UNIVERSITY OF IOANNINA
SCHOOL OF HEALTH SCIENCES
FACULTY OF MEDICINE

DEPARTMENT OF HYGIENE AND EPIDEMIOLOGY

Evidence from Systematic Reviews in Dental Research

Nikolaos Pandis

Dissertation Prepared for the Degree of
DOCTOR OF PHILOSOPHY

IOANNINA 2015

UNIVERSITY OF IOANNINA
SCHOOL OF HEALTH SCIENCES
FACULTY OF MEDICINE

DEPARTMENT OF HYGIENE AND EPIDEMIOLOGY

Evidence from Systematic Reviews in Dental Research

Nikolaos Pandis

Dissertation Prepared for the Degree of
DOCTOR OF PHILOSOPHY

IOANNINA 2015

Η έγκριση της διδακτορικής διατριβής από την Ιατρική Σχολή του Πανεπιστημίου Ιωαννίνων δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα Ν. 5343/32, άρθρο 202, παράγραφος 2 (νομική κατοχύρωση του Ιατρικού τμήματος).

Ημερομηνία αίτησης του κ. Πανδή Νικολάου: 28-9-2011

Ημερομηνία ορισμού Τριμελούς Συμβουλευτικής Επιτροπής: 725[α]/6-12-2011

Μέλη Τριμελούς Συμβουλευτικής Επιτροπής:

Επιβλέπων

Σαλαντή Γεωργία, Επίκουρη Καθηγήτρια Επιδημιολογίας

Μέλη

Ντζάνη Ευαγγελία, Επίκουρη Καθηγήτρια Υγιεινής με ιδιαίτερη έμφαση στην Επιδημιολογία

Worthington Helen, Professor of Evidence Based Care School of Dentistry, University of Manchester

Ημερομηνία ορισμού θέματος: 3-2-2012

«Επιστημονική τεκμηρίωση από συστηματικές ανασκοπήσεις στην Οδοντιατρική έρευνα»

ΟΡΙΣΜΟΣ ΕΠΤΑΜΕΛΟΥΣ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ : 780[α]/2-7-2015

1. Worthington Helen, Professor of Evidence Based Care School of Dentistry, University of Manchester
2. Δημολιάτης Ιωάννης Αναπληρωτής Καθηγητής Υγιεινής με έμφαση στην Ιατρική Εκπαίδευση του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
3. Ευαγγέλου Ευάγγελος Επίκουρος Καθηγητής Υγιεινής με έμφαση στην Κλινική και Μοριακή Επιδημιολογία του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
4. Ντζάνη Ευαγγελία, Επίκουρη Καθηγήτρια Υγιεινής με ιδιαίτερη έμφαση στην Επιδημιολογία του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
5. Σαλαντή Γεωργία, Επίκουρη Καθηγήτρια Επιδημιολογίας του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
6. Σπυρίδωνος Παναγιώτα Λέκτορας Ιατρικής Πληροφορικής του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
7. Μαυρίδης Δημήτριος Λέκτορα του Παιδαγωγικού Τμήματος του Παν/μίου Ιωαννίνων

Έγκριση Διδακτορικής Διατριβής με βαθμό **«ΑΡΙΣΤΑ»** στις **4-9-2015**

ΠΡΟΕΔΡΟΣ ΤΟΥ ΤΜΗΜΑΤΟΣ ΙΑΤΡΙΚΗΣ
Πασχόπουλος Μηνάς
Καθηγητής Μαιευτικής-Γυναικολογίας

Η Γραμματέας του Τμήματος
ΚΑΠΙΤΟΠΟΥΛΟΥ ΜΑΡΙΑ

Acknowledgements

I would like to thank the members of my committee Dr. Helen Worthington, Dr. Evangelia Tzani, and especially Dr. Georgia Salanti, my supervisor, who in a difficult time for her guided me and helped me on all aspect of my PhD process.
I would like to thank Loukia Spineli and Dimitris Mavridis for their invaluable help with the analysis and reporting of the first section on Network Meta-Analysis and Dr. Peter Miles and Dr. Emily Ong for providing individual patient data for the NMA section.

