instrument the questions from 1 to 9 can be answered

"yes"; "partially/can't tell" or "no". Item 10 judges the overall responses to the 9 questions on a scale of 1-7, where 1 indicates extensive flaws and 7, minimal flaws.

b. *AMSTAR – Appendix III.B*

The AMSTAR (A Measurement Tool for Assessing Multiple Systematic Reviews) (Shea *et al.*, 2007) was built upon a previously developed tool (OQAQ) and on a series of concepts derived from expert opinions (Sacks *et al.*, 1987). This tool was initially composed by 37 items of which 10 derived from the OQAQ (Oxman *et al.*, 1991), 24 from Sacks' paper (Sacks *et al.*, 1987) and 3 were added by the authors according to the recent advancement in research methodology (language bias, publication bias and publication status). A factor analysis was run to identify the relevant items to be included in the new tool. The results of the factor analysis were evaluated by a panel of eleven clinicians, methodologists, epidemiologists, and reviewers. Finally 11 components showing good face and content validity were selected. This 11-items instrument asks the reviewers to answer, for each item, "yes", "no", "can't answer" or "not applicable". "Can't answer" is used when the item is relevant to the research the but not described by the authors; "not applicable" is used when the item is not relevant, such as when a MA has not been possible or was not attempted by the authors. For each item, a full explanation of the rational for the answers is clearly reported in the checklist.

Shea and coll. (2007) reported that the AMSTAR tool can be applied to a wide variety of SRs, although they recognize that it has only been tested on SRs of RCTs evaluating treatment interventions.

The reliability of this instrument was measured in a later publication (Shea *et al.*, 2007), demonstrating satisfactory results: the levels of agreement of the single items ranged from moderate to almost perfect and the reliability of the total AMSTAR score was excellent. When assessing the validity, good construct validity was demonstrated by the convergence of the AMSTAR score with a global instrument considered as a "gold standard" (Shea *et al.*, 2007; 2008). Finally, regarding feasibility (the extent to which users are able to respond to the questions of the instrument), the AMSTAR proved to be an easy-to-use instrument, since an average time of 14.9 minutes were required to complete the checklist for each review.

Currently, the AMSTAR is employed in 9-10% of the overviews of SRs, and its use is constantly increasing. The wide acceptance for the AMSTAR tool is probably due to its availability and ease use, but also because it reflects methodological changes that have occurred after the development of the first tools.

Recently (Kung *et al.*, 2010), the AMSTAR was revised to obtain a tool able to quantify the quality of SRs, and the new proposed instrument was named R-AMSTAR (revised-AMSTAR) *(Appendix III.C)*. The developers of the revised instrument aimed to overcome the qualitative evaluation performed with the original AMSTAR tool by applying a score to each item. The score was based on whether critical reading of the SR revealed satisfactory or unsatisfactory coverage of the criterion. Each domain's score ranged between 1 and 4, and so the total R-AMSTAR ranged between 11 and 44.

Nevertheless, a recent SR re-assessed the measurement proprieties of both AMSTAR and R-AMSTAR (Pieper *et al.*, 2014). According to the

findings of this review, the AMSTAR was a reliable, valid and easy to administer tool, but still further investigation is needed to apply this instrument to SRs including study design other than RCTs. On the other hand, the evidence to assess the proprieties of the R-AMSTAR was still poor.

The AMSTAR has been tested only on SRs of RCTs evaluating treatment interventions, while a wide range of research questions have been addressed with observational studies (i.e. studies on risk factors). Pieper and coll. (2014) reported their experience in applying this instrument to SRs of non-randomized studies. Also in these circumstances, AMSTAR showed good measurement proprieties, even though some problems were faced with some items. For instance, to describe the characteristics of a wide range of studies might be more challenging than to describe the characteristics of an RCT (item 6).

Moreover, problems were encountered also with the items 7 and 8 (critical appraisal), due to the absence of a clearly recommended tool for quality assessment of different type of studies.

Nowadays, the group promoting the original AMSTAR tool is working on the development of a new instrument, the "Non randomized studies-AMSTAR" (NRS-AMSTAR), specifically dedicated to the SRs of observational studies (http://amstar.ca/Developments.php). The aim of the future tool is to extend and modify the original tool, in order to apply the concepts of the AMSTAR to the methods of observational studies.

III.2. Quality of the body evidence

a. GRADE

Judgments about recommendations in healthcare are complex, as clinical guidelines are only as good as the evidence on which they are based (Atkins *et al.*, 2004).

GRADE (Grading of Recommendations Assessment, Development, and Evaluation) is currently emerging as the dominant method for appraising the quality of the evidence of a specific outcome, and for evaluating the strength of recommendations for SRs and guidelines.

Generally, the quality of the evidence indicates 'the extent to which we can be confident that an estimate of effect is correct', while the strength of a recommendation indicates 'the extent to which we can be confident that adherence to the recommendation will do more good than harm' (Atkins *et al.*, 2004).

The GRADE working group was established in 2000 as an informal collaboration of people aiming to address the shortcomings of the existing models for grading the evidence in health care. Currently, this international group is formed by guidelines developers, systematic reviewers and clinical epidemiologists, belonging to different organizations (Guyatt, 2007).

This approach is now suggested by the Cochrane Collaboration for its use in SRs and by the World Health Organization (WHO) for guideline developers.

To implement the appropriate use of the GRADE approach, the Journal of Clinical Epidemiology has published a series of articles, between 2011

and 2013, describing step-by-step the principles of the GRADE and its the application (Grade Series).

The GRADE has been developed for reviews and guidelines that examine alternative management strategies or interventions (Guyatt *et al.*, 2011). When using GRADE, the evidence is not rated study by study, but is rated across studies for specific clinical outcomes. The outcomes of interest should be chosen by framing the question according to what is crucial for patients' care, and for each outcome the comparisons between different interventions should be addressed (Guideline 2).

The rating methods adopted with GRADE are transparent, explicit and systematic, and start from the *a-priori* assumption of "high" ranking to RCTs and "low" ranking to observational studies, as RCTs are generally less prone to biases.

The initial ratings can be "downgraded" or "upgraded" according to different variables.

Main reasons to downgrade the level of evidence are: presence of risk of bias, inconsistency of the evidence, indirectness of the evidence, imprecision of the evidence and presence of publication bias (Guidelines 3 to 8). The risk of bias is generally represented by lack of clearly randomized allocation sequence, lack of blinding, lack of allocation concealment, failure to adhere to intention-to-treat analysis, large losses to follow-up, incomplete or absent reporting of some outcomes. The inconsistency is the presence of significant and unexplained variability in results from different trials. The indirectness refers to the type of comparison performed in the study and to the limits of generalizability of the reported results. The imprecision deals with the range of the confidence interval. Publication bias is related to the extent to which

studies with "negative" findings remain unpublished.

On the other hand, several conditions may allow an upgrade of the rating of the quality of the evidence. In particular, the presence of a large magnitude of effect, the dose-response gradient and the control of confounding (Guideline 9) are reasons for upgrading.

Finally, the quality of the evidence for each relevant outcome is graded as "high", "moderate", "low" or "very low" (Guideline 11). However, it must be taken into account that high-quality evidence does not always imply a strong recommendation. To evaluate the strength of recommendations several factors beside the quality of the evidence must be considered, such as the balance between desirable and undesirable effects and the cost-effectiveness relationship (Guideline 14, 15). The principles of the GRADE approach are summarized in Figure III.1.
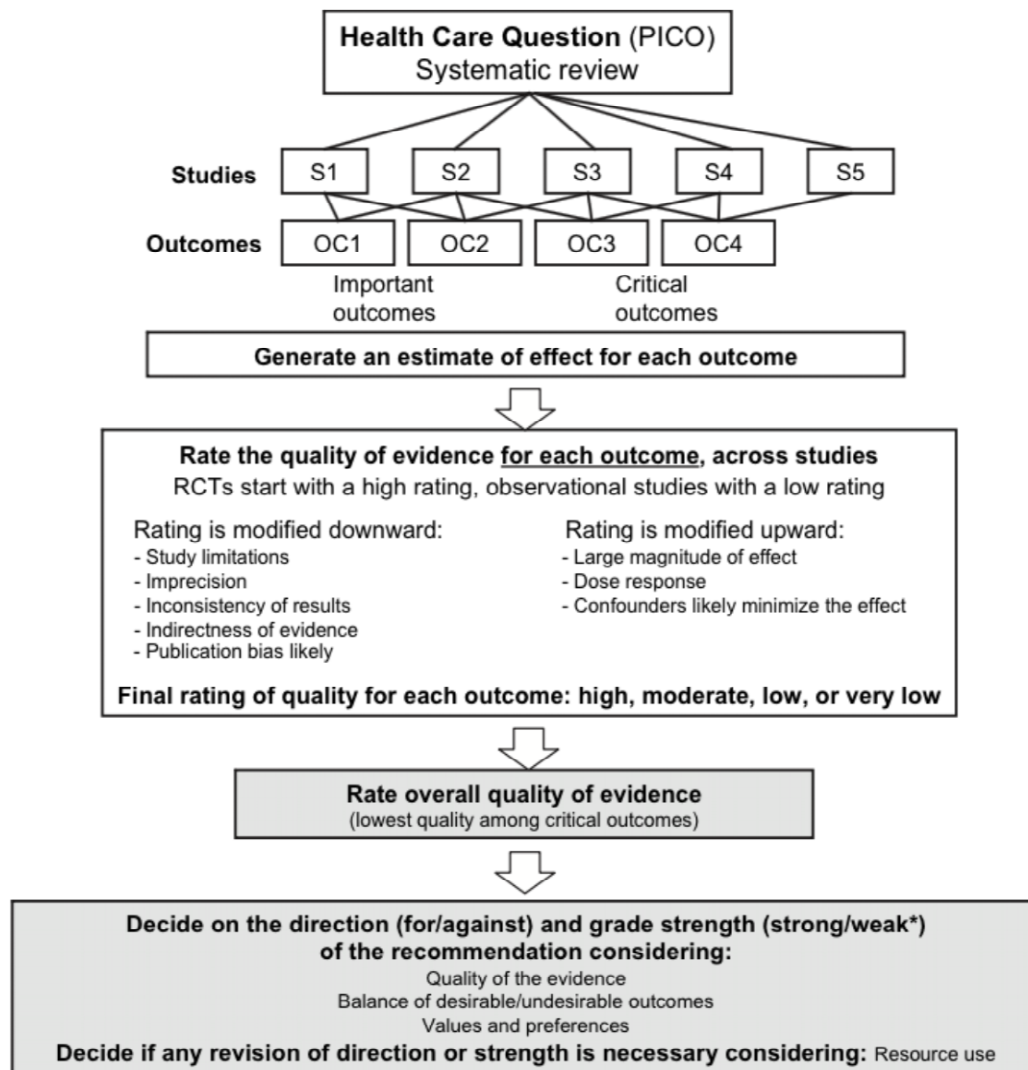
Figure III. 1 Schematic view of GRADE's process for developing recommendations as suggested by Guyatt *et al.*, 2011.

References

- Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magrini N, Mason J, Middleton P, Mrukowicz J, O'Connell D, Oxman AD, Phillips B, Schünemann HJ, Edejer T, Varonen H, Vist GE, Williams JW Jr, Zaza S; GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ. 2004; 328: 1490.

- Guyatt G. An Emerging Consensus on Grading Recommendations? Chin J Evid-based Med, 2007; 7: 1–8.

- Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. PLoS One. 2012; 7: e49667.

- Kung J, Chiappelli F, Cajulis OO, Avezova R, Kossan G, Chew L, Maida CA. From Systematic Reviews to Clinical Recommendations for Evidence-Based Health Care: Validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for Grading of Clinical Relevance. Open Dent J. 2010; 14: 84-91.

- Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009; 6: e1000097.

- Moher D, Soeken K, Sampson M, Campbell K, Ben Perot L, Berman B: Assessing the quality of reports of systematic reviews in pediatric complementary and alternative medicine. BMC Pediatr. 2002, 2: 2.

- Mulrow CD. The medical review article: state of the science. Ann Intern Med. 1987; 106: 485-8.

- Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA, Streiner DL. Agreement among reviewers of review articles. J Clin Epidemiol. 1991; 44: 91-8.

- Oxman AD, Guyatt GH. Guidelines for reading literature reviews. CMAJ. 1988 15; 138: 697-703.

- Oxman AD, Guyatt GH. The science of reviewing research. Ann N Y Acad Sci. 1993; 703: 125-33.

- Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol. 1991; 44: 1271-8.

- Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often

have limited rigor: a systematic review. J Clin Epidemiol 2012; 65: 1267-73.

- Pieper D, Buechter RB, Li L, Prediger B, Eikermann M. Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. Journal of Clinical Epidemiology. 2014 doi:10.1016/j.jclinepi.2014.12.009.

- Pieper D, Mathes T, Eikermann M. Can AMSTAR also be applied to systematic reviews of non-randomized studies? BMC Res Notes. 2014; 7: 609.

- Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med. 1987; 316: 450-5.

- Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, Ramsay T, Bai A, Shukla VK, Grimshaw JM. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One. 2007 26; 2: e1350.

- Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, Henry DA, Boers M. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol. 2009; 62: 1013-20.

GRADE SERIES:

• Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knotterus A. GRADE guidelines: A new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol. 2011; 380–382.

• Guyatt G, Oxman AD, Akl E, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, Debeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ. GRADE guidelines 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011; 64: 383–394.

• Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, Alderson P, Glasziou P, Falck-Ytter Y, Schunemann HJ.GRADE guidelines 2. Framing the question and deciding on important outcomes.J Clin Epidemiol. 2011; 64: 395-400.

• Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, Guyatt GH. GRADE guidelines 3: rating the quality of evidence - introduction.J Clin Epidemiol. 2011; 64: 401-406.

• Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, Norris SL, Williams JW Jr, Atkins D, Meerpohl J, Schünemann HJ.GRADE guidelines 4: rating the quality of evidence - risk of bias.J Clin Epidemiol. 2011; 64: 407–415.

• Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, Alonso-Coello P, Djulbegovic B, Atkins D, Falck-Ytter Y, Williams JW Jr, Meerpohl J, Norris SL, Akl EA, Schünemann HJ. GRADE guidelines 5: rating the quality of evidence - publication bias. J Clin Epidemiol. 2011; 64: 1277–1282.

• Guyatt G, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux P, Montori VM, Freyschuss B, Vist G, Jaeschke R, Williams JW Jr, Murad MH, Sinclair D, Falck-Ytter Y, Meerpohl J, Whittington C, Thorlund K, Andrews J, Schünemann HJ.GRADE guidelines 6. Rating the quality of evidence - imprecision. J Clin Epidemiol. 2011; 64: 1283-1293.

• Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Glasziou P, Jaeschke R, Akl EA, Norris S, Vist G, Dahm P, Shukla VK, Higgins J, Falck-Ytter Y, Schünemann HJ; The GRADE Working Group. GRADE guidelines: 7. Rating the quality of evidence - inconsistency. J Clin Epidemiol. 2011; 64: 1294-1302.

• Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Falck-Ytter Y, Jaeschke R, Vist G, Akl EA, Post PN, Norris

S, Meerpohl J, Shukla VK, Nasser M, Schünemann HJ; The GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence - indirectness. J Clin Epidemiol. 2011; 64: 1303-1310.

• Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, Atkins D, Kunz R, Brozek J, Montori V, Jaeschke R, Rind D, Dahm P, Meerpohl J, Vist G, Berliner E, Norris S, Falck-Ytter Y, Murad MH, Schünemann HJ; The GRADE Working Group.GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol. 2011; 64: 1311-1316.

• Brunetti M, Shemilt I, Pregno S, Vale L, Oxman AD, Lord J, Sisk J, Ruiz F, Hill S, Guyatt GH, Jaeschke R, Helfand M, Harbour R, Davoli M, Amato L, Liberati A,Schünemann HJ. The GRADE Working. GRADE guidelines: 10. Rating the quality of evidence for resource use. J Clin Epidemiology. 2013; 66: 140-150.

• Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, Atkins D, Kunz R, Montori V, Jaeschke R, Rind D, Dahm P, Akl EA, Meerpohl J, Vist G, Berliner E, Norris S, Falck-Ytter Y, Schünemann HJ. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. J Clin Epidemiol. 2013; 66: 151-157.

• Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, Brozek J, Norris S, Meerpohl J, Djulbecovic B, Alonso-Coello P, Post PN, Busse JW, Glasziou P, Christensen R, Schünemann HJ. GRADE guidelines 12. Preparing Summary of Findings tables-binary outcomes. J Clin Epidemiol. 2013; 66: 158-172.

• Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, Johnston BC, Karanicolas P, Akl EA, Vist G, Kunz R, Brozek J, Kupper LL, Martin SL, Meerpohl JJ, Alonso-Coello P, Christensen R, Schunemann HJ. GRADE guidelines 13. Preparing Summary of Findings tables-continuous outcomes. J Clin Epidemiology 2013; 66:173-83.

• Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, Nasser M, Meerpohl J, Post PN, Kunz R, Brozek J, Vist G, Rind D, Akl EA, Schünemann HJ. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol. 2013, 66: 719-725.

• Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, Rind D, Montori VM, Brito JP, Norris S, Elbarbary M, Post P, Nasser M, Shukla V, Jaeschke R, Brozek J, Djulbegovic B, Guyatt G. GRADE

guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol. 2013; 66: 726-35.

Appendix III.A: OQAQ (Overview Quality Assessment Questionnaire) (Oxman *et al.*, 1991)

Criteria for assessing the scientific quality of research overviews
Were the search methods reported?
Was the search comprehensive?
Were the inclusion criteria reported?
Was selection bias avoided?
Were the validity criteria reported?
Was validity assessed appropriately?
. Were the methods used to combine studies reported?
Were the findings combined appropriately?
Were the conclusions supported by the reported data?
. What was the overall scientific quality of the overview?

Appendix III.B: AMSTAR (A Measurement Tool for Assessing Multiple Systematic Reviews) (Shea *et al.,* 2007)

Items
Was an 'a priori' design provided?
Was there duplicate study selection and data extraction?
Was a comprehensive literature search performed?
Was the status of publication (i.e. grey literature) used as an inclusion criterion?
Was a list of studies (included and excluded) provided?
Were the characteristics of the included studies provided?
Was the scientific quality of the included studies assessed and documented?
Was the scientific quality of the included studies used appropriately in formulating conclusions?
Were the methods used to combine the findings of studies appropriate?
.Was the likelihood of publication bias assessed?
.Was the conflict of interest stated?

Appendix III.C: R-AMSTAR (Revised-AMSTAR) (Kung *et al.,* 2010).

Items	Criteria
Was an 'a priori' design provided?	(A) 'a priori' design (B) statement of inclusion criteria (C) PICO/PIPO research question (population, intervention, comparison, prediction, outcome) If it satisfies 3 of the criteria 4 If it satisfies 2 of the criteria 3 If it satisfies 1 of the criteria 2 If it satisfies 0 of the criteria 1
Was there duplicate study selection and data extraction?	(A) There should be at least two independent data extractors as stated or implied. (B) Statement of recognition or awareness of consensus procedure for disagreements. (C) Disagreements among extractors resolved properly as stated or implied If it satisfies 3 of the criteria 4 If it satisfies 2 of the criteria 3 If it satisfies 1 of the criteria 2 If it satisfies 0 of the criteria 1
Was a comprehensive literature search performed?	(A) At least two electronic sources should be searched. (B) The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). (C) Key words and/or MESH terms must be stated AND where feasible the search strategy outline should be provided such that one can trace the filtering process of the included articles. (D) In addition to the electronic databases (PubMed, EMBASE, Medline), all searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found. (E) Journals were "hand-searched" or "manual searched" (i.e. identifying highly relevant journals and conducting a manual, page-by-page search of their entire contents looking for potentially eligible studies) If it satisfies 4 or 5 of the criteria 4 If it satisfies 3 of the criteria 3 If it satisfies 2 of the criteria 2 If it satisfies 1 or 0 of the criteria 1
Was the status of publication (i.e. grey literature) used as an inclusion criterion?	(A) The authors should state that they searched for reports regardless of their publication type. (B) The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc. (C) "Non-English papers were translated" or readers sufficiently trained in foreign language (D) No language restriction or recognition of non-English articles If it satisfies 3 of the criteria 4 If it satisfies 2 of the criteria 3 If it satisfies 1 of the criteria 2 If it satisfies 0 of the criteria 1
Was a list of studies (included and excluded) provided?	(A) Table/list/or figure of included studies, a reference list does not suffice. (B) Table/list/figure of excluded studies1 either in the article or in a supplemental source (i.e.

	online). (Excluded studies refers to those studies seriously considered on the basis of title and/or abstract, but rejected after reading the body of the text) (C) Author satisfactorily/sufficiently stated the reason for exclusion of the seriously considered studies. (D) Reader is able to retrace the included and the excluded studies anywhere in the article bibliography, reference, or supplemental source If it satisfies 4 of the criteria 4 If it satisfies 3 of the criteria 3 If it satisfies 2 of the criteria 2 If it satisfies 1 or 0 of the criteria 1
Were the characteristics of the included studies provided?	(A) In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions AND outcomes. (B) Provide the ranges of relevant characteristics in the studies analyzed (e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.) (C) The information provided appears to be complete and accurate (i.e. there is a tolerable range of subjectivity here. Is the reader left wondering? If so, state the needed information and the reasoning). If it satisfies 3 of the criteria 4 If it satisfies 2 of the criteria 3 If it satisfies 1 of the criteria 2 If it satisfies 0 criteria
Was the scientific quality of the included studies assessed and documented?	(A) 'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant. (B) The scientific quality of the included studies appears to be meaningful. (C) Discussion/recognition/awareness of level of evidence (D) Quality of evidence should be rated/ranked based on characterized instruments. (Characterized instrument is a created instrument that ranks the level of evidence, e.g. GRADE[Grading of Recommendations Assessment, Development and Evaluation.]) If it satisfies 4 of the criteria 4 If it satisfies 3 of the criteria 3 If it satisfies 2 of the criteria 2 If it satisfies 1 or 0 of the criteria 1
Was the scientific quality of the included studies used appropriately in formulating conclusions?	(A) The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review (B) The results of the methodological rigor and scientific quality are explicitly stated in formulating recommendations. (C) To have conclusions integrated/drives towards a

	clinical consensus statement (D) This clinical consensus statement drives toward revision or confirmation of clinical practice guidelines If it satisfies 4 of the criteria 4 If it satisfies 3 of the criteria 3 If it satisfies 2 of the criteria 2 If it satisfies 1 or 0 of the criteria 1
Were the methods used to combine the findings of studies appropriate?	(A) Statement of criteria that were used to decide that the studies analyzed were similar enough to be pooled? (B) For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I). (C) Is there a recognition of heterogeneity or lack of thereof (D) If heterogeneity exists a "random effects model" should be used and/or the rationale (i.e. clinical appropriateness) of combining should be taken into consideration (i.e. is it sensible to combine?), or stated explicitly (E) If homogeneity exists, author should state a rationale or a statistical test If it satisfy 4 of the criteria 4 If it satisfy 3 of the criteria 3 If it satisfy 2 of the criteria 2; If it satisfy 1 or 0 of the criteria 1
.Was the likelihood of publication bias assessed?	(A) Recognition of publication bias or file-drawer effect (B) An assessment of publication bias should include graphical aids (e.g., funnel plot, other available tests) (C) Statistical tests (e.g., Egger regression test). If it satisfies 3 of the criteria 4 If it satisfies 2 of the criteria 3 If it satisfies 1 of the criteria 2 If it satisfies 0 of the criteria 1
.Was the conflict of interest stated?	(A) Statement of sources of support (B) No conflict of interest. This is subjective and may require some deduction or searching. (C) An awareness/statement of support or conflict of interest in the primary inclusion studies If it satisfies 3 of the criteria 4 If it satisfies 2 of the criteria 3 If it satisfies 1 of the criteria 2 If it satisfies 0 of the criteria 1

Chapter IV.

Overviews of SRs

One of the primary objectives of a SR is to summarize the information from several primary studies, in order to provide an answer to a clear clinical question related to diagnosis, prevention or treatment. This process results into a timesaving tool for those involved in providing health care. Due to the increasing spread of this research method, currently more than 8000 SRs have been published only by the Cochrane Collaboration (http://www.cochranelibrary.com/cochrane-database-of-systematic-reviews/index.html), without considering the SRs published in peer-reviewed papers. In 2010 Bastian and coll. (2010) highlighted that around 75 trials and 11 SRs were published every day, and a plateau in growth was not yet reached. The same authors updated this estimate few years later, and reported the result into a PubMed comment; their findings showed that, by the end of 2012, around 26 SRs per day were published.

In Figure IV.1 it is graphically represented the increasing trend of the number of SRs published from 1990 until 2014.
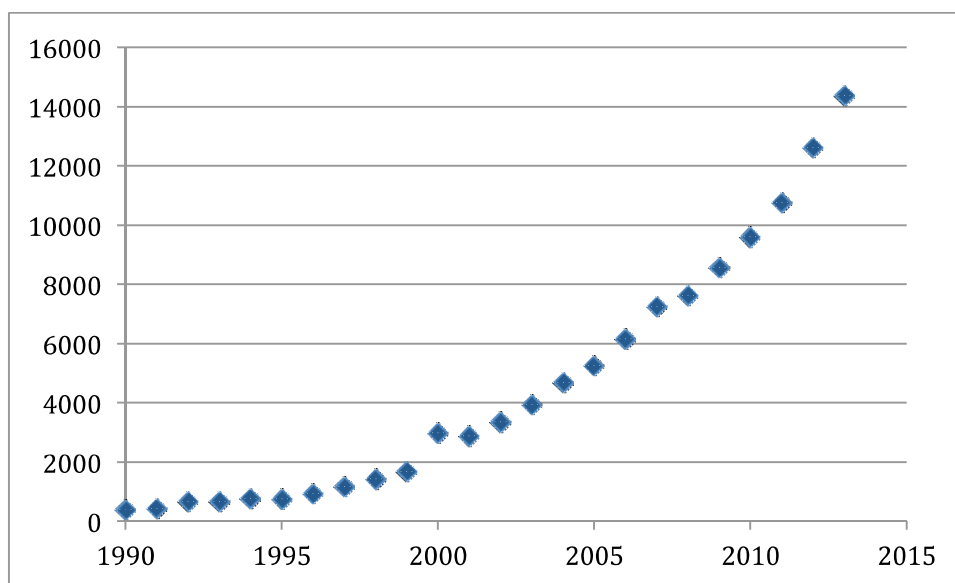
Figure IV. 1 Number of published SRs between 1990 and 2015. This estimate was calculated by searching on Medline (Entrez PubMed) the keyword "Review" [Publication Type] and applying the filter "Systematic Reviews".

Given this large amount of information, keeping up to date with recent literature remains a challenge for the clinicians, even when relying on SRs (Smith *et al.*, 2011; Silva *et al.*, 2012).

It is clear that, to invert this trend the research community should put greater efforts to conduct only SRs that address really relevant questions. Nonetheless, it is frequent that duplication in review efforts is performed, so producing multiple SRs on the same topic. Often when more than one SR is conducted on a given topic, they may vary in quality and scopes (Smith *et al.* 2011), and it is also likely that they provide different results (Jadad *et al.*, 1997).

Jadad and coll. (1997) proposed and algorithm to help the readers to interpret and to choose among discordant SRs (Figure IV.2).
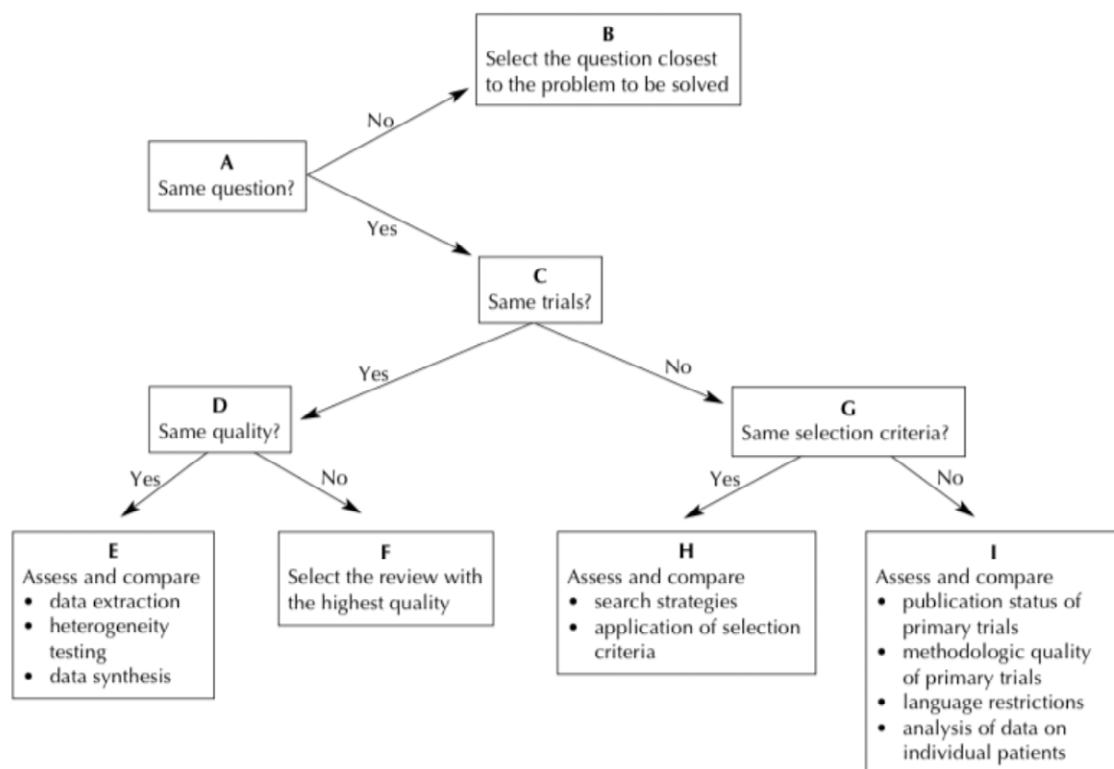
Figure IV. 2 Decision algorithm for interpreting discordant reviews. (Jadad *et al.,* 1997)

To overcome this critical issue, when a number of SRs in areas of priority exists, Overviews of Systematic Reviews (OoSR) have been proposed as new type of study, aiming to compare and contrast the results from multiple SRs.

This type of study, proposed as a 'friendly front end' to the evidence, has been introduced in 2009 by the Cochrane Collaboration (Backer and Oxman, 2009). Since than, the methodology for conducting OoSR has been constantly updated along with the other the topics reported in the Cochrane Handbook for Systematic Reviews of Interventions (Backer and Oxman, 2011).

To conduct an OoSR may have different aims: (1) to summarize evidence from more than one SR of *different interventions* for the same condition or problem; (2) to summarize evidence from more than one SR of the same intervention for the same condition or problem where *different outcomes* are addressed in different SRs; (3) to summarize evidence about adverse effects of an intervention from more than one SR of use of the intervention for one or more conditions; (4) to provide a *comprehensive overview* of an area.

The methodology of this type of study resembles the format of a SR of primary research, but applies these criteria (study selection, data extraction, quality assessment, data synthesis) to SRs. However, since OoSR represent a relatively new research design their methodology has not undergone extensive study. Indeed, a recent study identifying all the OoSR published in last decade (2000-2011) found a great variability of methodology between overviews, and pointed out the need for more guidance and standards to ensure methodological rigor to this approach (Hartling *et al.*, 2012). Similar results of heterogeneity in the methodology

of OoSR were pointed out, in the same year, by another group (Pieper *et al.*, 2012).

Despite the need to improve the methodological rigour and quality of reporting of OoSR, this type of study is growing interest among health care practitioner, due to the numerous advantages provided. Moreover, future challenge for OoSR is to conduct mixed treatment comparisons, by analyzing different intervention that have not been directly compared in head-to-head studies, in order to provide further help in the decision-making process.

IV.1.   Objectives of the research

As for all the other research fields, integrating the evidence into clinical practice is a great challenge for the orthodontists. Hence, due to the clear benefits provided by the OoSR in the evidence-based decision-making, we decided to apply this methodology in controversial orthodontic topics.

In particular, the aims of the current doctoral thesis were:

(1) to evaluate the methodological quality of SRs and MAs in orthodontics;

(2) to synthesize the reported results;

(3) to provide the strength of the reported evidence.

For this purpose this thesis focused on two main topics, highly debated in the orthodontic literature: functional orthopaedic treatment of Class II malocclusion and effects of palatal expansion.

References:

• Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med. 2010 Sep 21; 7: e1000326.

• Becker L, Oxman AD. Overviews of reviews. In Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.02 [updated September 2009]. Edited by: Higgins JP, Green S. Oxford: The Cochrane Collaboration; 2009.

• Becker LA, Oxman AD. Chapter 22: Overviews of reviews. In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011.

• Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. CMAJ. 1997 15; 156: 1411-6.

• Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often have limited rigor: a systematic review. J Clin Epidemiol. 2012; 65: 1267-73.

• Silva V, Grande AJ, Martimbianco AL, Riera R, Carvalho AP. Overview of systematic reviews - a new type of study: part I: why and for whom? Sao Paulo Med J. 2012; 130: 398-404.

• Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol. 2011; 11: 15.

Chapter V.

Evidence on Class II functional orthopaedic treatment

V.1. <u>Summary</u>

This Systematic Review (SR) aimed to assess the quality of SRs and Meta-Analyses (MAs) on functional orthopaedic treatment of Class II malocclusion and to summarise and rate the reported effects. Electronic and manual searches were conducted until June 2014. SRs and MAs focusing on the effects of functional orthopaedic treatment of Class II malocclusion in growing patients were included. The methodological quality of the included papers was assessed using the AMSTAR (Assessment of Multiple Systematic Reviews). The design of the primary studies included in each SR was assessed with Level of Research Design scoring. The evidence of the main outcomes was summarised and rated according to a scale of statements. 14 SRs fulfilled the inclusion criteria. The appliances evaluated were as follows: Activator (2 studies), Twin Block (4 studies), headgear (3 studies), Herbst (2 studies), Jasper Jumper (1 study), Bionator (1 study) and Fraenkel-2 (1 study). Four studies reviewed several functional appliances, as a group. The mean AMSTAR score was 6 (ranged 2–10). Six SRs included only controlled clinical trials (CCTs), three SRs included only randomised controlled trials (RCTs), four SRs included both CCTs and RCTs and one SR included also expert opinions. There was some evidence of reduction of the overjet, with different appliances except from headgear; there was some evidence of small maxillary growth restrain with Twin Block and headgear; there was some evidence of elongation of mandibular length, but the clinical relevance of this results is still questionable; there was insufficient evidence to determine an effect on soft tissues.

V.2. <u>Background</u>

Class II malocclusion is one of the most frequently encountered orthodontic issue as it occurs in about one-third of the population (McLain and Proffit, 1985). The efficacy of the functional orthopaedic treatments for such malocclusion is a widely debated topic, with controversial results in orthodontic literature (Aelbers and Dermaut, 2002).

Systematic Reviews (SRs) and Meta-Analyses (MAs) are generally considered appropriate study design for offering a strong level of evidence (Papadopoulos and Gkiaouris, 2007), especially on controversial topics. In addition, SRs are one of the best ways to stay up to date with current medical literature (Davidoff *et al.*, 2007) instead of reading an average of 17-20 articles per day (Lau *et al.*, 1998). A well-conducted SR aims to collect and synthesize all the scientific evidence on a specific topic, according to strict predetermined inclusion and exclusion criteria (Becker and Oxman, 2011). When possible, SRs might be integrated with MA to statistically contrast and combine results from different individual studies and to increase the statistical power of the analysis (Garg *et al.*, 2008).

Approaching the scientific literature using such methodology might reduce the possibility of systematic errors (bias) (Mulrow, 1984). However, the validity of the results of SRs or MAs might be influenced by different factors; among those, the lack of methodological quality of the individual studies included in the review (Jüni *et al.*, 2001), and the methodological flaws in the development of the SR or MA itself must be take into consideration. In 2010, it has been estimated that about 75

trials and 11 SRs of trials were being published every day (Liberati *et al.*, 2009). Moreover, it is likely to find different SRs on the same topic, conducted with different aims and   methodologies and leading to conflicting results (Bastian *et al.*, 2010).

In this scenario, the need of overviewing and comparing the results from the existent SRs in a single   paper takes place (Smith *et al.*, 2011). To point out the importance of such 'third level' of evidence, the Cochrane Collaboration has introduced the guidelines for Overview of Reviews (Becker and Oxman, 2011), to summarise multiple Cochrane reviews addressing the effects of two or more potential interventions for a single condition.

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

   •   to evaluate the methodological quality of SRs and MAs on functional orthopaedic treatment of Angle Class II malocclusion in growing patients. More specifically, to determine the methodological quality level of the SRs and MAs and to assess the design of the primary studies included in each SR or MA.

   •   to provide an overview of the reported effects of the treatments and to rate the evidence on which these results are based.

V.3. <u>**Materials and methods**</u>

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

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

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

V.3.1. Search strategy

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

Moreover, an exploration of the grey literature (unpublished studies) was performed among the conference abstracts of scientific congresses (European Orthodontic Society and International Association of Dental Research).

The following keywords were used and adapted according to the database rules: "Functional Orthodontic appliance", "Angle Class II", Malocclusion, Review, Systematic Review. The search strategies applied for each database are shown in Table V.1. A detail of the strategy adopted for PubMED search is reported in table V.2

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

Database	Search Strategy	Results
PubMed	(("Activator Appliances"[Mesh]) OR "Orthodontic Appliances, Functional"[Mesh] OR "Orthodontic Appliances, Removable"[Mesh]) AND ("Malocclusion, Angle Class II"[Mesh])) AND (Review* OR Meta-Analys*)	94
Cochrane Library	Malocclusion Angle Class II; Filter: Review	2
Scielo	Angle Class II Malocclusion AND (Review OR Meta-Analysis)	4
Lilacs	(tw:(Angle Class II Malocclusion)) AND (tw:(Review))	23

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

	Keywords	Results
#1	Malocclusion, Angle Class II/therapy[MeSH]	3920
#2	Activator Appliance [MeSH]	1426
#3	Orthodontic Appliances, Removable [MeSH]	4326
#4	Orthodontic Appliances, Functional[MeSH]	2486
#5	#2 OR #3 OR #4	5232
#6	#1 AND #4	1489
#7	#5 AND (Review* OR Meta-Analys*)	94

Table V. 2 Search strategy and keyword combinations used for Medline (via PubMED)

V.3.2.   Studies selection and data collection

Inclusion criteria:

- To be a Systematic Review or a Meta-Analysis;

- Studies on the effects of functional orthopaedic appliances on Class II skeletal malocclusion;

- Studies on growing patients.

Exclusion criteria:

- Dual publication;

- Systematic Review of SRs;

- SR updated in a later publication;

- Treatment protocol not involving functional orthopaedics.

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

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

V.3.3. Quality assessment of the included Systematic Reviews

For each included SR, the methodological quality was assessed using the AMSTAR (Assessment of Multiple Systematic Reviews) (Shea *et al.*, 2007). AMSTAR is composed by 11 items, each one can be answered 'Yes', when clearly done, 'No', when clearly not done, 'Not Applicable', when the item is not relevant, such as when a MA was not attempted by the authors, 'Can't answer', when the item is relevant, but not described by the authors. Each 'Yes' answer is scored 1 point, while the other answers are scored 0 point. According to the number of criteria met, the quality of the included paper was rated as 'Low' (AMSTAR ≤3); 'Medium' (AMSTAR 4–7); 'High' (AMSTAR ≥8) (Ryan *et al.*, 2011; 2014).

Moreover, to assess the design of the primary studs included in each SR the LRD (Level of Research Design scoring) was used (Cooke, 1996; Antes, 1998). The interpretation of such score, which is base on the hierarchy of evidence, is shown in Table V.3.

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

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

Table V. 3 Interpretation of the LRD scores. The scores are based on the type of studies included in the SR.

V.3.4. Synthesis of the results and rating of the evidence

The main results of the included SRs were summarised according to the appliances examined in the study. Afterwards, the evidence on which such results are based was rated according to a modified predetermined scale of statements (Ryan *et al.*, 2011; 2014). The statements applied took into account: the way the data were pooled (MA or narrative synthesis), the statistical significance of the result and the number of studies/participants on which the result was based. A full explanation of the statements adopted is reported in Table V.4.

Moreover, a downgrade of the rating was performed (i.e. from sufficient evidence to some evidence) whenever the quality of most of the individual studies addressing a specific outcome was low. The quality of the individual studies was not re-assessed, but reported as assessed by the authors of the reviews.

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

Table V. 4 Scale of Statements adopted to rate the evidence of the outcomes retrieved from each SR

V.4. <u>Results</u>

V.4.1. Papers selection

The updated electronic search of all databases resulted in 123 references. One article was retrieved from sources other than database, and it was a 'in press' paper provided by the authors. After duplicates were removed, 115 references were left. Eighty-six references were excluded because the topic was not pertinent or because they were not SRs. The remaining eligible 29 articles were entirely read, and 15 of them were excluded (Fig. V.1; Table V.5). The most common exclusion criterion was the absence of a systematic search strategy, especially among the oldest papers. The 14 SRs included and the data extracted from each SR are shown in Table V.6. One-third of the included SRs (5 of 14) were integrated with MA (Antonarakis and Kiliaridis, 2007; Chen *et al.*, 2002, Marsico *et al.*, 2011; Perillo et al., 2011; Thiruvenkatachari *et al.*, 2013). The number of patients included ranged from 59 to 1763. The diagnosis reported in most of the paper was generally 'Angle Class II malocclusion'; six SRs (Barnett *et al.*, 2008; Flores-Mir *et al.*, 2007; Flores-Mir *et al.*, 2006; Flores-Mir and Major, 2006a; Flores-Mir and Major, 2006b, Thiruvenkatachari *et al.*, 2013) more specifically evaluated Class II Division 1 malocclusion and only in one study (Jacob *et al.*, 2013) vertical facial growth was taken into account as inclusion criterion (Class II hyperdivergent patients). Six SRs (Antonarakis and Kiliaridis, 2007; Barnett *et al.*, 2008; Cozza *et al.*, 2006; Ehsani *et al.*, 2014; Jacob *et al.*, 2013; Perillo *et al.*, 2010) included only papers with a comparable Class II untreated group. The appliances

studied in the included SRs were as follows: Activator (Antonarakis and Kiliaridis, 2007; Flores-Mir and Major, 2006b); Twin Block (Antonarakis and Kiliaridis, 2007; Ehsani *et al.*, 2014; Flores-Mir and Major, 2006; Olibone *et al.*, 2006); headgear (Antonarakis and Kiliaridis, 2007; Chen *et al.*, 2002; Jacob *et al.*, 2013); Herbst (Barnett *et al.*, 2008; Flores-Mir *et al.*, 2007); Jasper Jumper (Flores-Mir *et al.*, 2006); Bionator (Flores-Mir and Major, 2006b); Frankel-2 (Perillo *et al.*, 2010). Four papers evaluated several functional orthopaedic appliances, as a group (Chen *et al.*, 2002, Cozza *et al.,* 2011; Marsico *et al.*, 2011; Thiruvenkatachari *et al.*, 2013). The primary outcome of most of the articles (7 SRs) was the effect of treatment on the mandible, measured through different cephalometric methods and reference points.
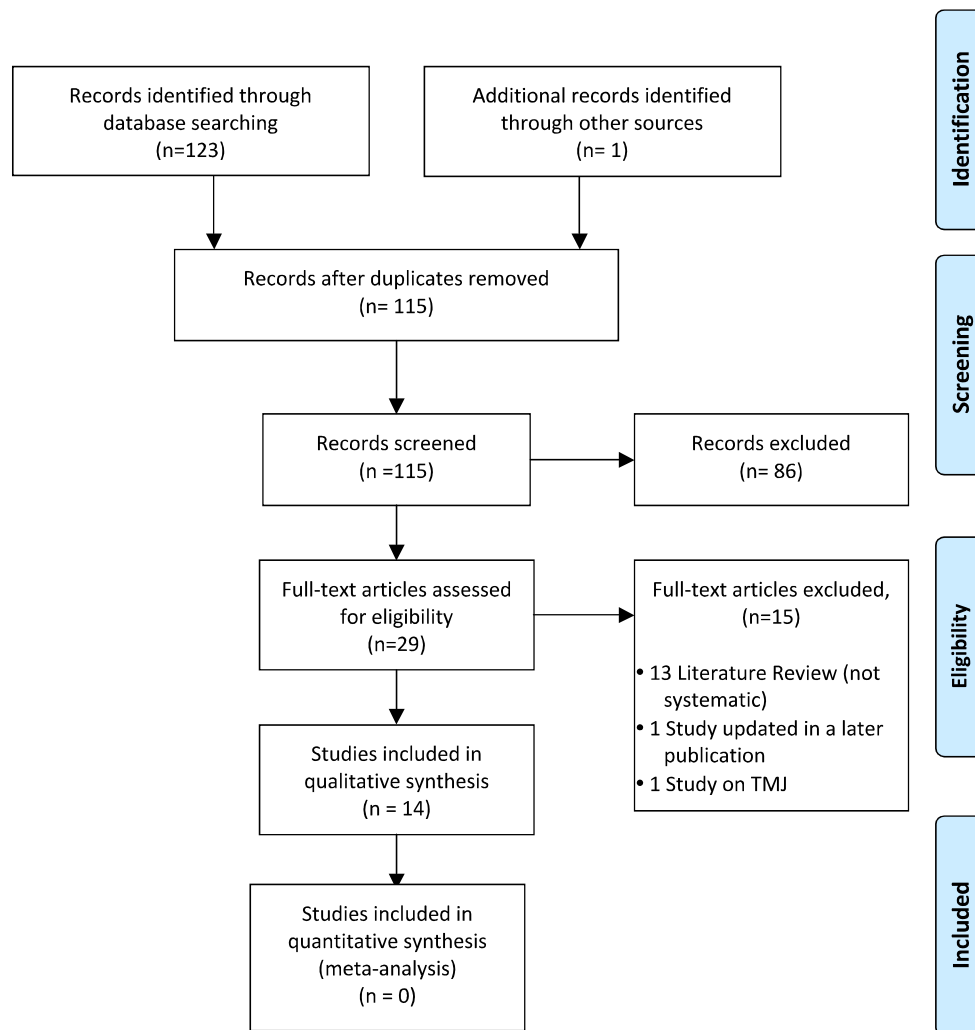
Figure V. 1 PRISMA Flow Diagram