I would like to thank Dr. Kerry Dwan for her guidance on the section on selective reporting and Jo Weldon from the COHG for providing the missing protocols for this section. I would like to thank Ms Nancy Santesso from the GRADE group for helpful comments during the preparation of part 3 on GRADE.
A special thanks to my friend and colleague Dr. Padhraig S. Fleming for his immense contributions to this work.
I feel fortunate that I was given the opportunity to work with such a highly esteemed group of scientists and learn from them.

I would like to thank Professor Christos Katsaros, University of Bern, Switzerland, for his personal support and encouragement.
Special thanks to my wife Brandi and daughter Emily for their support, patience and understanding during the PhD process which took time from them.

This research project has been funded by the Papavramidès Foundation, Bern, Switzerland.

Abstract

Scientific evidence balanced with clinical expertise and patient preferences and values are key components to evidence based health care. The existing body of evidence is important into guiding healthcare recommendations and clinical practice guidelines. It is imperative within healthcare fields to assess the quality of the existing body of evidence for not only to drive recommendations but also to determine areas which require further research.

The present thesis aims to provide empirical evidence related to the quality of evidence in the oral health field. The thesis is divided in three thematic sections starting from the more specific to the more general. In the first part the first attempt on network meta-analysis in the field of orthodontics is reported. Application of acceptable alternative interventions compared with standard therapy begs the question of which intervention is optimal. NMA allows comparing interventions not tested on head to head comparisons using a common comparator and facilitates ranking in terms of effectiveness of those interventions. NMA was applied in order to compare dental alignment efficiency of various orthodontic systems.

In the second part all systematic reviews and the corresponding protocols published by the Cochrane Oral Health Group were retrieved and were compared in order to identify discrepancies in the reported outcomes between the two reports. An assessment of outcome reporting bias was also made by looking at whether changes such as new outcome introduction and outcome upgrades were associated with statistically significant findings.

In the third and final section systematic reviews from 14 high impact general and specialty dental journals, and from Cochrane OHG, 2008-2013, were selected in order to assess the quality of the evidence on a variety of interventions. The assessment was done using the GRADE approach which rates the evidence on 5 domains: risk of bias, inconsistency, indirectness, imprecision, and publication bias.

****

Table of Contents

List of Tables

3.3 Frequencies and percentages of included review characteristics for Cochrane and non-Cochrane reviews at review level

3.4 Frequencies and percentages of characteristics of selected meta-analyses for detailed assessment for 41 Cochrane and 50 not Cochrane reviews at meta-analysis level.

3.5. Frequencies and percentages of overall and per domain GRADE rating for Cochrane and non-Cochrane meta-analyses.

List of Figures

3.4.2 Number of included non-Cochrane and Cochrane reviews by year of publications

3.4.3. Heterogeneity (I-square %) for non-Cochrane and Cochrane included meta-analyses

Chapter 1

Network Meta-Analysis is Orthodontics: Initial orthodontic alignment effectiveness with self-ligating and conventional appliances

1.1 Introduction

Fundamental to evidence-based orthodontic practice is the use of high quality research during the decision making process. Ideally, best evidence should be used and appraised to decide whether a theoretical approach might be applicable to a particular setting or individual patient. The scientific evidence is categorised into different levels based on the scientific priority it commands, ranging from clinical opinion at the lower level to systematic reviews of high quality clinical trials at the highest.

Systematic reviews of interventions aim to collect and accumulate high quality evidence on the effects of an intervention in a systematic, transparent and unbiased manner. This information may be combined qualitatively, or quantitatively in a meta-analysis. Quantitative analysis may produce a more precise estimate on the effectiveness and safety of a therapy. Additionally, systematic reviews may reconcile misunderstandings and existing controversies regarding therapies and expose unanswered questions, which may be addressed in subsequent trials [1].