Reference	Reason for the exclusion
- Harrison JE, O'Brien KD, Worthington HV. Orthodontic treatment for prominent upper front teeth in children. (2007)Cochrane Database Syst Rev. 18:CD003452.	Updated in a later publication
- Tadic N, Woods M. Contemporary Class II orthodontic and orthopaedic treatment: a review. (2007) Aust Dent J. 2007;52:168-74.	Non-systematic Review
- Popowich, K., Nebbe, B., Major, P.W. (2003) Effect of Herbst treatment on temporomandibular joint morphology: a systematic literature review. *American Journal of Orthodontics and Dentofacial Orthopedics* 123, 388-394	Focus on TMJ
- Jacobs, T., Sawaengkit, P. (2002) National Institute of Dental and Craniofacial Research efficacy trials of bionator class II treatment: a review. *The Angle Orthodontist* 72, 571-575	Non-Systematic Review
- Dyer, F.M., McKeown, H.F., Sandler, P.J. (2001) The modified twin block appliance in the treatment of Class II division 2 malocclusions. *Journal of Orthodontics* 28, 271-280	Non-systematic Review

- McSherry, P.F., Bradley, H. (2000) Class II correction-reducing patient compliance: a review of the available techniques. *Journal of Orthodontics* 27, 219-225	Non-systematic Review Focus on the classification of the appliances
- Collett, A.R. (2000) Current concepts on functional appliances and mandibular growth stimulation. *Australian Dental Journal* 45, 173-178	Non-systematic Review
- Rudzki-Janson, I., Noachtar, R. (1998) Functional appliance therapy with the Bionator. *Seminars in Orthodontics* 4, 33-45	Non-systematic Review
- Pancherz, H. (1997) The effects, limitations, and long-term dentofacial adaptations to treatment with the Herbst appliance. *Seminars in Orthodontics* 3, 232-243	Non-systematic Review
- Barton, S., Cook, P.A. (1997) Predicting functional appliance treatment outcome in Class II malocclusions-a review. *American Journal of Orthodontics and Dentofacial Orthopedics* 112, 282-286	Non-systematic Review
- Aelbers, C.M., Dermaut, L.R. (1996) Orthopedics in orthodontics: Part I, Fiction or reality--a review of the literature. *American Journal of Orthodontics and Dentofacial Orthopedics* 110, 513-519	Non-systematic Review
- McNamara JA Jr, Peterson JE Jr, Alexander RG.(1996) Three-dimensional diagnosis and management of Class II malocclusion in the mixed dentition. *Semininars in Orthodontics* 2:114-137.	Non-systematic Review
- Ball, J.V., Hunt, N.P. (1991) Vertical skeletal change associated with Andresen, Harvold, and Begg treatment. *European Journal of Orthodontics* 13, 47-52	Non-systematic Review Focus only on vertical growth modification
- Mills, J.R. (1991) The effect of functional appliances on the skeletal pattern. *British Journal of Orthodontics* 18, 267-275	Non-systematics Review
- Bishara, S.E., Ziaja, R.R. (1989) Functional appliances: a review *American Journal of Orthodontics and Dentofacial Orthopedics* 95, 250-258	Non-systematic Review

Table V. 5 References excluded after the full-text reading and reason for the exclusion

V.4.2. Quality of the included SRs and MAs

The Cohen's k coefficient for the AMSTAR items was 0.91, thus indicating very good inter-examiner agreement.

The AMSTAR score ranged from a minimum of 2 to a maximum of 10; the mean score was 6. The single AMSTAR items for each paper and the total AMSTAR scores are shown in Table V.7. Three papers were rated as 'low quality', 8 papers were rated as 'medium quality', and 3 papers were rated as 'high quality'. Six papers included only Clinical Controlled Studies (CCTs), three papers included only Randomised Controlled Studies (RCTs), four papers included both CCTs and RCTs, and one paper included also book chapter and expert opinions. The LRD scores

are shown in Table V.8.

V.4.3. Main outcomes and rating of the evidence

For this purpose, the papers showing low quality (Chen *et al.*, 2002; Jacob *et al.*, 2013; Olibone *et al.*, 2009) (AMSTAR <4) were excluded.

a. *Dentoalveolar effects.* Three SRs (Barnett *et al.,* 2008; Ehsani *et al.,* 2014; Flores-Mir *et al.*, 2007) studied the dentoalveolar effects of functional orthopaedic treatment, while two SRs (Antonarakis and Kiliaridis, 2007; Thiruvenjatachari *et al.,* 2013) focused only on OVJ changes. *Overjet (OVJ).* There is some evidence that functional appliances, considered as a group, significantly decrease the OVJ [-3.88 mm (Barnett *et al.,* 2008) to -4.17 mm (Thiruvenjatachari *et al.,* 2013)], with higher results for the Twin Block when assessed individually [-6.45 mm (Barnett *et al.,* 2008); -3.3 mm to -6-9 mm (Ehsani *et al.,* 2014)]. There is insufficient evidence to support a significant reduction of the OVJ (-4.6 to -5.6 mm) with Splint- Type Herbst appliance (Flores-Mir *et al.*, 2007). There is insufficient evidence to determine an effect of the headgear on the OVJ as controversial results are reported: no significant effect was found by Antonarakis and Kiliaridis (2007) while a small significant reduction was reported by Thiruvenkatachari *et al.* (2013) (-1.07 mm). *Upper and lower incisors.* There is some evidence of proclination of the lower incisors (L1.GoGn: +3.9°) and retroclination of the upper incisor (U1.Mx plane: -9.2°) with Twin Block (Ehsani *et al.,* 2014). There is insufficient evidence to support a proclination/anterior movement of the lower incisors with both Splint-Type (Flores-Mir *et al.*, 2007) and Crown-Banded-Type Herbst appliance (Barnett *et al.,* 2008). *Upper and lower*

molars. There is insufficient evidence to support a distal movement and intrusion of the upper molars and a mesial movement of the lower molars, reported with Splint-Type (Flores-Mir *et al.*, 2007) and Crown-Banded- Type Herbst appliance (Barnett *et al.,* 2008). There is insufficient evidence to determine a mesio-distal movement of upper and lower molars with Twin Block, due to the controversy of the findings (Ehsani *et al.,* 2014).

b.   *Maxillary skeletal effects.* Four SRs evaluated the effects of treatment on the upper jaw (Antonarakis and Kiliaridis, 2007; Barnett *et al.,* 2008; Ehsani *et al.,* 2014; Flores-Mir *et al.*, 2007). There is some evidence of a small maxillary growth restraint with Twin Block appliance [SNA: -0.7°(Ehsani *et al.,* 2014) to -1.03°(Antonarakis and Kiliaridis, 2007)] and with headgear [SNA: -1.01°(Antonarakis and Kiliaridis, 2007)]. There is some evidence of a non-significant effect with other activators, considered as a group (Harvold, Bionator, Schwarz) (Antonarakis and Kiliaridis, 2007). There is insufficient evidence to determine the effect of both Splint-type (Flores-Mir *et al.,* 2007) and Crown-Banded-Type Herbst Appliance (Barnett *et al.,* 2008) on the upper jaw, which is reported to be very low or even not significant.

c.   *Mandibular skeletal effects.* Seven SRs analysed the effects of functional orthopaedic treatment on the lower jaw (Antonarakis and Kiliaridis, 2007; Barnett *et al.,* 2008; Cozza *et al.,* 2006; Ehsani *et al.,* 2014; Flores-Mir *et al.*, 2007; Marsico *et al.,* 2011; Perillo *et al.*, 2010). There is some evidence of a significant advancement of mandibular position in relation to cranial base (SNB) with Twin Block appliance [1.2°(Ehsani *et al.,* 2014); 1.53°(Antonarakis and Kiliaridis, 2007)], while some evidence of a very small increase of the same angle was reported with other activators,

considered as a group [Harvold, Bionator, Schwarz; 0.66°(Antonarakis and Kiliaridis, 2007]. There is some evidence of mandibular length increasing after treatment with functional appliances, considered as a group, ranging between 0.8 and 4.7 mm as measured with Co-Gn (or Co-Pg) and between 1.2 and 2.2 mm as measured with Olp-Pg + OLp-Co (Cozza *et al.*, 2006). The same result was reported with an effect size of 0.61 (Marsico *et al.*, 2011). There is some evidence of a significant elongation of Co-Gn with Fraankel-2 (Perillo *et al.*, 2010) appliance and Twin Block (Ehsani *et al.,* 2014) appliance individually (1.07 mm year and 2.9 mm, respectively). There is insufficient evidence to support a significant mandibular length increasing with both Splint-type (Flores-Mir *et al.*, 2007) (0.7- to 2.7- mm) and Crown-Banded-Type (Barnett *et al.,* 2008) (1.6- to 2.2- mm) Herbst appliances.

d.   *Soft tissue effects.* Four SRs evaluated the effects of functional orthopaedic treatment on soft tissues (Ehsani *et al.,* 2014; Flores-Mir *et al.*, 2006; Flores-Mir and Major, 2006a; Flores-Mir and Major, 2006b). There is insufficient evidence to support an improvement in facial convexity after treatment with fixed appliances (Jasper Jumper (J) and Herbst (H)) (Flores-Mir *et al.*, 2006). In particular, the increase of the naso-labial angle (J) or the retrusion of subnasale point (H) and the protrusion of labrale inferious point (J) or the protrusion of the soft menton (H) are reported. There is insufficient evidence to determine an effect of Twin Block (Ehsani *et al.,* 2014; Flores-Mir and Major, 2006a) on soft tissues due to the controversy of the reported results: in fact, significant effects were reported in the one SR (Ehsani *et al.,* 2014) while non-significant findings were pointed out in another paper (Flores-Mir and Major, 2006a). There is insufficient evidence to determine an effect

on soft tissues with Activator and Bionator as controversial results are reported in one SR (Flores-Mir and Major, 2006b).

Author, Year	Study design, Diagnosis n°of patients	Intervention (I) Control groups (C)	Outcome measures	Quality tool and quality of the individual studies	Results	Authors' conclusions (C) Authors' comments on quality of studies (Q)
Antonarakis and Kiliaridis, 2007	SR and MA of 9 P CTs and RCTs; Class II; 670 subjects	I1: Act (HA; Schwarz; Bio); I2: TB; I3: EOT; I4: Combination (EOT/functional); C: Untreated Class II subjects	Maxillary effect (SNA); Mandibular effect (SNB); Inter-maxillary relation (ANB); Overjet	**Petren et al.:** Medium-High* (9/9)	Maxillary effect: Both I2 and I3 control maxillary growth; higher control with I3 (I3:1.03°, I2:1.01°), with lower homogeneity. No significant effect on SNA with I1 and I4. Mandibular effect: I1, I2 and I4 increase mandibular growth; greater effects and high homogeneity with I2(I1:0.66°; I2:1.53°;.I4: 1.05°). No significant results on SNB with I3. Intermaxillary relation: All I1, I2, I3 and I4 reduce ANB angle; highest reduction with I2 (I1:0.92°; I2:2.61°; I3:1.38°; I4:1.8°), highest homogeneity with I4. Overjet: I1, I2 and I4 show a significant decreasing of the OJ; highest decrease with I2 (I1:3.88mm; I2:6.45mm;	C: All appliances showed an improvement in sagittal intermaxillary relationships (decrease in ANB) when compared to untreated Class II subjects. The use of functional appliances and/or extraoral traction acts mostly in one of the two jaws (mandible for activators and combination appliances and maxilla for extraoral traction) while the twin block group, shows changes on both jaws. Besides the small sagittal skeletal base improvement influencing overjet, the dentoalveolar effect on overjet is brought about by palatal tipping of maxillary and labial tipping of mandibular

					I4:4.37), highest homogeneity with I4. No significant difference in OJ with I3.	incisors, respectively. Q: Heterogeneity of age, sample size, control groups, and appliances.
Barnett *et al.,* 2008	SR of 3 CCTs; Class II Division 1; 102 subjects	I1: Crown- or Banded Type Herbst C: Untreated Class II subjects	Dental and Skeletal cephalometric changes	-	I1 determines an increase of several mandibular sagittal skeletal variables (2-3 mm); minimal maxillary skeletal effects (few variables were statistically significant different), proclination/anterior movement of the lower incisors, retroclination/posterior movement of the upper incisors, extrusive and anterior direction of movement of the mandibular first molars; distal movement and intrusion of the maxillary first molars (clinically questionable). Overjet and overbite were also reduced;	C: Dental changes have more impact than skeletal changes. Q: No RCT, Poor methodological quality of the studies; frequent use of condylion as reference point for mandibular length measurement, which is well-known to be difficult to determine cephalometrically. Different landmarks/measurements, different group age ranges, different treatment duration

| Chen et al., 2002 | SR and MA of 6 RCTs; Class II; Not Reported | I1: Functional appliances (Bass, Bio, Fr-2, TB); C: No treatment and/or EOT | Mandibular growth (horizontal and vertical dimension) | - | I1 significantly increases only in Ar-Pg and Ar-Gn distances. No effect of the type of appliance. | C: There is the need to re-evaluate functional appliance use for mandibular growth enhancement. The clinical effect on mandibular length is little, and probably influenced by reference point. (Ar moves posteriorly and superiorly after functional therapy) The absence of statistically significant difference of angular values (SNB and LIA) was unexpected and might be because all appliances were analysed as a group.

Q: Methodological limitation; heterogeneity of skeletal age, treatment durations; lack of treatment-control match and patient compliance evaluation. |

Cozza et al., 2006	SR of 4 RCTs and 18 CTs (2 P and 16 R); Class II; 1763 subjects	I1: Functional appliances (Act, Bass, Bio, Fr-2, Herbst, MARA, TB)

C: Untreated Class II subjects | Mandibular sagittal position, Total mandibular length, Mandibular ramus height, Mandibular body length

Efficiency of the appliances | **Modified Jadad scale:** Low (3/22) Medium (13/22) Medium-High (6/22) | I1 increases mandibular growth in two third of the samples. Changes in mandibular position in relation to the cranial base were not clinically significant

The Herbst appliance showed the highest coefficient of efficiency (0.28 mm per month) followed by the Twin-block (0.23 mm per month). Lowest coefficient of efficiency for the Frankel appliance (0.09 per month). | C: The amount of supplementary mandibular growth appears to be significantly larger if the functional treatment if performed at the pubertal peak in skeletal maturation.

Q: Quality of the studies from low to medium/high, rare RCTs (neglecting the skeletal maturation). |
| Ehsani et al., 2014 | SR of 10 CTs (6 P and 4 R) and Meta-Analysis of 5 studies; Class II; 664 subjects | I1: TB

C: Untreated Class II subjects | Skeletal, Dental and Soft tissues effects | **Modified risk of bias:** High risk (1/10) Medium risk (5/10) Low risk (4/10) | Data from the meta-analysis: I1 controls maxillary growth (SNA: 0.8°), projected the low jaw slightly forward (SNB: 1.2°), increments the mandibular body length (CoGn: 3 mm), and increases the anterior facial dimensions (ALFH: 2 mm). Moreover, reduces the upper incisor | C: Dental effects are consistently reported. A clinically significant restraint of maxillary growth was not found. Although the mandibular body length is increased, the facial impact of it is reduced by the simultaneous increment of the face height. |

					proclination (U1-AnsPns: 9.2°) and increases the lower incisor inclination (L1-GoGn: 3.8 degree)	Q: Highly heterogeneous and biased studies (various measurements and treatment times, use of historical controls).
Flores-Mir *et al.*, 2007	SR of 3 CCTs; Class II div 1; Not Reported	I1: Splint-Type Herbst; C: No treatment	Dental and Skeletal cephalometric changes	**Self-produced checklist:** Low (3/3)	Skeletal Effects: I1 increases anteroposterior length of the mandible (0.7- to 2.9-mm), increases mandibular protrusion (1.2° to 2.9°), decreases intermaxillary discrepancy (-1,5° to -2.1° and -4.2-mm to -4.9mm), retrudes maxillary anteroposterior position (<1 mm) and increases posterior (1.4- to 2.5-mm) and anterior (1.2- to 3-mm) facial height. Dental effects: I1 reduces OJ (-4.6- to -5.6-mm) and OB (-2.5mm), determines mandibular incisor proclination (3.2° to 4.5°), protrusion (1.5- to 4-mm) and extrusion (5.3°), determines mesial movement of lower molars (0.8- to 3.6-mm) (no extrusion), and distal	C: The combination of several small (statistically significant) changes in different skeletal and dental areas produces the overall reported positive change, but they are not likely clinically significant. Q: Secondary level of evidence. Small sample size. Use of different variables and reference points of cephalometric analysis. No homogeneity in treatment and control groups (race, gender, age.). Few studies using control group that included Class II patients.

					movement of upper molars (2.5- to 5.4-mm) with intrusion (-0.9mm) and retroclination (5.6°). No significant changes for upper incisors.	
Flores-Mir *et al.*, 2006	SR of 5 CTs (1 P and 4 R); Class II div 1; 228 subjects	I1: Jasper-Jumper; I2: Herbst; C: No treatment	Soft tissue changes	**Self-produced checklist:** Low (2/5) Medium (3/5)	I1 increases naso-labial angle, retrudes the position of "Labrale Superious" relative to the vertical reference plane, and protrudes the position of "Labrale Inferious" relative to Esthetic Plane (E-plane); I2 generates a soft menton protrusion, a "Subnasale" retrusion, contradictory results regarding the anteroposition of the upper lip and no changes in the lower lip.	C: There is little of evidence on Jasper Jumper appliance and the results are contradictory. Herbst appliance determines a significant improvement in facial profile. This improvement is not the product of a more forward position of the lower lip but more likely a retrusion of the upper lip. On average, although fixed functional appliances produce some significant statistical changes in the soft tissue profile, the magnitude of the changes may not be perceived as clinically significant.

						Q: Low level of evidence; reference structures not always reliable.
Flores-Mir and Major, 2006a	SR of 2 CCTs; Class II div 1; 59 subjects	I1: TB; C: No treatment	Soft tissue changes	**Self-produced checklist:** Low (1/2) Medium (1/2)	I1 shows no evidence of change of facial convexity, lower lip, nose and soft tissue menton. Controversial changes in the upper lip for the position of *labrale superius* relative to the aesthetic line, which was in a more retruded position (-1.9 mm).	C: A few studies evaluated the soft tissue profile changes. The twin block appliance seems to not produce a soft tissue profile changes to be perceived as clinically significant. Three-dimensional quantification of soft tissue changes is required. Q: Low level of evidence.
Flores-Mir and Major, 2006b	SR of 10 CCTs and 1 RCT; Class II div 1; 540 subjects	I1: Act; I2: Bio; C: No treatment	Soft tissue changes	**Self-produced checklist:** Low (10/11) Medium (1/11)	Contradictory results for both I1 and I2	C: Although some statistically significant soft tissue changes were found, for both I1 and I2 the clinical significance is questionable. Q: Methodological weakness, low level of evidence.

Jacob et al., 2013	SR of 4 CCTs; Class II hyperdivergent patients; 221 subjects	I1:Extraoral high-pull headgear (maxillary splints/banded molars) C: Untreated Class II hyperdivergents	Skeletal changes (horizontal and vertical); Dental effects (molar eruption)	**Modified Antczak et al.:** Low – 4 points out of 10 (1/4) Medium – 6/7 points out of 10 (3/4)	I1 decreases ANB angle (from 0.9° to 1.5°), decreases overjet (2.6- to 6.5 mm). Statistically significant posterior displacement of the maxilla (0.1- to 0.5- mm), distalization of the maxillary molar (0.5 to 3.3 mm), maxillary molar intrusion (0.4- and 0.7- mm), retroclination (4.4° to 11.0°) and intrusion of the maxillary incisors (0.2- to 2.1- mm) were also reported with I1. No effects on the mandible.	C: High-pull headgear treatment improved the AP skeletal relationships, by displacing, the maxilla posteriorly but not the vertical skeletal relationships. Q: Greater attention to the design and report of studies should be given to improve the quality of such trials.
Marsico et al., 2011	SR and MA of 4 RCTs; Class II; 338 subjects	I1: Functional Appliances (Act, HA, Fr-2, Bio, TB); C: No treatment	Mandibular growth (total length)	**Assessment of risk of bias:** High risk (1/4) Unclear risk (1/4) Low risk (2/4)	I1 increases mandibular growth (1.79 mm in the annual mandibular growth) when compared with C, with statistical heterogeneity.	C: The treatment with functional appliances results in change of skeletal pattern (small increases of mandibular length); however, even if statistically significant, appear unlikely to be very clinically significant. The heterogeneity of the results can be attributed to the difference in sample dimension and

						to the use of different functional appliances). Several benefits must be attributed to the early treatment of Class II malocclusion with functional appliances. Q: Heterogeneity regarding cephalometric analyses, variables and reference points.
Olibone *et al.*, 2006	SR of 45 references (articles and book's chapters) Class II; Not reported	I1: TB; C: Not reported	Mandibular growth; Maxillary effect; Intermaxillary relation; Upper incisor; Lower incisor	-	I1 produces significant reduction of the SNA angle; retroclination of the upper incisors, increase of mandibular length and condyle growth; proclination of the lower incisors; improvement of maxilla-mandibular relation	C: The alterations were the combination of modifications on the condyle, the mandibular fossa, the basal bone and dentoalveolar alterations. Most of the authors recommend the use of the Twin-Block during the pubertal peak. Q: nr.

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

					5.22mm) and ANB (-0.63°) when comparing Late orthodontic functional treatment with no treatment.	of the evidence.
		AIM2: I1: Functional Appliances C: No treatment or different kind of appliance				

SR: Systematic Review; MA: Meta-Analysis CCT: Controlled Clinical Trial; *P:* Prospective; *R:* Retrospective; *RCT:* Randomized Controlled Trial; *Act:* Activator; *TB:* Twin Block; *EOT:* Extra Oral Traction; *Bio:* Bionator; *Fr-2:* Frankel-2; *MARA:* Mandibular Anterior Repositioning Appliance; *HA:* Harvold Activator; *AIBP:* Anterior Inclined Bite Plate; *stated by the authors. Quality not reported for the individual studies.

Table V. 6 Data extracted from the 14 Systematic reviews and Meta-Analyses included

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

Y= Yes; N= No; NA= Not Applicable; CA= Can't Answer

Table V. 7 Quality assessment according the AMSTAR items for each SR and Total AMSTAR scores. For each Yes answer: 1 point; all the other answers: 0 point.