Traditional meta-analysis focuses on comparison of two therapies or a comparison of one therapy with a control. When several interventions are being tested, performing multiple pair-wise comparisons restricts the evidence to just one part of the whole picture. A relatively recent development in meta-analysis allows the incorporation of all eligible evidence in a unified framework permitting simultaneous synthesis of all available data, under certain assumptions. These methods allow direct and indirect comparisons of diverse interventions in trials using the same outcome to be combined, increasing the amount of usable information to calculate pooled estimates. This type of meta-analysis has been termed Multiple Treatments Meta-analysis, Mixed Treatment Comparisons or Network Meta-Analysis (NMA) [2–7].

Network meta-analysis offers important advantages over conventional meta-analysis allowing interventions that have not been compared in any trial to be compared and permitting ranking of interventions in order of effectiveness. This technique can also improve the precision of the effect estimates, by reducing the width of the confidence intervals compared with direct evidence alone [3–8].

Systematic reviews using NMA are scarce in the dental literature and to our knowledge quantitative syntheses of this nature have not been reported in orthodontics [9–11]. NMA has primarily been undertaken within a Bayesian framework and has long been inaccessible to non-statisticians. However, the development of software routines and tutorials in STATA™ are likely to popularize the method [12–15]. The main purpose of this study is to to assess the effectiveness of initial orthodontic alignment using different bracket systems under the NMA framework will be combined.

1.2 Methods

Setting up the clinical research question and identifying the studies

A potentially important factor influencing the effectiveness of initial orthodontic alignment is the bracket system and specifically the mode of archwire ligation. The two main bracket systems include those in which wires are engaged using either elastics or ligatures (conventional brackets; CBs) and self-ligating brackets (SLBs), incorporating a self-engaging mechanism, either active or passive. The treatments we intend to compare are applicable, in principle, to any individual potentially in need of orthodontic therapy and hence the treatments can be considered as competing. Several trials have been published comparing self-ligating systems, such as In-Ovation-R, Damon, and Smart-Clip with conventional systems. The comparisons are typically limited to two groups such as Damon vs. conventional or Smart-Clip vs. conventional and so forth [16]. In this review we will use NMA to compare all interventions simultaneously allowing them to be ranked in order of effectiveness.

The following selection criteria were applied:

- Study design: Randomized clinical trials (RCTs), controlled clinical trials (CCTs), and split-mouth designs were to be included.
- Participants: Patients with full-arch, fixed and bonded orthodontic appliance(s).
- Interventions/comparators: Various types of self-ligating and conventional brackets.
- Outcome measures: Millimeters of crowding alleviated during the initial alignment stage per unit of time.

1.3 Search Strategy for Identification of Studies

The following electronic databases were searched: MEDLINE (1966 to December 2012, Appendix 1), and EMBASE (1980 to December 2012), the Cochrane Oral Health Group's Trials Register (December 2012), the Cochrane Central Register of Controlled Trials (CENTRAL, The Cochrane Library Issue 1, 2013) without language restrictions. Unpublished literature was searched electronically using ClinicalTrials.gov (www.clinicaltrials.gov), the National Research Register (www.controlled-trials.com), and Pro-Quest Dissertation Abstracts and Thesis database (www.proquest.com) using the term 'orthodontic'. Conference proceedings and abstracts were also accessed where possible. Authors were to be contacted to identify unpublished or ongoing clinical trials and to clarify results if necessary. Reference lists of the included studies were screened for relevant research.

1.4 Assessment of comparability and risk of bias in the included studies and data extraction

Assessment of research for inclusion in the review, assessment of risk of bias and extraction of data was performed and disagreements resolved by discussion. Seven criteria were considered to grade the risk of bias of individual studies including: random sequence generation; allocation concealment; blinding of participants and

personnel; blinding of assessors; incomplete outcome data; selective reporting of outcomes; and other potential sources of bias [1].