Authors, Year, Reference	LRD
Antonarakis and Kiliaridis, 2007	II-III
Barnett *et al.,* 2008	III
Chen *et al.,* 2002	II
Cozza *et al.,* 2006	II-III
Ehsani *et al.,* 2014	III
Flores-Mir *et al.,* 2007	III
Flores-Mir *et al.,* 2006	III
Flores-Mir and Major, 2006a	III
Flores-Mir and Major, 2006b	II-III
Jacob *et al.,* 2013	III
Marsico *et al.,* 2011	II
Olibone *et al.,* 2006	III-IV-V
Perillo *et al.,* 2010	II-III
Thiruvenjatachari *et al.,* 2013	II

Table V. 8 Study design of the primary studies included in each SR, as assessed according to the LRD scores.

V.5. <u>Discussion</u>

The present SR aimed to summarise the current evidence from the SRs and MAs on the orthopaedic functional treatment of Class II Malocclusion. In particular, the focus of the present study concerns the quality and the main results of the SRs and MAs addressing this issue.

V.5.1. Quality of the included SRs

Scientific and rigorous methods are employed in SRs to identify and summarise the literature, to minimise biases that come from narrative reviews. Nonetheless, as with all the other publications, the value of a SR depends on the way it is conducted and on the accuracy of the results (Liberati *et al.*, 2009). The methodological quality of the included SRs was assessed with the AMSTAR (Shea *et al.*, 2007). AMSTAR is a recent valid and reliable quality tool (Shea *et al.*, 2009), built upon expert opinion and empirical data collected with a previously developed tool (Shea *et al.*, 2007). The item 1 of the AMSTAR ('Was an "a priori" design provided'?) refers to a registered protocol of the review. The databases for protocol registration, such as PROSPERO (International Prospective Register of Systematic Review) (Booth *et al.*, 2011), have been recently introduced; therefore, in our study, due to a chronological limitation the presence of the protocol registration was neglected. Affirmative answer to the item 1 was assigned whenever clear predetermined research criteria were provided. Ensuring such approach avoids the review method to be influenced by reviewers' expectations (Garg *et al.*, 2008). The AMSTAR scores of the SRs included in the current study showed a wide range of values, between 2 and 10, with an

average value of 6. Common factors for the included review to lose point in the AMSTAR score were as follows: not performing a grey literature search (item 4), not assessing the publication bias (item 10) and not providing the conflict of interest of the authors (item 11). However, AMSTAR score have to be carefully interpreted as the single AMSTAR items may have different weights in the overall quality of a SR (List *et al.*, 2010). For instance, reporting the conflict of interest (item 11) has a low impact on the methodology of a SR. On the other hand, the assessment of the scientific quality of the primary study included (item 7) has to be considered a key item, as this evaluation allows the identification of flaws in the primary literature. In 10 of 14 SRs, the quality of the individual studies was documented and reported. Modified Jadad Scale (Jadad *et al.*, 1994) and Assessment of risk of bias (Higgings *et al.*, 2011) were the most used tools, together with self-produced checklists based on the key of interest, which are also considered valid instruments (Jüni et al., 2001; Flores-Mir *et al.*, 2006; Flores-Mir and Major, 2006a; Flores-Mir and Major, 2006b). Among the included studies, only the Cochrane review (Thiruvenkatachari *et al.*, 2013) adopted the GRADE approach (Atkins *et al.*, 2004) suggested from the Cochrane collaboration as system for grading the quality of evidence and providing the strength of recommendation.

The paper with the highest AMSTAR score (AMSTAR 10) is a Cochrane Review (Thiruvenkatachari *et al.*, 2013). This result is in accordance with what previously pointed out in several studies (Jadad *et al.*, 1998; 2000; Fleming *et al.*, 2013) when comparing the methodology of Cochrane SRs with that of SRs published in paper-based journals; the authors found that the SRs published by the Cochrane Collaboration present less flaws

and better methodological quality. These findings suggest that standardised instructions and several peer-review levels improve the methodological soundness of literature.

The AMSTAR score evaluates whether a SR is conducted in appropriate way, but still it neglects information regarding the individual articles included in the SR. To overcome this issue, the AMSTAR score was integrated with the LRD score. The Level of Research Design Scoring has been previously adopted in SR of SRs (List and Axelsson, 2010), and it assigns a score to the design of the individual studies according to the hierarchy of evidence (Cooke, 1996; Antes, 1998).

Only one SR (Olibone *et al.*, 2006) included non-controlled studies, book chapters and expert opinions (LRD III-IV-V). This SR showed also the lowest AMSTAR score (AMSTAR 2) and presented a structure closer to a narrative review than to a SR, without providing any definite conclusion. However, it was included in our study because the methodology of the literature search reflects some of the principles of a SR.

Most of the included reviews (6 SRs) included only CCTs. Even if RCTs are considered the best way to investigate the efficacy of dental interventions and to compare different treatment alternatives (Pocock, 1996), and MAs of RCTs are considered one of the highest level of evidence (Garg *et al.*, 2008; Hadorn *et al.*, 1996; Guyatt *et al.* 2000) only 3 of the included SRs (Chen *et al.*, 2002; Marsico *et al.*, 2011; Thiruvenkatachari *et al.*, 2013) investigated only RCTs. The number of RCTs included in these SRs was variable [6 for Chen *et al.* (2002), 4 for Marsico *et al.* (2011) and 17 for Thiruvenkatachari *et al.* (2013)] and only 2 studies overlapped in the 3 searches, because of different inclusion and

exclusion criteria. Interestingly, one of the SR of RCTs (Chen *et al.*, 2002) was judged of low quality with the AMSTAR score (AMSTAR 3), demonstrating that even the results of a SR of RCTs, which pretends to be the highest level of evidence, have to be carefully interpreted as major methodological flaws can affect the quality of the SR.

V.5.2. Main outcomes and rating of the evidence

To not provide a simple narrative summary of the results and to assess the quality of body evidence, a predetermined scale of statements was adopted for each of the outcomes analysed. This instrument has been previously adopted in a Cochrane SR of SRs (Ryan *et al.*, 2011; 2014), to not re-assess the quality of the studies included within reviews. In the current study, it was not possible to adopt the GRADE approach (Atkins *et al.*, 2004) as suggested by the Cochrane Collaboration, as 'Summary of findings' tables were not reported in any of the included SRs, except for the Cochrane SR (Thiruvenkatachari *et al.*, 2013) and frequently raw data were not available. The difficulties encountered in our study when synthesising the data extracted from the included SRs and MAs were mainly due to the variability of the inclusion criteria and to the heterogeneity of samples, outcomes, cephalometric landmarks and analysis. Our study pointed out a strong weakness in the initial diagnosis of skeletal Class II malocclusion. All the included SRs set 'Class II malocclusion' as inclusion criterion, but none of them clearly stated how the diagnosis was performed. It was observed that treatment success with functional appliances depends on a great number of confounding variables, including the severity of the baseline conditions.

Underestimating this factor does not guarantee generalisation of the conclusion, as the sample might not properly represent the target population (Khorsan and Crawford, 2014). Results from SRs and MAs should be the cornerstone for developing practice guidelines, but due to the limited and biased evidence of the primary studies, the clinical recommendations are always reported to be weak. The most frequently reported flaws of the primary studies were as follows: methodological limitations, absence of a control-matched untreated group, variability of the treatment timing, small sample size and variability of cephalometric analysis and landmarks.

a.   *Dentoalveolar effects.* Dentoalveolar effects. According to the results provided by the included SRs and MAs, there is a good consensus in literature regarding the effect of reduction of the OVJ after functional orthopaedic treatment. Nevertheless, if the results of the functional appliances in general and of the Twin Block in particular are supported by a good level of evidence, it is not so for the Splint-Type Herbst appliance. Indeed, the SR by Flores-Mir *et al.* (2007) that provides results on this outcome is based only on three references judged of low quality by the authors. Regarding the headgear, the evidence supporting the effect on the OVJ was considered insufficient, due to the controversy of results. These controversies are probably related to the different study selection [all studies (Antonarakis and Kiliaridis, 2007) vs. RCTs (Thiruvenjatachari *et al.*, 2013)] and to the different inclusion criteria of the studies assessing this outcome. In fact, Antonarakis and Kiliaridis (2007) chose as diagnostic criterion the Class II malocclusion, while Thiruvenkatachari *et al.* (2013) selected the participants as they presented prominent upper front teeth. Therefore, it is likely to observe a greater

dental movement when the starting position of the teeth is altered.

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

Little information is reported about the long-term effects after functional treatment. In one SR (Antonarakis and Kiliaridis, 2007), it is reported that skeletal changes seem to be more temporary than dentoalveolar changes, which are more stable.

b. *Maxillary skeletal effects.* Regarding the evidence provided on maxillary growth restraint, few significant values were reported and most of them were too small to be considered clinically relevant. The best effect of SNA reduction seems to be achieved with headgear (Antonarakis and Kiliaridis, 2007), while Twin Block shows variable results between significant and non-significant (Antonarakis and Kiliaridis, 2007; Ehsani *et al.,* 2014). Non-significant values of maxillary growth control were reported with both Splint-Type and Crown- Banded Herbst, but the evidence supporting this result is insufficient due to the small number of primary studies (2 or 3 studies) on which this result is based. In addition, the quality of the individual studies was low in the SR by Flores-Mir *et al.* (2007), and even not assessed in the study by Barnett *et al.* (2008). Therefore, the current evidence from SRs is not adequate to suggest or discourage the use of Herbst appliance for maxillary skeletal growth control.

c. *Mandibular skeletal effects.* Enhancement of mandibular length and/or achievement of a more forward position of the mandible, albeit still widely discussed, are frequently desired outcomes as most of the skeletal Class II malocclusion are due to a mandibular retrusion (McNamara, 1981).

Addressing all functional appliances as a group Cozza *et al.* (2006) reported a wide range of significant and non-significant findings, providing results which are scarcely applicable in the daily practice. The variability of the results in this SR is probably due to the inclusion of retrospective studies, which are susceptible to selection bias, and studies with historical samples, which suffer from the secular growth trends, occurred within the craniofacial region over the past century (Antoun *et al.*, 2015). Moreover, data from treatment with removable and fixed appliances were pooled in this review: this choice can influence the results as the two techniques differ for working hours, length of treatment time, optimal treatment timing and mode of bite-jumping (Shen *et al.*, 2005). Considering the primary studies included in this SR, in which the pubertal peak was included in the treatment timing, clinical significance of supplementary mandibular elongation (>2 mm) was reported in all studies except one. According to this finding, the authors of this SR support the hypothesis that the short-term supplementary mandibular growth appears to be significantly larger when the functional treatment is performed at the adolescent growth spurt. Even though all the SRs and MAs included in our study set the treatment of growing subjects as inclusion criterion, none of them put efforts in assessing the skeletal age. Only in one MA (Antonarakis and Kiliaridis, 2007), the studies were included only if the age of the participants was reported. Barnett *et al.* (2008) and Flores-Mir *et al.* (2007) reported a significant elongation of the mandible with Crown-Banded and Splint-Type Herbst Appliance, respectively, but the literature supporting these outcomes was judged to be insufficient due to the small number and low quality of the primary studies. Comparing the effect of Acrylic-Splint Herbst

with Crown or Banded Herbst Appliance, the differences seem to be small and not relevant, but more research is needed on this issue.

In the MA by Perillo *et al.* (2011) on Frankel-2 appliance, a significant but small increase of mandibular total length was found. However, the sensitivity analysis pointed out a negative correlation between the quality of the included studies and the retrieved results, making questionable the clinical relevance of the findings.

The most recent MAs (Marsico *et al.,* 2011) points out an effect size of the treatment of 0.61 when comparing Class II subjects treated with different functional appliances with untreated control groups. This finding is the result of the standardisation of different cephalometric measures of mandibular length, which accounts differently for jaw divergence (Co-Pg, Co-Gn and Olp-Pg+OLp-Co). In addition, the amount of mandibular length reported as the result of the conversion of the effect size (1.79 mm) is higher than that reported in the individual studies included in the SR. This controversy pointed out that major flaws could affect also a MA of RCTs rated of high quality with the AMSTAR score.

d.   <u>*Soft tissues effects.*</u> Regarding soft tissues, better results seem to be obtained with fixed functional appliances than with removable, especially when Herbst appliance is used (Flores-Mir *et al.,* 2006; Flores-Mir and Major, 2006a; Flores-Mir and Major, 2006b). The authors report the improvement of the profile to be mainly due to the retrusion of the upper lip, rather than to the protrusion of the lower lip. However, all the SRs assessing this outcome reported controversial results based on the low-quality primary studies; hence, this evidence has to be considered insufficient.

In addition, none of the primary studies included in the three SRs assessed the changes in facial profile by means of three-dimensional scanning, which is considered a reliable, non-invasive and free of radiation technique for assessing facial form (Rongo *et al.*, 2014). Due to the superimposition of the hard tissues, conventional cephalometric analyses are considered not adequately capable to detect the soft tissue structure, so the results regarding the soft tissues effects might have been underestimated.

V.6.   **Future research**

According to our findings, the registration of the protocol and the implementation of the use of PRISMA guidelines (Liberati *et al.*, 2009) might improve the methodological quality of future SRs. In addition, the use of the GRADE as tool to assess the quality of the primary studies and to provide the strength of recommendation can give a substantial contribution to the clinical conclusions and give more values to the future evidence from SRs of SRs. Moreover, it seems more useful for future SRs to analyse more homogeneous group of patients (selected according initial diagnosis, skeletal maturation and vertical growth pattern) and appliances, as reporting an aggregate pooled effect might be misleading if there are important reasons to explain variable treatment effects across different types of patients (Garg *et al.*, 2008). Finally, the evidence from the included SRs and MAs demonstrates that more research is needed on long-term effects of functional orthopaedic treatment.

V.7.   <u>Conclusions</u>

• The SRs on functional orthopaedic treatment of Class II malocclusion present a heterogeneous methodological quality. Only two SRs were judged of high quality.

• Three of the 14 papers analysed, include only RCTs and numerous SRs report a low quality of the individual studies. Clinicians should be aware of the existent tool to assess strength and weakness of the SRs and MAs, to adequately recognise whenever limited information can be obtained from such studies.

• In general, there is still no sufficient evidence to suggest or to discourage the orthopaedic functional treatment in Class II patients. The lack of definite evidence is mainly due to the small number of primary studies for each outcome and the low quality of most of the individual studies.

• There is some evidence of reduction of OVJ with several functional appliances, except from Herbst appliance, due to the poor quality of literature, and headgear, due to the controversial results reported with this appliance.

• There is some evidence of a small maxillary growth control with headgear and Twin Block. In the short term, there is some evidence of mandibular length increasing after treatment with several functional appliances, but not with Herbst appliance, which presents poor quality of literature. However, the clinical relevance of the reported results is still questionable and long-term data are not available.

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

References

• Aelbers CMF, Dermaut LR. Orthopedics in orthodontics: part I, fiction or reality? A review of literature. American Journal of Orthodontics and Dentofacial Orthopedics. 2002; 121: 560-62.

• Antes G. Evidence-based medicine. Internist (Berl). 1998; 39: 899–908.

• Antonarakis GS, Kiliaridis S. Short-term anteroposterior treatment effects of functional appliances and extraoral traction on class II malocclusion. A meta-analysis. The Angle Orthodontist. 2007; 77: 907-14.

• Antoun JS, Cameron C, Sew Hoy W, Herbison P, Farella M. Evidence of secular trends in a collection of historical craniofacial growth studies. European Journal of Orthodontics. 2015; 37: 60-6.

• Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magrini N, Mason J, Middleton P, Mrukowicz J, O'Connell D, Oxman AD, Phillips B, Schünemann HJ, Edejer T, Varonen H, Vist GE, Williams JW Jr, Zaza S. GRADE Working Group. Grading quality of evidence and strength of recommendations. BMC Medical Research Methodology. 2004; 19: 1490.

• Barnett GA, Higgins DW, Major PW, Flores-Mir C. Immediate skeletal and dentoalveolar effects of the crown- or banded type Herbst appliance on Class II division 1 malocclusion. The Angle Orthodontist. 2008; 78: 361-9.

• Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? Plos Medicine. 2010; 7: e1000326.

• Becker LA, Oxman AD. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. 2011. Higgins JPT, Green S (editors).

• Booth A, Clarke M, Ghersi D, Moher D, Petticrew M, Stewart L. An international registry of systematic review protocols. The Lancet. 2011; 377: 108-9.

• Chen JY, Will LA, Niederman R. Analysis of efficacy of functional appliances on mandibular growth. American Journal of Orthodontics and Dentofacial Orthopedics. 2002; 122: 470-6.

• Cooke IE. Finding the evidence. In: Cooke IE, Sackett DL. Ed. Ballière's. Clinical obstetrics and gynecology. International Practice and Research. 1996; 10: 561-7.

• Cozza P, Baccetti T, Franchi L, De Toffol L, McNamara JA Jr. Mandibular changes produced by functional appliances in Class II malocclusion: a systematic review. American Journal of Orthodontics and Dentofacial Orthopedics. 2006; 129: e1-12.

• Davidoff F, Haynes B, Sackett D, Smith R. Evidence-based medicine: a new journal to help doctors identify the information they need. British Medical Journal. 1995; 31: 1085-6.

• Ehsani S, Nebbe B, Normando D, Lagravere MO, Flores-Mir C. Short-term treatment effects produced by the Twin-block appliance: a systematic review and meta-analysis. (in press) European Journal of Orthodontics. 2014; 22. doi: 10.1093/ejo/cju030

• Fleming PS, Seehra J, Polychronopoulou A, Fedorowicz Z, Pandis N. A PRISMA assessment of the reporting quality of systematic reviews in orthodontics. The Angle Orthodontist. 2013; 83: 158-63.

• Flores-Mir C, Ayeh A, Goswani A, Charkhandeh S. Skeletal and dental changes in Class II division 1 malocclusions treated with splint-type Herbst appliances. A systematic review. The Angle Orthodontist. 2007; 77: 376-381.

• Flores-Mir C, Major MP, Major PW. Soft tissue changes with fixed functional appliances in Class II division 1. The Angle Orthodontist. 2006; 76: 712-20.

• Flores-Mir C, Major PW. A systematic review of cephalometric facial soft tissue changes with the Activator and Bionator appliances in Class II division 1 subjects. European Journal of Orthodontics. 2006; 28: 586-93.

• Flores-Mir C, Major PW. Cephalometric facial soft tissue changes with the twin block appliance in Class II division 1 malocclusion patients. A systematic review. The Angle Orthodontist. 2006; 76: 876-81.

• Garg AX, Hackam D, Tonelli M. Systematic review and meta-analysis: when one study is just not enough. Clinical Journal of the American Society of Nephrology. 2008; 3: 253-60.

• Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor CD, Wilson MC, Richardson WS. Users' guides to the medical literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. Evidence- Based Medicine Working Group. The Journal of American Medical Association. 2000; 284: 1290-96.

• Hadorn DC, Baker D, Hodges JS, Hicks N. Rating the quality of evidence for clinical practice guidelines. Journal of Clinical Epidemiology. 1996; 49: 749-54.

• Higgins JP, Altman DG, Gozsche PC, Jüni, P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L., Sterne JA. The Cochrane Collaboration's tool for assessing risk of bias in randomized trials. BMC Medical Research Methodology. 2011; 343: d5928.

• Jacob HB, Buschang PH, dos Santos-Pinto A. Class II malocclusion treatment using high-pull headgear with a splint: a systematic review. Dental Press Journal of Orthodontics. 2013; 18: e1-7.

• Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, Moher D. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. The Journal of American Medical Association. 1998; 280: 278-80.

• Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M, Stevens R. Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. BMC Medical Research Methodology. 2000; 26: 537-40.

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

• Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. British Medical Journal. 2001; 323: 42-6.

• Khorsan R, Crawford C. How to Assess the External Validity and Model Validity of Therapeutic Trials: A Conceptual Approach to Systematic Review Methodology. Evidence Based Complementary and Alternative Medicine: eCAM. 2014; 2014: 694804

• Khorsan R, Crawford C. How to Assess the External Validity and Model Validity of Therapeutic Trials: A Conceptual Approach to Systematic Review Methodology. Evidence Based Complementary and Alternative Medicine: eCAM. 2014; 2014: 694804

• Lau J, Ioannidis JPA, Schmid CH. Summing up evidence: one answer is not always enough. The Lancet. 1998; 351: 123-7.

• Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for

reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. British Medical Journal. 2009; 339: b2700.

• List T, Axelsson S. Management of TMD: evidence from systematic reviews and meta-analyses. Journal of Oral Rehabilitation. 2010; 37: 430-451.

• Marsico E, Gatto E, Burrascano M, Matarese G, Cordasco G. Effectiveness of orthodontic treatment with functional appliances on mandibular growth in the short term. American Journal of Orthodontics and Dentofacial Orthopedics. 2011; 139: 24-36.

• McLain JB, Proffit WR. Oral health status in the United States: prevalence of malocclusion. Journal of Dental Education. 1985; 49: 386-96.

• McNamara JA Jr. Components of class II malocclusion in children 8-10 years of age. The Angle Orthodontist. 1981; 51: 177-202.

• Mulrow CD. Rationale for systematic reviews. British Medical Journal. 1994; 3; 597-99.

• Olibone V, Lys L, Guimarãe AS, Atta JY. Influência do aparelho propulsor Twin Block no crescimento mandibular: revisão sistemática da literatura/ Influence of the Twin Block propulsor appliance in the mandibular growth: systematic review of literature. Revista Dental Press de Ortodontia e Ortopedia Facial. 2006; 11: 19-27.

• Papadopoulos MA, Gkiaouris I. A critical evaluation of meta-analyses in orthodntics. American Journal of Orthodontics and Dentofacial Orthopedics. 2007; 131: 589-99.

• Perillo L, Cannavale R. Ferro F, Franchi L, Masucci C, Chiodini P, Baccetti T. Meta-analysis of skeletal mandibular changes during Frankel appliance treatment. European Journal of Orthodontics. 2011; 33: 84-92.

• Pocock SJ. Clinical Trials – A Practical Approach. 1996. Chichester: John Wiley & Sons Ltd.

• Rongo R, Antoun JS, Lim YX, Dias G, Valletta R, Farella M. Three-dimensional evaluation of the relationship between jaw divergence and facial soft tissue dimensions. The Angle Orthodontist. 2014; 84: 788-94.

• Ryan R, Santesso N, Hill S, Lowe D, Kaufman C, Grimshaw J. Consumer-oriented interventions for evidence-based prescribing and medicines use: an overview of systematic reviews. Cochrane Database of Systematic Reviews. 2011; 11: CD007768.

• Ryan R, Santesso N, Lowe D, Hill S, Grimshaw J, Prictor M, Kaufman C,

Cowie G, Taylor M. Interventions to improve safe and effective medicines use by consumers: an overview of systematic reviews. Cochrane Database of Systematic Reviews. 2014; 29: CD007768.

• Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Medical Research Methodology. 2007; 15: 10.

• Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, Henry DA, Boers M. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. Journal of Clinical Epidemiology. 2009; 62: 1013-20.

• Shen G, Hägg U, Darendeliler M. Skeletal effects of bite jumping therapy on the mandible - removable vs. fixed functional appliances. Orthodontics and Craniofacial Research. 2005; 8: 2-10.

• Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Medical Research Methodology. 2011; 11: 15.

• Thiruvenkatachari B, Harrison JE, Worthington HV, O'Brien KD. Orthodontic treatment for prominent upper front teeth (Class II malocclusion) in children. Cochrane Database of Systematic Review. 2013; 11: CD003452.

Chapter VI.

Evidence on dental and skeletal effects of palatal expansion techniques

VI.1. <u>Summary</u>

This Systematic Review (SR) aimed to assess the quality of SRs and Meta-Analyses (MAs) on the dentoalveolar and skeletal effects of palatal expansion techniques, to summarize the main results, and to evaluate the quality of the evidence of the reported results. Electronic and manual searches were conducted until February 2015. SRs and MAs focusing on the dentoalveolar and skeletal effects of maxillary expansion performed with fixed appliances in growing patients were included. The methodological quality of the included papers was assessed using the AMSTAR (A Measurement Tool to Assess Systematic Reviews). The design of the primary studies included in each SR and MA was assessed with the LRD (Level of Research Design scoring). The main outcomes were summarized and the quality of the evidence within SRs was rated according to a pre-determined scale of levels of evidence. Twelve SRs/MAs fulfilled the inclusion criteria. The appliances evaluated were as follows: Hyrax (bonded or banded) (8 studies), Haas (bonded or banded) (7 studies), Quad Helix (QH) (6 studies), Minne-expander (3 studies), Nitanium maxillary expander (2 studies), bone-anchored maxillary expander (3 studies), and expansion arch (1 study). The mean AMSTAR score was 6.8 (ranged 4–10). Two papers included only studies with randomization (Randomized Clinical Trials – RCTs), four papers included only Clinical Controlled Studies (CCTs), two papers included RCTs and CCTs, one paper included CCTs and non-controlled trials, one paper included RCTs and non-controlled trials, and two papers included RCTs, CCTs and non-controlled trials. The current evidence from SRs and MAs provided high quality results as concerned the dentoalveolar short-term expansion obtained with rapid maxillary expansion (RME) at maxillary molar, premolar, canine and mandibular molar region. There was moderate evidence regarding the short-term dentoalveolar expansion achieved with slow maxillary expansion (SME) technique and the

long-term effect of RME. No differences of dentoalveolar effects between the two expansion modalities (RME vs. SME) were also reported with moderate evidence. There was low evidence of skeletal transversal effects of RME and long-term dentoalvolar effects of SME. Short-term skeletal effects of SME were reported with very low evidence. Finally, effects other than transversal (vertical, sagittal) were all reported to be small and supported by very low evidence.

KEYWORDS: palatal expansion technique, crossbite, rapid maxillary expansion, slow maxillary expansion, review literature as topic, evidence-based dentistry, adolescent, growth and development

VI.2. <u>Background</u>

Posterior Crossbite (PXB) is a malocclusion affecting canine, premolar and molar region, in which the buccal cusps of the maxillary teeth occlude lingually to the buccal cusps of the corresponding mandibular teeth (Björk, 1964). Its prevalence ranges between 8 and 22% of children in primary/mixed dentition (Lindner and Modeer, 1985; Kurol and Bergland, 1992; Tscill *et al.*, 1997; Lux *et al.,* 1997). Several anatomical and myofunctional findings related to untreated PXB have been reported (Bell and Kiebach, 2014), such as asymmetric condylar positioning (Hesse *et al.,* 1997), asymmetric mandibular growth (Pirtiniemi *et al.*, 1990), dental discrepancies and dental asymmetries with Class II tendency on the crossbite side (Allen *et al.*, 2003), but restoration of normal growth and function have been extensively documented after crossbite treatment (Myers *et al.*, 1980; Lam *et al.*, 1999).

Maxillary expansion is one of the treatments of choice proposed for the resolution of PXB, especially in case of skeletal constriction of the upper jaw, with the intent of increasing the transverse widths of the maxilla through the opening of the mid-palatal suture (Lagravere *et al.*, 2005; Martina *et al.*, 2012). This purpose can be achieved with different modalities (Castañer-Peiro, 2006).

When plenty of primary literature on a given crucial topic exists, Systematic Reviews (SRs) and Meta-Analyses (MAs) are considered the gold standard to provide pre-filtered evidence (Mulrow, 1994; Higgins and Green, 2011; Moher *et al.*, 2009). However, due to the rapid diffusion of such approach, clinicians and researchers who previously had to deal with the huge amount of primary studies, now have to

manage the number of SRs and MAs on the same topic, frequently variable in quality and objectives (Smith *et al.*, 2011) Moreover, it is likely to find different results provided by SRs of different quality level (Moher *et al.*, 2002). Hence, when a number of SRs and MAs exist in priority scientific areas, an Overview of reviews is the type of research suggested to summarize and appraise multiple results (Smith *et al.*, 2011; Higgins and Green, 2011).

Such approach has been previously adopted in the orthodontic field to synthesise the evidence on the controversial issue of Class II orthopaedic functional treatment (D'Antò *et al.*, 2015), but to the best of our knowledge this is the first SR of SRs on the effects of palatal expansion techniques.

In one previous SR the results of published MAs in orthodontics were critically summarized and discussed (Papadopoulos and Gkiaouris, 2007). Among those, MAs on transversal problems were reported, but palatal expansion was not considered as main aim of the SR, and the authors used for the literature search generic keywords embracing the whole orthodontic field.

Differently, the focus of the current study was to conduct a comprehensive literature search exclusively on the effects of palatal expansion, and to collect the evidences from both SRs and MAs. More specifically, the aims of the present study were: (1) to evaluate the methodological quality of SRs and MAs on dental and skeletal effects of palatal expansion, and (2) to summarize the reported effects of treatment by appraising the evidence on which the results are based.

VI.3. Materials and methods

V.3.1. Studies selection and data collection

For the current study all the SRs and MAs concerning dental and skeletal effects of palatal expansion techniques were analysed. The databases investigated for the systematic literature search were: Medline (Entrez PubMed, www.ncbi.nlm.nih.gov), the Cochrane Library (www.cochranelibrary.com), Latin American and Caribbean Health Sciences (LILACS, http://lilacs.bvsalud.org), Scientific Electronic Library Online (SciELO, http://www.scielo.org), Web of Knowledge (WOK, https://webofknowledge.com/) and Scopus (http://www.scopus.com/). The survey covered the period from the starting of the databases up to June 2014. The search was later updated covering the period between June 2014 and February 2015. In order to include possible overlooked papers a further hand-search of orthodontic journals (*European Journal of Orthodontics*, *American Journal of Orthodontics and Dentofacial Orthopedics* and *The Angle Orthodontist*) was performed, starting from the first volume available on the digital archives. Moreover, an effort of exploration of the grey literature (unpublished studies) was performed among the conference abstracts published on WOK and Scopus databases and on the databases of scientific congresses (European Orthodontic Society and International Association of Dental Research).

The search strategies applied for each database are shown in Table VI.1. The detailed PubMed search is reported in Table VI.2.

The *inclusion criteria* were: (1) Systematic Reviews or Meta-Analyses; (2)

Studies assessing dentoalveolar and/or skeletal effects of palatal expansion techniques; (3) Treatment performed with fixed orthodontic expansion appliances. The *exclusion criteria* were: (1) Dual publication; (2) Systematic Reviews of SRs; (3) Surgically–assisted rapid maxillary expansion (SARME); (4) Cleft lip/palate diagnosis or craniofacial syndrome diagnosis; (5) Expansion treatment performed in association with protraction headgear/facemask therapy; (6) Updated publications; (7) SRs or MAs focusing on treatments strategies others than fixed appliances (grinding/removable appliances).

Two operators read all titles and abstracts to identify the potentially eligible papers. Whenever the information provided by only title and abstract arouse doubts, the reference was included for full-text reading. Due to language limitations two out of six databases (LILACS and SciELO) were analysed by one operator (R.B.). Subsequently, the full-texts of the references that seemed to fulfil the inclusion criteria were entirely read, and only the papers that completely satisfied all the inclusion criteria were selected. When a paper addressed different interventions (e.g. data about fixed and removable appliances) the SRs or MAs was included, discarding the un-necessary data.

To identify any further relevant missing paper, the reference lists of the included SRs and MAs were analysed.

Data about authors, year of publication, study design, number of subjects, diagnosis, intervention, expansion modality, outcome, methods, quality of the primary studies, results, author's conclusions and author's comments on quality of the studies were extracted from all the included SRs and MAs. When relevant data were not available in the publication, the authors of the review were contacted to obtain further information.

All the stages of study selection and data extraction were independently run by two operators (R.B. and F.D.). Disagreements between the two examiners were discussed and solved to reach consensual decision. If necessary, a third operator (V.D.) was contacted for the final decision.

Database	Search Strategy	Results
Pubmed	(((((((("Palatal Expansion Technique"[Mesh]) OR (maxillary expansion OR palatal expansion))) AND (("Meta-Analysis" [Publication Type]) OR "Review" [Publication Type]))) NOT "Craniofacial Abnormalities"[Mesh]) NOT "Malocclusion, Angle Class III/therapy"[Mesh]) NOT "Orthognathic Surgery"[Mesh]) NOT "Cleft Palate"[Mesh]	119
WOK	TOPIC: (palatal expansion OR maxillary expansion) AND TOPIC: (review OR meta-analysis) NOT TOPIC: (craniofacial syndrom*) NOT TOPIC: (surg*) NOT TOPIC: (angle class III) NOT TOPIC:(cleft palate)	105
Scopus	(TITLE-ABS-KEY (palatal expansion OR maxillary expansion) AND TITLE-ABS-KEY (review OR meta-analysis) AND NOT TITLE-ABS-KEY (craniofacial syndrom*) AND NOT TITLE-ABS-KEY (surg*) AND NOT TITLE-ABS-KEY (cleft palate) AND NOT TITLE-ABS-KEY (angle class iii))	69
SCIELO	(palatal expansion OR maxillary expansion) AND (review OR meta-analysis)	2
Lilacs	(palatal expansion OR maxillary expansion) AND (review OR meta-analysis)	18
Cochrane Library	MeSH descriptor: [Palatal Expansion Technique] explode all trees	1

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

	Search strategy	Results
#1	"Palatal Expansion Technique"[Mesh]	2048
#2	maxillary expansion OR palatal expansion	2977
#3	#1 OR #2	2977
#4	#3 AND ("Meta-Analysis" [Publication Type]) OR "Review" [Publication Type]	151
#5	#4 NOT "Craniofacial Abnormalities"[Mesh] NOT "Malocclusion, Angle Class III/therapy"[Mesh] NOT "Orthognathic Surgery"[Mesh] NOT "Cleft Palate"[Mesh]	119

Table VI. 2 PubMed search strategy

V.3.2. Quality of the included SRs and MAs

For each included SR and MA the methodological quality was assessed using "A Measurement Tool to Assess Systematic Reviews" (AMSTAR) (Shea e*t al.*, 2007). AMSTAR is composed by 11 items, each can be answered "Yes", when clearly done, "No", when clearly not done, "Not Applicable", when the item is not relevant, such as when a MA was not attempted by the authors, "Can't answer", when the item is relevant but not described by the authors. Each "Yes" answer is scored 1 point, while the other answers are scored 0 point. According to the number of criteria met, the quality of the included papers was rated as "Low" (total score ≤3); "Medium" (total score from 4 to 7); "High" (total score ≥8) (Ryan *et al.*, 2011; 2014).

In addition, to assess the design of the primary studies included in each SR or MA, the Level of Research Design scoring (LRD) was used (Antes, 1998; Cooke, 1996). The interpretation of such score, based on the hierarchy of evidence, is shown in Table VI.3.

For each included SR or MA two investigators (R.B. and F.D.) independently assessed the methodological quality, with no blinding for

the authors of the review.

The inter-examiner reliability for the AMSTAR scores was calculated by means of Cohen's k coefficient. Nonetheless, disagreements and discrepancies on the AMSTAR scoring were discussed and solved to reach a unanimous score.

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

Table VI. 3 Interpretation of the LRD scores. The scores are based on the type of studies included in the SR

V.3.3. Synthesis of the results and quality of the body evidence

The main outcomes of the included SRs and MAs were summarized according to: timing (short- or long-term effects), structure involved (dentoalveolar or skeletal effects), direction of the effect (transversal, vertical or sagittal), expansion modality (Slow Maxillary Expansion - SME or Rapid Maxillary Expansion - RME) and appliance.

For each outcome, the quality of the body evidence was rated according to a pre-determined, objective and transparent set of levels of evidence, which resemble the criteria considered within the GRADE approach (Guyatt *et al.*, 2008). The criteria used for the assessment of the level of evidence for each outcome were based on the systematic evaluation of: the way the data were pooled (MA or SR with narrative synthesis), the

number of studies, the number of participants, and the quality of the primary studies assessing the outcome. To ensure the reproducibility of this assessment, full explanation of the method and the cut-offs for decision-making are reported in Table VI.4. The quality of the individual studies was not re-assessed, but reported as assessed by the authors of the reviews.

	Type of study	n. of participants	n. of studies	Quality of primary studies
No downgrade	Meta-Analysis	>200	>10	low for <50% of the included studies
1 downgrade	Qualitative Synthesis	100-200	6-10	low between 50 and 75% of the included
2 downgrades		0-99	1-5	low for >75% of the included studies

If the "Quality of primary" study was not reported we were conservative and assumed as 1 downgrade. If the "n. of participants" was not reported we assumed the same downgrade of the "n. of studies".

Table VI. 4 Objective criteria for rating the evidence of the reported outcomes.

high	no or 1 downgrade
moderate	2-3 downgrade
low	4-5 downgrade
very low	>5 downgrade

Table VI. 5 Levels of evidence according to the number of downgrades.

The evidence from each outcome was classified as: *very low, low, moderate* or *high*, according to the number of downgrades (Table VI.5). It was assumed that for *high* evidence further research was very unlikely to change our confidence in the estimate of effect; for *moderate* evidence further research was likely to have an important impact on our confidence in the estimate of effect and may change the estimate; for *low* evidence further research was very likely to have an important impact on our confidence in the estimate of effect and was likely to change the estimate; for *very low* evidence any estimate of effect was very uncertain.

VI.4.   <u>Results</u>

VI.4.1.   Studies selection

The updated electronic search of all databases resulted in 314 references. After duplicates were removed, 224 references were left. After reading title and abstract 193 references were excluded, mainly because the topic was not pertinent to the current study or because they were narrative reviews. The remaining eligible 31 articles were entirely read and 19 of them were excluded (Fig. VI.1; Table VI.6). The most common exclusion criterion was the absence of a systematic search strategy. The 12 SRs included and the data extracted from each study are reported in Table VI.7. (Schiffman and Tuncay, 2001; Petren *et al.*, 2003; Lagravere *et al.*, 2005a; Lagravere *et al.*, 2005b; Lagravere *et al.*, 2005c; Lagravere *et al.*, 2006; De Rossi *et al.*, 2008; Zuccati *et al.*; 2011; Zhou *et al.*, 2013; Bazargani *et al.*, 2013; Lione *et al.*, 2013; Agostino *et al.*, 2014). Four out of 12 studies were integrated with MA (Schiffman and Tuncay, 2001; Lagravere *et al.*, 2006; Zhou *et al.*, 2013; Agostino *et al.*, 2014). The total number of subjects included ranged from 89 to 997. Two studies did not report the number of subjects (Schiffman and Tuncay, 2001; Zuccati *et al.*, 2011) while in another study this information was only partially available (Lione *et al.*, 2013).

The initial diagnosis was 'unilateral or bilateral posterior crossbite' in two studies (Schiffman and Tuncay, 2001; Lagravere *et al.*, 2005b) while unspecified posterior crossbite, transverse discrepancy or constricted arches were reported in four studies (Lagravere *et al.*, 2005c; Zuccati *et al.*, 2011; Zhou *et al.*, 2013; Agostino *et al.*, 2014), and no initial diagnosis was specified in half of the studies. In one SR only primary studies

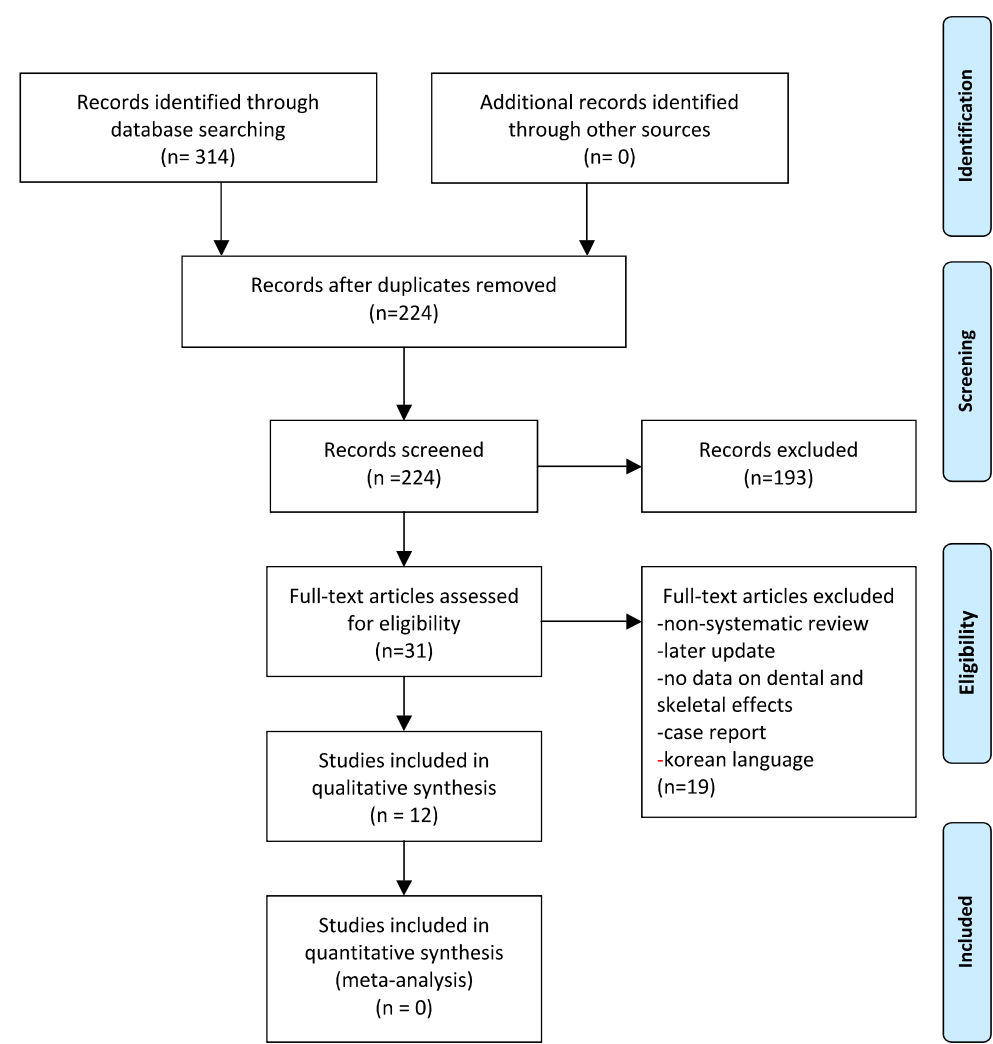performed in early and mixed dentition were included (Petren *et al.*, 2003).

Figure VI. 1 PRISMA Flow Diagram

Reference	Reason for the exclusion
- Bell RA. A review of maxillary expansion in relation to rate of expansion and patient's age. Am J Orthod. 1982; 81: 32-7.	Narrative Review
- Bishara SE, Staley RN. Maxillary expansion: clinical implications. Am J Orthod. 1987; 91: 3-14.	Narrative Review
- Cerrone Júnior G., Santiago R.C., Martins M.T., de Paula M.V.Q. Controversies about midpalatal suture and maxillary expansion: Literature Review. Pesquisa Brasileira em Odontopediatria e Clinica Integrada. 2011; 11: 297-303.	Narrative Review
- De Coster T. Orthopedic expansion of the maxilla. Orthod Fr. 2006; 77: 253-64.	Narrative Review
- Glineur R, Boucher C, Balon-Périn A. Interceptive treatments (ages 6-10) of transverse deformities: posterior crossbite. Orthod Fr. 2006 Jun; 77:249-52.	Narrative Review
- Gola G, Masini M, Purro C, De Martini S. Biological fundamentals and general therapeutic ideas on transverse expansion of the upper jaw (panoramic literature review on the state of the art). Parodontol Stomatol (Nuova). 1984; 23:133-40.	Narrative Review
- Hodges S, Cunningham S. Transverse expansion in the upper arch. Prim Dent Care. 1998; 5: 63-7.	Narrative Review
- James G, Strokon D. Cranial strains and malocclusion VIII: palatal expansion. Int J Orthod Milwaukee. 2009; 20: 15-30.	Narrative Review
- Jones SP, Waters NE. The quadhelix maxillary expansion appliance: Part II: Clinical characteristics. Eur J Orthod. 1989; 11:195-9.	Narrative Review
- Kennedy DB, Osepchook M. Unilateral posterior crossbite with mandibular shift: a review. J Can Dent Assoc. 2005; 71: 569-73.	Narrative Review
- McNamara JA. Maxillary transverse deficiency. Am J Orthod Dentofacial Orthop. 2000; 117: 567-70.	Narrative Review
- Roveta A.L., Bolasco L. Disjunction with different elements of skeletal anchorage. Actas odontol. 2013; 10: 19-27.	Narrative Review
- Saldarriaga J.R. Rapid palatal expansion. Revista de la Facultad de Odontología Universidad de Antioquia. 1991; 2: 89-95.	Narrative Review
- Sorel O. Rapid palatal expansion for the treatment of maxillary constriction. Rev Stomatol Chir Maxillofac. 2004; 105: 26-36.	Narrative Review
- Harrison JE, Ashby D. Orthodontic treatment for posterior crossbites. Cochrane Database Syst Rev. 2000: CD000979.	Updated in a later publication
- Harrison JE, Ashby D. Orthodontic treatment for posterior crossbites. Cochrane Database Syst Rev. 2001: CD000979.	Updated in a later publication
- Romanyk DL, Lagravere MO, Toogood RW, Major PW,	No data on dental or skeletal

Carey JP. Review of Maxillary Expansion Appliance Activation Methods: Engineering and Clinical Perspectives. Journal of Dental Biomechanics 2010; 2010: 496906.	effects
- Zhu JF, Crevoisier R, King DL, Henry R, Mills CM. Posterior crossbites in children. Compend Contin Educ Dent. 1996; 17: 1051-4.	Case Report
- Ko CH, Lim SH, Yoon YJ, Kim KW. A meta analysis of maxillary expansion: comparisons of intercanine/intermolar expansion and rapid/slow expansion. The Korean Journal of Orthodontics. 2004; 34: 23-32	Article in Korean

Table VI. 6 References excluded after the full-text reading and reason for the exclusion

The appliances studied were: Hyrax (bonded or banded) (Schiffman and Tuncay, 2001; Petren *et al.*, 2003; Lagravere *et al.*, 2006; De Rossi *et al.*, 2008; Zhou *et al.*, 2013; Bazargani *et al.*, 2013; Lione *et al.*, 2013; Agostino *et al.*, 2014), Haas (bonded or banded) (Petren *et al.*, 2003; Lagravere *et al.*, 2005b; Lagravere *et al.*, 2006; Zhou *et al.*, 2013; Agostino *et al.*, 2014), Quad Helix (QH) (Schiffman and Tuncay, 2001; Petren *et al.*, 2003; Lagravere *et al.*, 2005c; Zhou *et al.*, 2013; Lione *et al.*, 2013; Agostino *et al.*, 2014), Minne-expander (Lagravere *et al.*, 2005c; Zhou *et al.*, 2013; Agostino *et al.*, 2014); Nitanium maxillary expander, (Lagravere *et al.*, 2005c; Lione *et al.*, 2013), bone-anchored maxillary expander (Bazargani *et al.*, 2013; Lione *et al.*, 2013; Agostino *et al.*, 2014), and expansion arch (Agostino *et al.*, 2014).

The main outcomes were: short-term dental changes (Petren *et al.*, 2003; Lagravere *et al.*, 2005c; Lagravere *et al.*, 2006; Bazargani *et al.*, 2013), long-term dental changes (Lagravere *et al.*, 2005b; Zuccati *et al.*, 2011; Zhou *et al.*, 2013), short-term skeletal changes (Lagravere *et al.*, 2005c; Lagravere *et al.*, 2006; De Rossi *et al.*, 2008; Bazargani *et al.*, 2013; Lione *et al.*, 2013), and long-term skeletal changes (Lagravere *et al.*, 2005a; Schiffman and Tuncay, 2001; Zuccati *et al.*, 2011).

Author, year	Study design, total n°of subjects	Diagnosis	Intervention (I) or Appliance (A) and control groups (C)	Expansion Technique	Outcome measures	Methods/ Measurement	Quality of the primary studies	Results	Author's Conclusion (C) and Author's Comments on the Quality of the Papers (Q)
Schiffman & Tuncay, 2001	MA of 2 RCTs, 2 P CCTs and 2 R CCTs; n. of subjects NR	unilateral or bilateral posterior crossbite	I1: Hyrax I2: QH C: Untreated/ Absent/Alterna tive Treatments	SME and RME	Long-term skeletal transversal effects (stability)	NR	**Self-produced meta-analytic score:** *unclear*	The mean expansion was 6.00 mm±1.29. While wearing retention in the short-term (<1 yr) the 78.5% (4.71mm) of the original expansion was manteined. While wearing retention in the long term retention (> 1 yr) the 92% of the original expansion was manteined. Post-retention data show a total loss of 35.5% of the original transverse increase. Studies reporting short-term post-expansion data maintained 75% (3.88 mm), while	C: Maxillary expansion stability is minimal. Post retention data show that the residual expansion is no greater than what has been documented as normal growth. Q: nr

								longer-term post-expansion data (>50 months) demonstrated a mean loss of 40% of the expansion (residual expansion 2.4). I1 shows the 50% of relapse, while I2 around 64%.	
Petren *et al.*, 2003 ~	SR of 2 RCTs, 5 P CCTs, 5 R CCTs * 695 subjects	Primary and early mixed dentition with posterior crossbite	I1: QH I2: RME (Hyrax/Haas) C: no treatment/alternative treatment	unclear	Dental effects (of the early treatment)	Dental cast, posteroanterior radiographs, lateral radiographs, Clinical Examination, Photos	**Self-produced checklist:** *8 low quality, 4 medium quality*	100% or close to 100% of success rate when using I1 or I2. Spontaneous correction was found to occur in 16% to 50% of the untreated control groups. Highest amount of correction with I1 (Mmw: from 3.3- to 6.4mm and Mcw from 1.3-to 5.2mm), followed by I2 (Mmw 5.5 mm, Mcw 3.2 mm). Regarding stability after retention, higher	C: QH and RME are effective in the early mixed dentition at a high success rate. However, there is no scientific evidence available that shows which of the treatment modalities is the most effective. There is limited evidence for stability of crossbite correction at least 3 years post-treatment.

								values with I2 (Mmw: 5.4mm, Mcw: 3.3mm) followed by I1 (Mmw: from 3.6- to 5.1 mm Mcw: from 2.2- to 3.3mm)	Q: Primary studies lack of power because of small sample size, bias, and confounding variables, lack of method error analysis, blinding in measurements, and deficient or lack of statistical methods
Lagravere et al., 2005a	SR of 3 R CCTs; 113 subjects	-	I1: Type of appliance NR C: no treatment/fixed appliance	RME	Long-term skeletal transversal, anteroposterior and vertical effects	Posteroanterior radiographs, lateral radiographs	**Self-produced checklist:** *2 low (<50% of checks); 1 medium (>50% of checks)*	<u>Transverse changes</u>: Statistical increase of lateronasal width. Statistically significant increase of maxillary width in patients treated before growth peak. <u>Anteroposterior changes:</u> No significant difference <u>Vertical changes:</u> A statistically significant long-term difference was present in the SN-PP (0.88) and	C: Long-term stability of transverse skeletal maxillary increase is better in skeletally less mature individuals (prepubertal growth peak). Long-term transverse skeletal maxillary increase is approximately 25% of the total appliance adjustment (dental expansion) in prepubertal

								SN-Gn (0.88) angles when comparing RME vs. C. Mandibular plane reduction (-0.85°) was lower than that reported in C.	adolescents but not significant for postpubertal adolescents. RME did not produce significant anteroposterior or vertical changes. Q: Lack of description of a statistical estimation process for the sample size, dropouts, and intra- and interexaminer reliability. Long term RCT are required.
Lagravere et al., 2005b	SR of 3 R CCTs and 1 P CCT; 412 subjects	unilateral and bilateral posterior crossbite	I1:Haas (+fixed treatment) C:untreated	RME	Long-term dental effects (Mmw; Mpw; Mcw; mmw; mcw; OVJ; molar extrusion;	Dental cast, posteroanterior radiographs, lateral radiographs	-	I1 increases Mmw (between 4.8-and 2.7-mm), Mpw (between 4.7-and 3.7-mm) and Mcw (between 2.5- and 2.2-mm) and mmw (between 5.4- and 0.7-mm)	C: Clinically significant long-term maxillary molar width increase can be achieved with I1. Because of crown tipping, the amount of reported long-

					incisor inclination and OVB.)			and mcw (between 1.8- and 0.8-mm) I1 decreases OVJ (0,6-mm), while no statistically significant difference were found for molar extrusion, incisor inclination and OVB when compared with C.	term width increase varied with the reference point used for measurements. More transverse dental arch changes were found after puberty compared with before puberty. The difference may not be clinical significant (0.8 mm). No anteroposterior or vertical dental changes were associated with RME with Haas expander. Q: Lack of clear statement regarding the retention protocol and different landmarks for measurements.

Lagravere et al., 2005c	SR of 5 P CTs and 4 non-controlled CTs; 89 subjects	Constricted arches	I1: Minne-expander (banded or bonded) I2: QH I3: Nitaniun maxillary expander	SME	Skeletal and dental effects	Dental cast, posteroanterior radiographs, lateral radiographs	-	I1 increase the transversal width, with a skeletal response from the 28 to the 50% of the total expansion. No difference between bonded and banded I1. I2 can determine small percetage of skeletal expansion, but the major effect is dentoalveolar. More skeletal effecta are obtained in younger patients. I3 effects similar to those of I1.	C: No strong conclusion can be made on dental or skeletal changes after SME. Q: Absence of control group for all of the included studies
Lagravere et al., 2006	MA of 3 non-controled CTs, 9 P CCTs and 2 RCTs; 335 subjects	-	I1: Hyrax (with or withuot acrylic bite plates) I2: Haas C: Absent / Alternative treatments	RME	Immediate transversal anteroposterior and vertical dental and skeletal effects	Dental cast posteroanterior radiographs, lateral radiographs	**Self-produced checklist:** *14 low (<50% of checks)*	Transverse dental changes. Increase of Mmw (from 6.04- to 6.74-mm), maxillary intermolar mesioapex root width (4.44 mm), maxillary intermolar angulation	C: The greatest changes were in the maxillary transverse plane and they were more dental than skeletal. Few vertical and anteroposterior changes were statistically

								(approximately 3.10 degrees), Mcw (5.35 mm; when measured from the crown apex). PA cephalometric radiographs showed a 3.9 mm increase in the maxillary interincisal apex width and a 2.98 mm increase in the midline diastema. Non-statistically significant mmw changes. <u>Vertical and anteroposterior dental changes</u>. Statistically significant extrusion of the maxillary molar cusp (0.53 mm), increase in overjet (1.29 mm) and change in angulation of the maxillary incisor to sella nasion (SN) plane (0.86	significant, and none was clinically relrvant. Q: All of the studies had methodological problems regarding sample selection, description and statistical approach.

								degrees). <u>Transverse skeletal changes</u> Significant increase of nasal cavity width(intercondyle width ; 2.14mm) and left jugale-right jugale (interalveolar width 2.73mm) <u>Vertical and anteroposterior skeletal changes.</u> Statistically significant changes in the mandibular plane (with respect to the palatal plane - 1.65mm and SN plane - 1.97mm). No stastically significant changes of the palatal plane.	
De Rossi *et al.*, 2008	SR of 4 P CCTs; 152 subjects	-	I1: BREMA C: Alternative treatments	RME	Vertical and sagittal skeletal effects	Lateral radiographs	-	I1 produces downward movement of the maxilla and downward and backward rotation of the mandible,	C: Vertical effects are only partially controlled with bonded devices, there is no consensus in the literature

								although the vertical effects seem to be smaller than those of the banded appliances.	regarding the maxillary sagittal displacement after RME; there is not sufficient evidence to support the use of BRMEA to control the undesirable effects of RME. Q: nr.
Zuccati *et al.*, 2011	SR of 12 RCTs; n. of subjects NR	Posterior Crossbite	I1:QH, C: Untreated control/Alternative treatments	NR	Long-term transversal effects (stability)	Dental casts, lateral radiographs, CBCT, CT	**Cochrane collaboration tool for assessing risk of bias:** *unclear*	I1 shows stable results after 3 years from the treatment.	C: Many treatments appear to be successful in the short ter, but challenges remain in the search for better long-term outcomes. Q: small sample size, bias and confounding variables, lack of blinding in measurements, and deficient statistical methods

Zhou *et al.*, 2013	MA of 2 RCTs and 12 CCTs; 997 subjects	Transverse discrepancy	I1: RME I2: SME A1: QH, A2: Hyrax A3: Haas A4: Minne-expander C: no treatment/alternative treatment	RME and SME	Long-term dental effects (post-expansion/ retention and net change): Mmw, Mcw, Mpw and mmw	Dental cast, lateral radiographs	**Cochrane collaboration tool for assessing risk of bias:** *3 low quality, 9 medium quality, 2 high quality*	I1vsControl: Mmw: Signficant increase post expansion (4.45 mm) and in the net change (2.49 mm). Mcw: Signficant increase Post expansion (2.58 mm) and in the net change (2.27 mm). Mmw: Signficant increase Post expansion (0.49 mm) and in the net change (0.06 mm). I2vsControl: Mmw: Signficant increase Post expansion (4.09 mm) followed by a non-significant relapse (−0.40 mm). Signficant net change (3.58 mm). Mcw: Signficant increase Post expansion (2.7 mm) followed by a non-significant relapse (−0.41 mm).	C: SME is effective in expanding maxillary arch, while we cannot determine its effectiveness in mandibular arch expansion. RME is effective in expanding both maxillary and mandibular arches. Furthermore, SME is superior to RME in expanding molar region of maxillary arch. Q: Limited number of high-quality studies, different treatment strategies (expansion only or expansion plus fixed orthodontic treatment), and different measurement

| | | | | | | | | | Signficant net change (2.64 mm). Mpw: Signficant increase Post expansion (3.86 mm) followed by a non-significant relapse ($-$0.16 mm). Signficant net change (3.52 mm). mmw: Signficant increase Post expansion (1.19 mm) followed by a significant increase in the retention period (0.65 mm). Signficant net change (2.02 mm). I1vsI2 Mmw no significant changes in the post expansion and retention period; net change significanty higher in SME. Mcw, Mpw and mmw: no significant difference. | landmarks. |

| Bazargani et al., 2013 | SR of 2 RCTs and 8 P non-controlled CTs; 256 subjects | - | I1: Haas
I2: Hyrax
I3: bone-anchored maxillary expander,
I4: tooth-anchored maxillary expander

C: Absent / Alternative treatment | RME | Immediate dental and skeletal effect | 3d images from CBCT | **Self-produced checklist:** *8 low quality, 2 medium quality* | <u>Dental Structures</u>: Buccal tipping of the first molars between 7.5° and 1.0°. <u>Midpalatal suture:</u> The mean posterior expansion ranged from 1.6 to 4.33 mm; the mean anterior expansion ranged from 1.52 to 4.33 mm, corresponding also to 22%– 53% of total screw expansion. <u>Nasal Cavity:</u> Expansion range from 1.2- and 2.73-mm. <u>Circummaxillary Sutures:</u> Overall small changes (0.30- and 0.45-mm). <u>Spheno-occipital Synchonrosis:</u> mean expansion of 0.6 mm. <u>Orbital structures:</u> small increase in volume (0.72 mL) and width (1.09 mm) | C: Midpalatal suture amounted to 20%–50% of the total screw expansion. RME produced immediate significant changes in transverse dimensions of the nasal cavity, circummaxillary sutures, spheno-occipital synchondrosis, and aperture width.

Q: Frequent shortcomings in study design, sample size, and inadequate selection description..The majority of the articles were judged to be of low quality therefore, no evidence-based conclusions could |

								but clinically insignificant.	be drawn from these studies.
Lione *et al.*, 2013**	SR of 13 R non-controlled CTs, 2 P CTs, 9 R CCTs 3 P CCTs; 4 RCTs 807 subjects in 28 articles ***	-	I1: Hyrax (bondend or banded) I2: Haas (bonded or banded I3: Bone anchored expander I4: NiTi expander I5: QH C: Absent/no treatment/alter native treatment	RME	Skeletal transversal and vertical effects	Dental casts, posteroanterior radiographs, lateral radiographs, tridimensional radiographic techniques;	**Self-produced checklist:** *20 low quality, 8 medium quality, 2 medium-high quality*	Midpalatal suture opening range from 1.1 to 4.8 mm in the anterior region and from 1.2 to 4.3 mm in the posterior region. At the end of the active phase, RME resulted in a slight inferior movement of the maxilla (SN-PNS +0.9 mm; SN-ANS +1.6 mm). This downward movement of the maxilla and premature dental contacts are responsible for the mandible rotation in a downward and backward direction with a mean increase in the following variables: SN-MP, 1.7°; PP-MP, 1.5°;	C: RME always opened the midpalatal suture in growing subjects, regardless the type of appliance. The vertical changes were small (less than 2 mm or 2°, and may be not considered clinically relevant) and transitory. Q: The methodology of these investigations was generally of low and medium quality therefore the findings should be interpreted with caution. Small sample size with no consecutive cases, bias and

								SN-GoGn, 1.1°.	confounding variables, lack of method error analysis, blinding in measurements, deficiency or lack of statistical methods and absence of power analysis. A very serious limitation of most studies was the lack of an adequate untreated control group.
Agostino *et al.,* 2014~	Cochrane MA of 15 RCTs; 657 subjects	posterior crossbite	I1: Hyrax (tooth or tooth-tissue borne) I2: Hass (tooth or tooth-tissue) I3: Bone-anchored maxillary expander I4: Minne Expander (bondend or banded) I5: QH I6: expansion arch	RME and SME	Compariso n between appliances for the correction of crossbite	Dental cast, posteroanterior radiographs, lateral radiographs	**Cochrane collabora tion tool for assessing risk of bias:** *2 low risk of bias, 7 high risk of bias, 6 unclear risk of bias*	No statistical significant differences between RME appliances: -tooth borne and tooth-tissue borne I1 -tooth-tissue borne I2 and tooth borne I1 -tooth borne I1 and I3 -bondend and bande I4 No statistical significa	C: The evidence was of very low quality and was insufficient to allow the conclusion that any one intervention is better than another for any of the outcomes in this review. Q: The sample sizes were consistently small, More

| | | | | C: Alternative tratment | | | | | | differences between RME and SME, or Semi-Rapid and RME were also reported. No statistical significant differences between SME appliances: -I5 and I6 | randomised controlled trials are required to address the question of what is the best treatment for posterior crossbites. 'Correction of crossbite' needs to be the primary outcome for all studies addressing this research question. |

R: retrospective; P: Prospective; CCT: Controlled Clinical Trial; CT: Clinical Trial; RCT Randomized Clinical/Controlled Trials; RME: Rapid Maxillary Expansion; SME: Slow Maxillary Expansion; BREMA: Bonded Rapid Maxillary Expansion Appliance; QH: Quad-Helix; CBCT: Cone-Beam Computer Tomography; CT: Computed Tomography; NR: Not Reported; Mmw: Maxillary intermolar width; Mpw: Maxillary interpremolar width; Mcw: Maxillary intercanine width; mmw: mandibular intermolar width; mcw: mandibular intercanine width

~This paper also reports data about grinding and/or removable appliances that are not pertinent to our review
*2 RCTs, 2 P CCTs and 1 R CCT only assess the effects of grinding
** The main outcomes of this study are the side effects of the maxillary expansion. Data about skeletal effects are also reported
*** Two studies reported the number of extracted teeth instead of the number of subjects

Table VI. 7 Data extracted from the 12 SRs and MAs included

VI.4.2. Quality of the included SRs and MAs

The Cohen's k coefficient for the AMSTAR was 0.93, thus indicating excellent inter-examiner agreement.

The AMSTAR score ranged from a minimum of 4 to a maximum of 10 (mean score 6.8). The AMSTAR items for each paper and the total AMSTAR scores are shown in Table VI.8. None of the papers was rated as "low quality", 7 papers were rated as "medium quality" and 5 papers were rated as "high quality".

Two papers included only studies with randomization (Randomized Clinical Trials – RCTs), four papers included only Clinical Controlled Studies (CCTs), two papers included RCTs and CCTs, one paper included CCTs and non-controlled trials, one paper included RCTs and non-controlled trials, and two papers included RCTs, CCTs and non-controlled trials. The LRD scores are shown in Table VI.9.

	Schiffman & Tuncay, 2001	Petren *et al.*, 2003	Lagravere *et al.* 2005a	Lagravere *et al.* 2005b	Lagravere *et al.* 2005c	Lagravere *et al.* 2006	De Rossi *et al.* 2008	Zuccati *et al.*, 2011	Zhou *et al.* 2013	Bazargani *et al.*, 2013	Lione *et al.*, 2012	Agostino *et al.*, 2014
Was an 'a priori' design provided?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Was there duplicate study selection and data extraction?	CA	CA	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Was a comprehensive literature search performed?	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Was the status of publication (i.e. grey literature) used as an inclusion criterion?	N	N	N	N	N	N	N	N	N	N	N	Y
Was a list of studies (included and excluded) provided?	N	N	Y	Y	N	Y	Y	Y	N	N	N	Y
Were the characteristics of the included studies provided?	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y
Was the scientific quality of the included studies assessed and documented?	Y	Y	Y	N	N	Y	N	Y	Y	Y	Y	Y
Was the scientific quality of the included studies used appropriately in formulating conclusions?	CA	Y	Y	NA	NA	Y	NA	Y	Y	Y	Y	Y
Were the methods used to combine the findings of studies appropriate?	Y	Y	Y	NA	NA	Y	NA	Y	Y	Y	Y	Y
Was the likelihood of publication bias assessed?	Y	NA	NA	NA	NA	N	NA	NA	Y	NA	NA	N
Was the conflict of interest stated?	N	N	N	N	N	N	N	Y	Y	Y	N	Y
Total AMSTAR Score	**5**	**6**	**8**	**5**	**4**	**8**	**5**	**8**	**9**	**7**	**7**	**10**

Y= Yes; N= No; NA= Not Applicable; CA= Can't Answer

Table VI. 8 Quality assessment according the AMSTAR items for each SR and Total AMSTAR scores. For each Yes answer: 1 point; all the other answers: 0 point.

Authors, Year, Reference	LRD
Schiffman & Tuncay, 2001	II-III
Petren *et al.*, 2003	III*
Lagravere *et al.*, 2005a	III
Lagravere *et al.*, 2005b	III
Lagravere *et al.*, 2005c	II-III-IV
Lagravere *et al.*, 2006	II-III-IV
De Rossi *et al.*, 2008	III
Zuccati *et al.*, 2011	II
Zhou *et al.*, 2013	II-III
Bazargani *et al.*, 2013	II-IV
Lione *et al.*, 2013	III-IV
Agostino *et al.*, 2014	II

*considering the studies assessing the outcome pertinent to the current review

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

Table VI. 9 Study design of the primary studies included in each SR, as assessed according to the LRD scores.

VI.4.3.   Synthesis of the results and quality of the body evidence

a. <u>Short-term effects</u>

- *Dentoalveolar transverse effects of SME*

Maxillary intermolar width. Three studies (Petren *et al.*, 2003; Lagravere *et al.*, 2005c; Zhou *et al.*, 2013) assessed the changes in maxillary intermolar width after SME with QH appliance. The evidence ranged between *very low* and *moderate*. All the studies reported a significant increase of this width, ranging between a minimum of 3.3 mm to a maximum of 6.4 mm. One study (Lagravere *et al.*, 2005c) assessed the same outcome also with Minne-expander and Nitanium maxillary expander, and a significant increase was reported with a *low* level of evidence.

One study (Agostino *et al.*, 2014), aiming to directly compare different expansion appliances, pointed out no significant difference in maxillary intermolar expansion rate when comparing bonded Minne-expander versus banded Minne-expander; the evidence for this result was *very low*.

Maxillary intercanine width. Two studies (Petren et al., 2003; Zhou *et al.*, 2013) assessed the changes in intercanine width after SME with QH appliance. The evidence was *low* and both studies reported a significant increase ranging from a minimum of 1.3 mm to a maximum of 5.2 mm.

Mandibular intermolar width. One study (Zhou *et al.*, 2013) addressed this outcome. A *moderate* level of evidence of a significant increase of 0.49 mm was reported.

No data about maxillary intepremolar width were reported in the included SRs and MAs.

- *Dentoalveolar sagittal and vertical effects of SME*

No data about sagittal and vertical dentoalveolar effects of SME were reported in the included SRs and MAs.

- *Dentoalveolar transverse effects of RME*

Maxillary intermolar width. Three studies (Petren *et al.*, 2003; Lagravere *et al.*, 2006; Zhou *et al.*, 2013) assessed the changes in maxillary intermolar width after RME with banded Hyrax, acrylic Hyrax or Haas appliance. The evidence ranged from *very low* to *high* and a significant increase of the width was reported in all studies, ranging from a minimum of 4.09 mm to a maximum of 6.74 mm.

One study (Agostino *et al.*, 2014), aiming to directly compare different expansion appliances, pointed out no significant difference in maxillary intermolar expansion rate when comparing banded Hyrax (tooth borne) versus bonded Hyrax (tooth/tissue borne), tooth-tissue borne Haas versus tooth borne Hyrax, Hyrax tooth-borne expander versus bone-anchored expander and four-point banded Hyrax versus two-point banded Hyrax. The evidence was *very low* for all the comparisons.

Maxillary intercanine width. Three studies (Petren *et al.*, 2003; Lagravere *et al.*, 2006; Zhou *et al.*, 2013) assessed the changes in maxillary intercanine width after RME with banded Hyrax, acrylic Hyrax or Haas appliance. The evidence ranged from *very low* to *high* and a significant increase of the width was reported in all studies, ranging from a minimum of 2.70 mm to a maximum of 5.35 mm.

Maxillary interpremolar width. One study (Zhou *et al.*, 2013) assessed this outcome after RME with banded Hyrax, acrylic Hyrax or Haas appliance. A *high* evidence of a significant effect of 3.86 mm of increasing was reported.

Mandibular intermolar width. Two studies (Petren *et al.*, 2003; Zhou *et al.*,

2013) assessed the changes in mandibular intermolar width after RME with banded Hyrax, acrylic Hyrax or Haas appliance. The evidence was *low* in one SR (Petren *et al.*, 2003), which reported a non-significant effect of 0.49 mm, while *high* evidence of significant effect of 1.19 mm was reported in another paper (Zhou *et al.*, 2013).

- *Dentoalveolar sagittal and vertical effects of RME*

One study (Lagravere *et al.*, 2006) assessed sagittal and vertical changes after RME with Hyrax and Haas expander. A significant increase (0.53 mm) of the vertical distance between maxillary molar cusp and Palatal Plane (PP) and a significant increase of the Overjet (OVJ) (1.29 mm) were found, with a *very low* level of evidence.

- *Dentoalveolar transverse effects of SME vs. RME*

One study (Zhou *et al.*, 2013) reported no significant difference between SME, performed with QH or Minne-expander, and RME, performed with Hyrax expander. This outcome was assessed as concerned to maxillary intermolar width (-0.23 NS, *moderate* evidence), maxillary intercanine width (0.69 NS, *moderate* evidence), maxillary intepremolar width (1.35 NS, *low* evidence) and mandibular intermolar width (-0.68 NS, *low* evidence). Regarding maxillary intermolar width, the same result was pointed out in a more recent MA (Agostino *et al.*, 2014), when comparing treatment performed with two-bands expander, again with *very low* level of evidence.

- *Dentoalveolar sagittal and vertical effects of SME vs. RME*

No data about the direct comparison of SME vs. RME in terms of dentoalveolar sagittal and vertical effects were reported in the included SRs and MAs.

- *Skeletal transverse effects of SME*

One study (Lagravere *et al.,* 2006) assessed skeletal transverse effects after SME with QH, Minne-expander or Nitanium maxillary expander. For all the appliances an increase of the transverse dimensions was reported, with a *very low* level of evidence.

- *Skeletal sagittal and vertical effects of SME*

No data about sagittal and vertical skeletal effects of SME were reported in the included SRs and MAs.

- *Skeletal transverse effects of RME*

Three studies (Lagravere *et al.,* 2006; Bazargani *et al.,* 2013; Lione *et al.,* 2013) assessed the skeletal changes after RME, measured at different landmarks. One SR (Lagravere *et al.,* 2006) reported an increase of the distance between left and right jugale points (2.73 mm), with a *low* level of evidence. *Low* level of evidence was also found as concerned the effect of RME on the posterior region of the midpalatal suture (between 1.20 mm and 4.33 mm) and on the anterior region of the midpalatal suture (between 1.10 mm and 4.80 mm) (Bazargani *et al.,* 2013; Lione *et al.,* 2013). Finally, *low* evidence of skeletal expansion measured at canine was reported (from 1.5 mm to 4.3 mm) (Lione *et al.,* 2013).

- *Skeletal sagittal and vertical effects of RME*

<u>Upper jaw.</u> Vertical movements of the maxilla after RME with Hyrax and Haas expander were assessed in two studies (Lagravere *et al.,* 2006; Lione *et al.,* 2013). *Low* evidence of a non-significant effect was reported in one MA (Lagravere *et al.,* 2006), while *very low* evidence a small significant increase of the Sella-Nasion/Palatal Plane angle (1.7°) was reported in another SR (Lione *et al.,* 2013). *Very low* evidence of downward rotation of the maxilla was reported also in one SR (De Rossi *et al.,* 2008) assessing the effects of bonded expansion appliances; in the

same study unclear results are provided regarding skeletal sagittal movement.

<u>Lower jaw.</u> Vertical movements of the mandible after RME with Hyrax and Haas expander have been assessed in two studies (Lagravere *et al.,* 2006; Lione *et al.,* 2013). A downward rotation of the mandible has been reported in all the studies, based on different landmarks, with a level of evidence ranging between *low* and *very low*. One SR (De Rossi *et al.,* 2008) assessed both vertical and sagittal mandibular changes after RME with bonded expansion appliance, pointing out a *very low* evidence of downward ration and backward movement of the lower jaw.

<u>Other structures.</u> *Very low* evidence of small increase of several circummaxillary sutures and of the spheno-occipital synchondrosis was reported (Bazargani *et al.,* 2013).

- *Skeletal transverse, sagittal and vertical effects of SME vs. RME*

No data about the direct comparison of SME vs. RME in terms of skeletal transverse, sagittal and vertical effects were reported in the included SRs and MAs.

b. <u>Long-term effects</u>

- *Dentoalveolar transverse effects of SME*

Maxillary intermolar width. Three studies (Schiffman & Tuncay, 2001; Petren *et al.,* 2003; Zhou *et al.,* 2013) assessed the long-term changes of maxillary intermolar width after SME performed with QH. There was *low* evidence that around the 95% of initial expansion amount was maintained while wearing retention (Schiffman & Tuncay, 2001). After the retention phase, the amount of expansion maintained ranged between 3.6 mm and 5.1 mm and the evidence varied from *low* to *moderate*.

Maxillary intercanine width. Two studies (Petren *et al.*, 2003; Zhou *et al.*, 2013) assessed the long-term changes of maxillary intercanine width after SME performed with QH. A residual expansion amount post-retention was reported, ranging from a minimum of 2.2 mm to a maximum of 3.7 mm, with a *low* level of evidence.

No data about maxillary intepremolar width were reported in the included SRs and MAs.

Mandibular intermolar width. One study (Zhou *et al.*, 2013) assessed the long-term changes of mandibular intermolar width after SME performed with QH. *Moderate* evidence of a very small amount of residual expansion was reported (0.06 mm).

-   *Dentoalveolar sagittal and vertical effects of SME*

No long-term data are reported in the included SRs and MAs.

-   *Dentoalveolar transverse effects of RME*

Maxillary intermolar width. Five studies assessed this outcome (Schiffman & Tuncay, 2001; Petren *et al.*, 2003; Lagravere *et al.*, 2005a; Zuccati *et al.,* 2011; Zhou *et al.*, 2013). While wearing retention, approximately the 85% of the initial expansion amount was maintained after RME with Hyrax expander; this data was reported with a *low* level of evidence (Schiffman & Tuncay, 2001). After retention, one SR (Zuccati *et al.,* 2011) evaluated the effects of RME with QH, reporting a residual expansion between 3.4 mm and 4.6 mm, with a *low* level of evidence. Assessing the same outcome after treatment with Hyrax, Haas and acrylic splint expander, the expansion maintained ranged between 3.58 mm (Zhou *et al.*, 2013) (*moderate* evidence) and 5.4 mm (Petren *et al.*, 2003) (*very low* evidence). When only Hyrax expander was analysed (Schiffman & Tuncay, 2001), 3.56 mm of residual expansion (around the

49% of the initial expansion) were reported with *low* evidence. Considering the effect of treatment with Haas expander in association with fixed edgewise orthodontic appliances the amount of expansion maintained ranged between 4.00 mm and 4.80 mm (Lagravere *et al.*, 2005b), reported with *low* to *very low* level of evidence.

Maxillary intercanine width. Four studies (Petren *et al.*, 2003; Lagravere *et al.*, 2005b; Zuccati *et al.*, 2011; Zhou *et al.*, 2013) assessed the long-term changes of the intercanine width after RME. One study (Zuccati *et al.*, 2011) reported with *low* evidence a residual expansion between 1.40 mm and 3.20 mm. After treatment with Hyrax, Haas and acrylic splint expander the residual expansion ranged between 2.64 mm (Zhou *et al.*, 2013) (*moderate* evidence) and 3.30 mm (Petren *et al.*, 2003) (*very low* evidence). Considering the effect of treatment with Haas expander in association with fixed edgewise orthodontic appliances, the amount of expansion maintained ranged between 2.30 mm and 2.50 mm (Lagravere *et al.*, 2005b), presented with a *low* or *very low* level of evidence.

Maxillary interpremolar width. Two studies (Lagravere *et al.*, 2005b; Zhou *et al.*, 2013) evaluated interpremolar width. One MA (Zhou *et al.*, 2013) pointed out *high* evidence of a residual width of 3.52 mm after RME with Hyrax, Haas and acrylic splint expander. *Low/very low* evidence of residual expansion between 4.20 mm and 4.70 mm was found after treatment with Haas appliance combined with edgewise orthodontic appliance (Lagravere *et al.*, 2005b).

Mandibular intermolar width. Two studies (Lagravere *et al.*, 2005b; Zhou *et al.*, 2013) evaluated long-term changes of mandibular intermolar width. *Moderate* evidence of residual expansion of 2.02 mm after treatment with Hyrax, Haas and acrylic splint expander was reported in one MA (Zhou

et al., 2013). *Low/very low* evidence of residual expansion between 0.70 mm and 2.50 mm was found after treatment with Haas appliance combined with edgewise orthodontic appliance (Lagravere *et al.*, 2005b).

<u>Mandibular intercanine width.</u> One SR (Lagravere *et al.*, 2005b) assessed long-term changes of mandibular intercanine width after RME performed with Haas appliance in association with edgewise fixed appliance. *Low* or *very low* evidence of a residual expansion between 0.80 mm and 1.50 mm was found.

-   *Dentoalveolar sagittal and vertical effects of RME*

Long-term sagittal and vertical changes after RME with Haas expander combined with edgewise fixed appliance have been assessed in one SR (Lagravere *et al.*, 2005b). *Very low* evidence of a slight reduction of the OVJ (-0.6 mm) and insignificant molar extrusion were reported.

-   *Skeletal transverse, sagittal and vertical effects of SME*

No data on these outcomes were reported in the included SRs and MAs.

-   *Skeletal transverse, sagittal and vertical effects of RME*

One study (Lagravere *et al.*, 2005a) assessed long-term transverse, sagittal and vertical changes after RME, but no information on the type of appliance were reported. Considering maxillary width *low* evidence of residual expansion of 3 mm in early maturation group and insignificant values in late maturation group were reported.

Very low evidence of non significant sagittal changes at both maxilla and mandible are reported in one study (Lagravere *et al.*, 2005a), except for a significant retruded position of the A point (-1.05°).

-   *Skeletal transverse, sagittal and vertical effects of SME vs. RME*

No data on these outcomes were reported in the included SRs and MAs.

VI.5. <u>Discussion</u>

The present Systematic Review aimed to summarize the current evidence from SRs and MAs on dental and skeletal effects produced by different palatal expansion techniques. In particular, the focus of the present study concerned the quality and the main results of the SRs and MAs addressing this issue.

VI.5.1. Quality assessment of the included SRs and MAs

Since SRs and MAs can provide the highest level of evidence, these studies are generally considered the cornerstone of the evidence-based health care. Nevertheless, as with all other publications, the value of SRs depends on the quality of the methodology adopted and on the quality of the reporting. The methodological quality depends on the extent to which adequate scientific methods are adopted to minimize biases (Shea *et al.*, 2007). Since SRs can still have good quality of reporting but low methodological quality (Pieper *et al.*, 2014), it is crucial for the readers to critically appraise the results of a SR with appropriate instruments. The AMSTAR score is a valid, reliable and easy to use tool for assessing the methodological quality of SRs (Shea *et al.*, 2007; 2009). Indeed, according to an analysis of published overview of SRs, the AMSTAR is one of the most used quality assessment tool (Hartling *et al.*, 2012).

None of the included SRs and MAs demonstrated low quality, as assessed with the AMSTAR tool; the lowest score was 4 points (Lagravere *et al.*, 2005c).

The only item that presents "Yes" answer for all the included SRs and

MAs was the *item 1 ("Was an 'a priori' design provided?")*. This item refers to the existence of a registered protocol, or ethics approval, or pre-determined/a priori published research objectives. A published protocol was present only in two included MAs (Zhou *et al.*, 2013; Agostino *et al.*, 2014). However, since databases for protocol registration (i.e. PROSPERO - International Prospective Register of Systematic Review; Booth *et al.*, 2011) have been recently introduced, in our study we took into consideration the chronological limitation and attributed affirmative answer to the *item 1* whenever the research question and inclusion criteria seemed to be clearly established before the conduction of the review. Such approach should avoid the review method to be influenced by reviewers' expectations (Garg *et al.*, 2008).

The *item 2 ("Was there duplicate study selection and data extraction?")* refers to the study selection and data extraction procedure, which should be performed by at least two operators. This procedure was frequently clearly stated for the study selection, but often unclear for what concerns to data extraction (Lagravere *et al.*, 2005a; Lagravere *et al.*, 2005b; Lagravere *et al.*, 2005c; Lagravere *et al.*, 2006; De Rossi *et al.*, 2008; Zuccati *et al.*, 2011). On the contrary, in two SRs (Schiffman & Tuncay, 2001; Petren *et al.*, 2003) the authors clearly stated that two investigators performed data extraction, but this was unclear for the study selection.

Common items to loose points in the AMSTAR total score were: *item 4 ("Was the status of publication (i.e. grey literature) used as an inclusion criterion?")*, checked in 1 out of 12 papers; *item 10 ("Was the likelihood of publication bias assessed?")*, checked in 2 out of 12 papers; *item 11 ("Was the conflict of interest stated?")*, checked in 3 out of 12 papers. The *item 10* was rated as 'Not Applicable' whenever a MA was not performed. Taking in mind that

each AMSTAR item has different weight in the total score (List and Axelsson, 2010), not reporting the conflict of interest has a clear lower impact on the quality of a SR than introducing a publication bias. Publication bias is a major threat to the validity of any type of review and occurs when studies with statistically significant or clinically relevant results are more likely to be published than studies with non-significant or unfavourable results (Rothstein *et al.*, 2005). When the statistical comparison (MA) of the primary studies was not performed, detection of publication bias was considered 'Not Applicable'. Hence, including data from unpublished trials seems to be one obvious way to avoid the influence of publication bias when a MA is not possible (Higgings and Green, 2011), and must be critically improved in future SRs.

The other item that was unchecked in half of the papers was the *item 5* (*"Was a list of studies (included and excluded) provided?"*), due to the frequent absence of reporting of the excluded references.

The paper with the highest AMSTAR score was the only Cochrane SR available on palatal expansion (Agostino *et al.*, 2014). This result is in accordance with what has been already extensively proved regarding the quality of Cochrane systematic reviews, being better than that of non-Cochrane reviews published in peer-reviewd papers (Jadad *et al.*, 1998; Moher *et al.*, 2007; Fleming *et al.*, 2013). These findings underline the importance of several levels of peer-review and standardized guidelines to improve the quality of the literature.

Interestingly, the Cochrane SR was the only paper to have been updated. It is well know that SRs are more helpful if they appraise the most recent literature (Moher *et al.*, 2008), and more efforts should be put by the reviewers in keeping constant update also of non-Cochrane reviews.

To gain a wider perspective on the evidence provided by the included SRs and MAs, information regarding the design of the individual studies were collected, according the hierarchy of evidence, using the LRD scoring (Cooke, 1996; Antes, 1998). This instrument has been previously adopted in a SR of SRs (List and Axelsson, 2010) since is an easy outline of the level of evidence of the primary studies included in each SR.

The limit of this tool is the weak definition of what "controlled studies" are, as it generally included both untreated control groups and alternative treatment groups. When assessing the effects of an orthodontic/orthopedic procedure, clinicians cannot discard normal growth as a confounding factor (Lagravere *et al.*, 2005c), especially for long-term results. Therefore, the presence of an untreated control group gives more value to the results than the comparison with a different treatment modality.

In absence of the highest level of evidence, clinicians have to make decisions based on lower levels of evidence, even though they are more prone to biases (Lagravere *et al.*, 2005c). Four of the analysed SRs included also studies without control group (case-series): this choice dramatically reduces the strength of the conclusions, but still it underlines important weaknesses in the primary literature on specific topics.

RCTs are suggested as the best way to compare the efficacy of two or more interventions, but not without difficulties (Pocock, 1996). In 7 of the included SRs the literature search resulted into the inclusion of RCTs, but only 2 of these reviews were entirely focus on randomized studies. From the total number of 20 RCTs included within the SRs analysed, 5 RCTs overlapped in 3 SRs and 6 RCTs overlapped in 2 SRs.

The lack of overlap is probably due to the different inclusion criteria of the reviews, and reflects the differences in comparison and outcomes of the primary studies.

VI.5.2. Synthesis of the results and quality of the body evidence

When pooling the data from the included SRs and MAs, it was impossible to perform a MA due to the high heterogeneity of methodologies and endpoints. The first diversity between the included studies was in the *"head-to-head"* comparisons, performed with different appliances or expansion modalities (RME or SME). Even within the same type of expansion modality, variable rates of screw activation were found. The second main problem was the difference in outcomes; when the same outcome was reported (i.e. short term dentoalveolar transversal effects) difference in materials (radiographs or dental cast), landmarks and measurements were numerous. Finally, the lack of clarity of the endpoints also affected the comparability of the results. Indeed, in order to adequately contrast the results from multiple studies, the crossbite resolution with a goal of over-expansion should be reported as pre-determined inclusion criterion.

A factor that limited the external validity of the results of the included SRs and MAs was the unclearness of the initial diagnosis of crossbite and the definition of the baseline condition in terms of amount of transverse discrepancy, number of emiarches affected (unilateral or bilateral) and origin of the crossbite (skeletal or exclusively dentoalveolar); moreover, only in two SRs (Petren *et al.*, 2003; Agostino *et al.*, 2014) the authors

specified also the sagittal diagnosis among the inclusion criteria, clearly stating the exclusion of papers assessing the effect of expansion in patients with a Class III malocclusion.

The main limit of the current SR of SRs was the use of a scale of statements for the assessment of the quality of the body evidence, which has not been previously tested or validated, but proposed as modified from existent tools adopted in recent Cochrane Overviews of SRs (Ryan *et al.*, 2011; 2014; Pollock *et al.*, 2014). The proposed tool partially resembles the concepts of study limitation, publication bias, imprecision, inconsistency and indirectness required for the GRADE levels of evidence (Guyatt *et al.*, 2008), and contemporary provides a transparent, objective and reproducible assessment process. The "cut-offs" were established by reaching consensus among the authors of the review. We were unable to adopt the original GRADE level of evidence as suggested by the Cochrane Collaboration due to several factors: this approach was used only in one of the included MAs, frequently raw data of primary studies were not available to provide the "Summary of Findings" table, and this approach was considered restrictive for the current study, as it starts with an high rating of RCTs that are difficult to perform in the orthodontic field.

a. High evidences

According to the findings of the current study few outcomes presented high quality of the body evidence, and all the highest evidence were related to the dentoalveolar short-term transversal effects of RME. In particular, short-term increase of transversal width of maxillary molar, premolar and canine region and mandibular molar region are reported with high evidence after expansion with Hyrax or Haas expander (Zhou

et al., 2013).

Even if pooling data from different appliances in the same SR or MA provided better evidence regarding the expansion modality (RME) as more data were pooled, it must be taken into account that this can limit the applicability of the result since the appliances might differ in biomechanics and so provide different effects.

When comparing the transversal increase obtained at the intermolar width (4.09 mm) with that of canine (2.70 mm) or premolar regions (3.86 mm), we found that the greatest expansion was obtained at the molars with progressively reduced expansion in the anterior part of the arch. This can be explained by the fact that the appliances used for the expansion exert their force directly on the posterior teeth used for anchorage.

Even though the appliances evaluated did not exert their force directly on the inferior teeth, mandibular intermolar width was a frequent reported outcome. It has been argued that the occlusal forces delivered by maxillary molars could produce a movement of mandibular molars during expansion (Gryson, 1977). One MA (Zhou *et al.*, 2013) pointed out a short-term expansion of the lower molars of 1.19 mm after treatment with Hyrax or Haas expander, indicating a spontaneous adaptation of the occlusion.

     b. <u>Moderate evidences</u>

Moderate evidence was found regarding short term dentoalveolar effects of SME, performed with QH. An increase of 4.45 mm is reported at maxillary molar width (Zhou *et al.*, 2013), while data on the intercanine width are not supported by the same level of evidence. Differently from what pointed out for RME, the effect of the SME on the lower molars

was reported to be small and clinically not relevant (0.49 mm), thus indicating lesser spontaneous adaptation of the lower dentition, probably due to the different speed in movement.

Focusing on the amount of expansion achieved with the two expansion modalities, the findings of the current review could provide the general impression that the results obtained with RME are better than that achieved with SME. However, the direct comparison of the two techniques (SME performed with QH and Minne expander vs. RME performed with Hyrax expander) pointed out with *moderate* evidence that no statistical significant difference exists in the short term, as concerned to the expansion reached at maxillary intermolar and intercanine width (Zhou *et al.*, 2013). Therefore, the choice between the two activation modalities is still committed to clinicians' expertise.

Also long-term dental effects of RME with Haas or Hyrax expander presented *moderate* evidence: transversal increase of 3.58 mm, 3.52 and 2.64 mm was reported at intermolar, interpremolar and intercanine width, respectively (Zhou *et al.*, 2013). These measurements, as compared to those performed immediately after the expansion, showed relapse to some extent. Interestingly, Zhou and co-workers (2013) assessed also the dental changes while wearing retention, pointing out no significant differences in the retention period.

Despite the primary role of the retention in the maintenance of the expansion result, the studies assessing long-term dental effects frequently lacked of clear statement of retention protocol. It is strongly suggested for future studies to standardize and adequately report the retention techniques and the follow-up periods, in order properly assess the amount of relapse. Also the orthodontic treatment with fixed appliances

after the palatal expansion could have played a significant role in the long-term results. Without orthodontic intervention, natural dental arch width and arch perimeter loss from late adolescence to fifth/sixth decade of life might occur (Carter and Mcnamara, 1998). Therefore, a clear description of a later phase of fixed orthodontics, which is frequently neglected, should be reported in future studies.

In one MA (Schiffman & Tuncay, 2001) it was argued that, due the long-term expansion relapse (around the 40% of the initial expansion within 5 years), the final amount of maxillary transverse increase achieved in treated patients is not far from the expansion that would have occurred thanks to the normal growth. However, this conclusion is a limited inference since it is based on the results of normal growth, which might be reliable for patients without transverse discrepancies (Björk and Skieller, 1977), but it is not for subjects presenting PXB. Therefore, the comparison between long-term stability of expansion therapy and maxillary growth should be performed with untreated subjects with the same baseline conditions.

    c. <u>Low evidences</u>

Skeletal effects of RME have been extensively investigated and this technique seemed to be an effective procedure that always produced immediate transverse skeletal changes on the maxilla by opening the midpalatal suture, regardless of the type of palatal expander. However, the differences in landmarks and measurements adopted, prevented to pool data from numerous studies, therefore the best evidence found was *low* evidence.

Interestingly, the results of two included SRs (Lagravere *et al.*, 2005b; Lione *et al.*, 2013) suggested that RME treatment was able to induce

significantly more favourable skeletal changes in the transverse plane when it was initiated before the pubertal growth peak. Therefore, the role of treatment timing might be a crucial point to achieve better short-term skeletal modification after palatal expansion, and so it has to be taken into account for future clinical studies on this outcome.

The mean expansion changes in the midpalatal suture after RME in the posterior region ranged from 1.20 mm to 4.33 mm, while in the anterior region ranged between 1.10 mm to 4.80 mm. These results suggested that there was no consistent evidence on whether the midpalatal sutural opening was parallel or triangular (Bazargani *et al.*, 2013; Lione *et al.*, 2013).

Long-term dental effects of SME are also reported with a *low* level of evidence, due to the small number of primary studies and small sample size of the papers assessing this outcome. Partial maintenance of the expansion was always reported, even though a variable amount of relapse was present (Schiffman & Tuncay, 2001; Petren *et al.*, 2003; Zhou *et al.*, 2013). Limitations to this consideration are similar to those pointed out for the RME: lack of retention protocol, lack of definition of the follow-up period and absence of description of fixed orthodontic treatment.

Regarding the comparison between the two expansion techniques (RME vs. SME), non-significant data were reported with *low* evidence regarding short-term maxillary interpremolar width and mandibular intermolar width, and long-term results. The only significant value found was in the long-term effect of maxillary intermolar width, which resulted to be larger after SME than after RME, with a very limited clinical relevance (0.75 mm) (Zhou *et al.*, 2013).

d. <u>Very low evidence</u>

Short-term skeletal effects of the SME, assessed with QH, Minne-expander and Nitanium maxillary expander, were reported in only one SR (Lagravere *et al.*, 2005c), with a *very low* level of evidence. Unspecific increase of maxillary transversal dimension was pointed out, with no details on the amount of the increase or on the landmarks considered. This lack of evidence was mainly due to the small sample size of the primary study and to the lack of quality assessment performed in the review.

Regarding the short-term skeletal effects of RME, the circummaxillary sutures, in particular the zygomatico-maxillary and fronto-maxillary sutures, and the spheno-occipital syncondrosis appeared to be affected by maxillary expansion, but the changes in these structures were overall small and supported by *very low* evidence (Bazargani *et al.*, 2013). This effect should be further investigated since it can be particularly relevant in skeletal Class III patients, where the opening of the circummaxillary sutures might promote the advancement of the maxilla (Bazargani *et al.*, 2013).

Concerning immediate vertical and sagittal skeletal changes after RME treatments, *very low* evidence of more backward and downward position of mandible and of downward position of the maxilla were reported. It has been argued that the movement of the maxilla during the expansion can cause premature dental contacts, which can be responsible for the mandible rotation. This can be further explained by the fact that bonded expanders are reported to cause less downward and backward displacement of the mandible than banded appliances, probably due to the absence of occlusal interference, but these alterations are not

completely eliminated (De Rossi *et al.*, 2008). Even if changes in vertical dimension were too small to be considered clinically relevant, in a patient with a vertical growth pattern bonded appliances can be an option choice.

Long-term transversal skeletal effects of RME were studied only in one SR (Lagravere *et al.*, 2005a), and the reported evidences were generally *very low* due to the scarce quality of primary literature. In fact, the authors stated that numerous studies were excluded because of the absence of a control group that is essential to compare the normal growth changes with the changes of the treated group. Interestingly, as for the dentoalveolar effects, one SR reported higher stability in transverse skeletal maxillary increase in the group treated before pubertal peak than in the group treated after pubertal peak (Lagravere *et al.*, 2005a), but this finding needs to be further investigated.

The overall vertical and sagittal skeletal changes after RME were reported to be insignificant or not clinically relevant, but all supported by *very low* evidence.

Only one MA (Agostino *et al.,* 2014) assessed the differences in the effectiveness of several expansion appliances, reporting *very low* evidence of non-significant results. In our opinion, this seems a crucial point in the evidence-based decision-making that should be further investigated in order to provide the best treatment option.

VI.6.   Conclusions

• The quality of the SRs and MAs on dentoalveolar and skeletal effects of palatal expansion ranged between medium and high (from 4 to 10).

• Two studies (one MA and one SR) included only RCTs (LRD II).

• Palatal expansion is an effective procedure for the resolution of dental posterior crossbite since it always produces an increase of transverse dimension in the short-term. The evidence supporting these outcomes is overall lower for SME than for RME due to the lesser quality and lower number of primary studies.

• An increase of skeletal transverse width of the maxilla is always present, but reported to be lower than dentoalveolar expansion. Due to the *low* quality of the evidence for RME and *very low* quality for SME, more research is needed to confirm this finding.

• Dentoalveolar long-term expansion data are supported by *moderate* evidence for RME and *low* evidence for SME. However, these results are still unclear due to the variability in retention protocol, follow-up period and absence of untreated control groups.

• Skeletal long-term expansion data are poor and supported by *very low* evidence.

• There is *moderate* evidence supporting the non-significant difference between RME and SME regarding dentoalveolar effects.

• There is *very low* evidence supporting the non-significant difference between expansion appliances within the same expansion modality.

• Important gaps in the evidence provided by SRs and MAs on palatal expansion are: short- and long-term comparison of the skeletal

effects of RME vs. SME; long-term assessment of the skeletal effects of SME and evaluation of vertical and sagittal dental and skeletal short- and long-term effects of SME.

References

•   Agostino P, Ugolini A, Signori A, Silvestrini-Biavati A, Harrison JE, Riley P. Orthodontic treatment for posterior crossbites. Cochrane Database Syst Rev. 2014; 8: CD000979.

•   Allen D, Rebellato J, Sheats R, Ceron AM. Skeletal and dental contributions to posterior crossbites. Angle Orthod. 2003; 73: 515-24.

•   Antes G. Evidence-based medicine. Internist (Berl). 1998; 39: 899–908.

•   Bazargani F, Feldmann I, Bondemark L. Three-dimensional analysis of effects of rapid maxillary expansion on facial sutures and bones. Angle Orthod. 2013; 83:1074-82.

•   Bell RA, Kiebach TJ. Posterior crossbites in children: Developmental-based diagnosis and implications to normative growth patterns. Seminars in Orthodontics. 2014, 20: 77-113.

•   Björk A, Skieller V. Growth of the maxilla in three dimensions as revealed radiographically by the implant method. Br J Orthod. 1977; 4: 53-64.

•   Björk A. Sutural Growth of the Upper Face Studied by the Implant Method. Rep Congr Eur Orthod Soc 1964; 40: 49-65.

•   Booth A, Clarke M, Ghersi D, Moher D, Petticrew M, Stewart L. An international registry of systematic review protocols. The Lancet. 2011; 377: 108-109.

•   Carter GA, McNamara JA Jr. Longitudinal dental arch changes in adults. Am J Orthod Dentofacial Orthop. 1998; 114: 88-99.

•   Castañer-Peiro A. Interceptive orthodontics: the need for early diagnosis and treatment of posterior crossbites. Med Oral Patol Oral Cir Bucal. 2006; 11: E210-4.

•   Cooke IE. Finding the evidence. In: Cooke IE, Sackett DL. Ed. Ballière's. Clinical obstetrics and gynecology. International Practice and Research. 1996; 10: 561-67.

•   D'Antò V, Bucci R, Franchi L, Rongo R, Michelotti A, Martina R. Class II functional orthopaedic treatment: a systematic review of systematic review. Journal of Oral Rehabilitation, *accepted for publication.*

•   De Rossi M, Salviti De Sá Rocha R.A., Duarte Gavião M.B. Effects of bonded rapid maxillary expansion appliance (BREMA) in vertical and

sagittal dimensions: a systematic review. Braz J Oral Sci. 2008; 7: 1571-74.

• Fleming PS, Seehra J, Polychronopoulou A, Fedorowicz Z, Pandis N. A PRISMA assessment of the reporting quality of systematic reviews in orthodontics. The Angle Orthodontist. 2013; 83: 158-63.

• Garg AX, Hackam D, Tonelli M. Systematic review and meta-analysis: when one study is just not enough. Clinical Journal of the American Society of Nephrology. 2008; 3: 253-60.

• Gryson JA. Changes in mandibular interdental distance concurrent with rapid maxillary expansion. Angle Orthod. 1977; 47: 186-92.

• Guyatt GH, Oxman AD, Vist G, Kunz R, Falck-Ytter Y, for the GRADE Working Group. Rating quality of evidence and strength of recommendations GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008; 336: 924-6.

• Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. PLoS One, 2012; 7: e49667.

• Hesse K, Artun J, Joondeph DR, Kennedy DB. Changes in condylar position and occlusion associated with maxillary expansion for correction of functional unilateral posterior crossbite. Am J Orthod Dentofacial Orthop. 1997; 111: 410-18.

• Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

• Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, Moher D. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. The Journal of American Medical Association. 1998; 280: 278-80.

• Kurol J, Bergland L. Longitudinal study and cost-benefit analysis of the effect of early treatment of posterior crossbites in the primary dentition. Eur J Orthod. 1992; 14 173-79.

• Lagravère MO, Heo G, Major PW, Flores-Mir C. Meta-analysis of immediate changes with rapid maxillary expansion treatment. J Am Dent Assoc. 2006; 137: 44-53.

- Lagravere MO, Major PW, Flores-Mir C. Long-term skeletal changes with rapid maxillary expansion: a systematic review. Angle Orthod. 2005; 75: 1046–52.

- Lagravère MO, Major PW, Flores-Mir C. MO, Major PW, Flores-Mir C. Long-term dental arch changes after rapid maxillary expansion treatment: a systematic review. Angle Orthod. 2005; 75: 155-61.

- Lagravère MO, Major PW, Flores-Mir C. Skeletal and dental changes with fixed slow maxillary expansion treatment: a systematic review. J Am Dent Assoc. 2005; 136:194-9.

- Lam PH, Sadowsky C, Omerza F. Mandibular asymmetry and condylar position in children with unilateral posterior crossbite. Am J Orthod Dentofacial Orthop. 1999; 115: 569–75.

- Lindner A., Modeer T. Relation between sucking habits and dental characteristics in preschool children with unilateral crossbite. Scand J Dent Res. 1989; 97: 278–283.

- Lione R, Franchi L, Cozza P. Does rapid maxillary expansion induce adverse effects in growing subjects? Angle Orthod. 2013; 83: 172-82.

- List T, Axelsson S. Management of TMD: evidence from systematic reviews and meta-analyses. Journal of Oral Rehabilitation. 2010; 37: 430-51.

- Lux CJ, Ducker B, Pritsch M, Komposch G, Niekusch U.Occlusal status and prevalence of occlusal malocclusion traits among 9-year old schoolchildren. Eur J Orthod. 2009; 31: 294–99.

- Martina R, Cioffi I, Farella M, Leone P, Manzo P, Matarese G, Portelli M, Nucera R, Cordasco G. Transverse changes determined by rapid and slow maxillary expansion--a low-dose CT-based randomized controlled trial. Orthod Craniofac Res. 2012; 15: 159-68.

- McLain JB, Proffit WR. Oral health status in the United States—prevalence of malocclusion J Dent Educ. 1985; 49: 386–96.

- Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009; 6: e1000097.

- Moher D, Soeken K, Sampson M, Campbell K, Ben Perot L, Berman B. Assessing the quality of reports of systematic reviews in pediatric complementary and alternative medicine. BMC Pediatr 2002; 2: 3.

- Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. PLoS

Medicine 2007; 4: e78.

- Moher D, Tsertsvadze A, Tricco AC, Eccles M, Grimshaw J, Sampson M, Barrowman N. When and how to update systematic reviews. Cochrane Database Syst Rev. 2008 23: MR000023.

- Mulrow CD. Rationale for systematic reviews. BMJ 1994; 309: 597-599.

- Myers DR, Barenie JT, Bell RA, Williamson EH. Condylar position in children with functional posterior crossbites before and after correction. Pediatr Dent. 1980; 2: 190–94.

- Papadopoulos MA, Gkiaouris I. A critical evaluation of meta-analyses in orthodontics. Am J Orthod Dentofacial Orthop. 2007; 131: 589-99.

- Petrén S, Bondemark L, Söderfeldt B. A systematic review concerning early orthodontic treatment of unilateral posterior crossbite. Angle Orthod. 2003; 73: 588-96.

- Pieper D, Buechter RB, Li L, Prediger B, Eikermann M. Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. J Clin Epidemiol. 2014 30. doi: http://dx.doi.org/10.1016/j.jclinepi.2014.12.009.

- Pirtiniemi P, Kanotomaa T, Lahtela P. Relationship between craniofacial and condyle path asymmetry in unilateral crossbite patients. Eur J Orthod. 1990; 12: 408–13.

- Pocock SJ Clinical Trials – A Practical Approach. 1996 Chichester: John Wiley & Sons Ltd

- Pollock A, Farmer SE, Brady MC, Langhorne P, Mead GE, Mehrholz J, van Wijck F. Interventions for improving upper limb function after stroke. Cochrane Database Syst Rev. 2014; 11: CD010820.

- Rothstein HR, Sutton AJ, Borenstein M. Publication bias in meta-analysis. 2005. Wiley Eds.

- Ryan R, Santesso N, Hill S, Lowe D, Kaufman C, Grimshaw J. Consumer-oriented interventions for evidence-based prescribing and medicines use: an overview of systematic reviews. Cochrane Database of Systematic Reviews. 2011; 11: CD007768.

- Ryan R, Santesso N, Lowe D, Hill S, Grimshaw J, Prictor M, Kaufman C, Cowie G, Taylor M. Interventions to improve safe and effective medicines use by consumers: an overview of systematic reviews. Cochrane Database of Systematic Reviews. 2014; 29: CD007768.

- Schiffman PH, Tuncay OC. Maxillary expansion: a meta analysis. Clin

Orthod Res. 2001; 4: 86-96.

- Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, Ramsay T, Bai A, Shukla VK, Grimshaw JM. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One, 2007; 2: e1350.

- Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Medical Research Methodology. 2007; 15: 10.

- Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, Henry DA, Boers M. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. Journal of Clinical Epidemiology. 2009; 62: 1013-20.

- Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Medical Research Methodology. 2011; 11: 15

- Tscill P, Bacon W, Sonko S. Malocclusion in the deciduous dentition of Caucasian children. Eur J Orthod. 1997; 19: 361–67.

- Zhou Y, Long H, Ye N, Xue J, Yang X, Liao L, Lai W. The effectiveness of non- surgical maxillary expansion: a meta-analysis. Eur J Orthod. 2014; 36: 233-42.

- Zuccati G, Casci S, Doldo T, Clauser C. Expansion of maxillary arches with crossbite: a systematic review of RCTs in the last 12 years. Eur J Orthod. 2011; 35: 29-37.

Chapter VII.

General conclusions

In conclusion, the quality of the SRs and MAs in orthodontics is variable. Even though this kind of studies are considered a "gold standard" to access the best pre-filtered evidence, clinicians and researchers should be aware of the existent quality assessment tools and acquire further skills to address the evidence of the studies, before applying the reported results into their clinical practice. Furthermore, according to our findings also SRs of RCTs demonstrated considerable flaws.

Moreover, it was interesting to notice that Cochrane reviews always showed better methodological quality and more updated results than reviews published in peer-reviewed journals.

Further efforts should be put by reviews authors to follow the proposed guidelines for review reporting (PRISMA) and to adequately assess the quality of the primary studies included. These implementations might easily allow the readers to appraise the results reported in the SR and to evaluate whether these results are applicable in their clinical practice. As concerned to the latter issue, there is still a need of a tool to properly assess the quality of primary studies in orthodontics. Frequently self-produced not validated checklists were used, due to the lack of an instrument that could appropriately encompass the keys of interest of an orthodontic research study.

As concerned to clinical recommendations, the only evaluated outcome reporting high quality of the evidence was the short-term transversal dentoalveolar expansion achieved with RME technique (Haas or Hyrax) (from 4.09 mm at maxillary intermolar width to 3.86 mm at maxillary intercanine width).

Also for SME a significant effect of maxillary molar width increase was found in the short-term (4.45 mm), but with moderate evidence. However, when directly compared, moderate evidence of no differences between the two expansion modalities was found.

Long-term results are still uncertain due to the lack of reporting confounding variables (retention protocol, orthodontic fixed multibracket phase and natural growth). Even though relapse to some extent was present, up to moderate evidence of significant increase of transverse dimensions was pointed out when comparing RME (Hyrax or Haas) with alternative treatments or untreated controls. Similar results were found also for SME (Quad-Helix), but with low quality of the evidence.

Short-term skeletal transverse increase was reported with both RME and SME; however, the quality of the evidence was low for RME and very low for SME.

Dentoalveolar and skeletal effects on vertical and sagittal planes were generally reported to be small, not likely to be clinically relevant and supported by very low evidence.

Finally, no differences between expansion appliances within activation modality were found; however, the evidence supporting this finding was very low.

Regarding the effects of the functional orthopaedic treatment of Class II malocclusion, a substantial reduction of the OVJ can be expected with all functional appliance, and in particular with Twin-Block, as there was some evidence to support these findings. Similar results were pointed out also with Herbst appliances, but with insufficient evidence.

Some evidence of maxillary sagittal growth control was reported with headgear and Twin Block, but these findings were generally small (approximately -1° of SNA angle). Maxillary growth restraint seemed to be not achieved with Herbst appliances, but further research is needed to confirm this finding.

Overall, mandibular length increase was achieved with all the appliances evaluated in the included SRs and MAs, but the clinical relevance is still questionable due to the variably of the findings and the lack of long-term results.

Chapter VIII.

Annexes

VIII.1.   <u>Index of abbreviations</u>

- AMSTAR: A Measurement Tool to Assess Systematic Reviews

- CCT: Controlled Clinical Trial

- EBD: Evidence-Based Dentistry

- EBM: Evidence-Based Medicine

- GRADE: Grading of Recommendations Assessment, Development, and Evaluation

- LILACS: Latin American and Carribean Health Science

- LRD: Level of Research Design

- MA: Meta-analysis

- MOOSE: Meta-analysis of Observational Studies in Epidemiology

- NR: Narrative Review

- OoSR: Overview of Systematic Review

- OQAQ: Overview Quality Assessment Questionnaire

- OVJ: Overjet

- PICO: Patient/Problem, Intervention, Comparison, Outcome

- PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

- PXB: Posterior Cross Bite

- QH: Quad-Helix

- QUORUM: Quality of Reporting of Meta-Analysis

- RCT: Randomized Clinical Trial

- RME: Rapid Maxillary Expansion

- SciELO: Scientific Electronic Library Online

- SME: Slow Maxillary Expansion

- SR: Systematic Review

- WOK: Web Of Knowledge

VIII.2. <u>Figure Legends</u>

- Figure I. 1 The Evidence-Based Medicine Triad

- Figure I. 2 The *5-steps* model for EBM

- Figure II. 1 The evidence pyramid

- Figure II. 2 Structured questions for systematic reviews as proposed by Khan *et al.*, 2003. (Copyright © 2003, The Royal Society of Medicine)

- Figure II. 3 Graphical representation of a meta-analysis. Example of a forest plot of four studies assising the effect of a given intervention as compared to a control group.

- Figure III. 1 Schematic view of GRADE's process for developing recommendations as suggested by Guyatt *et al.*, 2011

- Figure IV. 1 Number of published SRs between 1990 and 2015. This estimate was calculated by searching on Medline (Entrez PubMed) the keyword "Review" [Publication Type] and applying the filter "Systematic Reviews".

- Figure IV. 2 Decision algorithm for interpreting discordant reviews. (Jadad *et al.*, 1997)

- Figure V. 1 PRISMA Flow Diagram

- Figure VI. 1 PRISMA Flow Diagram

VIII.3.   Table Legends

- Table II.1 Main differences between Systematic Reviews and Narrative Reviews

- Table V. 1 Search strategy for each database and relative results

- Table V. 2 Search strategy and keyword combinations used for Medline (via PubMED)

- Table V. 3 Interpretation of the LRD scores. The scores are based on the type of studies included

- Table V. 4 Scale of Statements adopted to rate the evidence of the outcomes retrieved from each SR

- Table V. 5 References excluded after the full-text reading and reason for the exclusion

- Table V. 6 Data extracted from the 14 Systematic reviews and Meta-Analyses included

- Table V. 7 Quality assessment according the AMSTAR items for each SR and Total AMSTAR scores. For each Yes answer: 1 point; all the other answers: 0 point.

- Table V. 8 Study design of the primary studies included in each SR, as assessed according to the LRD scores.

- Table VI. 1 Search strategy for each database and relative results

- Table VI. 2 PubMed search strategy

- Table VI. 3 Interpretation of the LRD scores. The scores are based on the type of studies included in the SR

- Table VI. 4 Objective criteria for rating the evidence of the reported outcomes

- Table VI. 5 Levels of evidence according to the number of downgrades

- Table VI. 6 References excluded after the full-text reading and reason for the exclusion

- Table VI. 7 Data extracted from the 12 SRs and MAs included

- Table VI. 8 Quality assessment according the AMSTAR items for each SR and Total AMSTAR scores. For each Yes answer: 1 point; all the other answers: 0 point

- Table VI. 9 Study design of the primary studies included in each SR, as assessed according to the LRD scores.

VIII.4.   <u>Publications</u>

- D'Antò V., **Bucci R.**, Franchi L., Rongo R., Michelotti A., Martina R. Class II functional orthopaedic treatment: a systematic review of systematic reviews. JOR. 2015 [accepted for publication]

- Rongo R., Valletta R., **Bucci R.**, Bonetti GA., Michelotti A., D'Antò V. Does clinical experience affect the reproducibility of cervical vertebrae maturation method? Angle Orthod. 2015 Jan 26. [Epub ahead of print]

- **Bucci R**, Rongo R, Zito E, Galeotti A, Valletta R, D'Antò V. Cross-cultural adaptation and validation of the Italian Psychosocial Impact of Dental Aesthetics Questionnaire (PIDAQ). Qual Life Res. 2015 24: 747-52.

- Paduano S, Cioffi I, Rongo R, Cupo A, **Bucci R**, Valletta R. Orthodontic management of congenitally missing maxillary lateral incisors: a case report. Case Rep Dent. 2014, 2014: 731074.

"The future belongs to those who believe in the beauty of their dreams"

E. Roosevelt