A data extraction form was developed to record study design; observation period; participants; interventions; outcomes; and outcome data of interest, including use of extractions. The clinical and methodological heterogeneity of included studies was to be judged by assessing the treatment protocol, particularly participants and setting, intervention details, such as materials used, follow-up, timing of data collection and measurement techniques.

The primary outcome assessed was the amount of tooth movement in millimetres per month. This outcome was calculated in each primary study by dividing the amount of crowding resolved by the number of days of follow-up during initial alignment and then scaled to provide a monthly rate.

For split mouth designs, the mean difference from the paired observations was calculated along with the corresponding standard deviation. The relative treatment effect was calculated as the difference in the monthly rate of alignment between the compared interventions. In instances where the standard deviation of the mean difference was not reported in the studies, the following formula was used to approximate it

$$\sqrt{sd_{before}^2 + sd_{after}^2 - 2 \cdot r \cdot sd_{before} \cdot sd_{after}}$$

where sd_{before} and sd_{after} are the standard deviations before and at the end of the alignment stage, respectively, and r is the correlation coefficient between the before and after measurements. The correlation coefficient was calculated from available individual patient data where necessary [17–20]. In order to reduce potential bias due to lack of randomization in CCTs, estimates adjusted for the impact of potential effect modifiers (such as bracket type, age, gender, and angle classification as covariates) were calculated from individual patient data supplied by the trial authors [17–20] .

1.5 Evaluation of the assumptions underlying NMA: when you should and when you should not consider NMA

The combination of direct and indirect evidence is considered reasonable when the selected treatments are exchangeable and any study participant is eligible to receive any of the competing interventions. The key assumption for applying NMA is that there are no important differences in the distribution of possible effect modifiers across trials investigating various comparisons, other than the particular treatments being assessed [5, 21–24]. The synthesis of studies directly comparing two treatments makes sense only when the studies are sufficiently similar in terms of important clinical and methodological characteristics (effect modifiers). Only when this is the case can we assume that the intervention effects are transitive, i.e. that we can estimate a comparison via a common comparator (Figure 1.5.1).

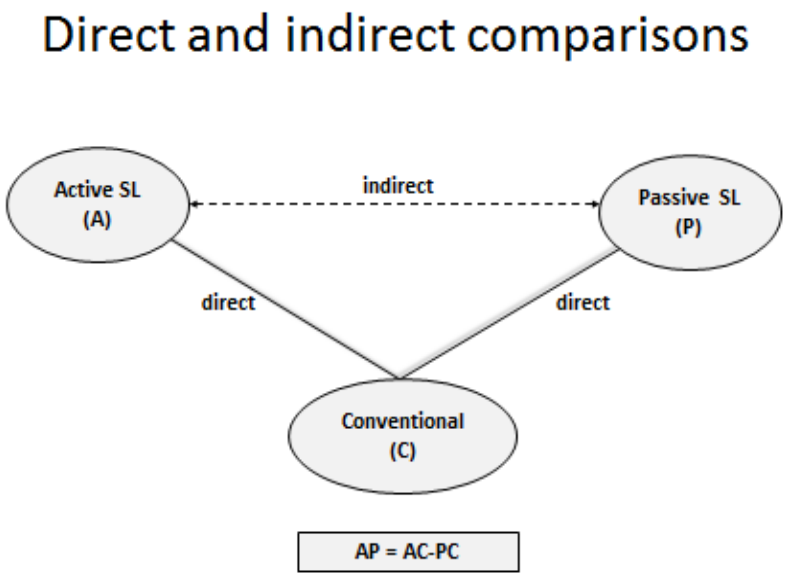

Figure 1.5.1

For example, a valid indirect comparison for active versus passive appliances requires that the studies considering active versus conventional appliances and passive versus conventional appliances studies are similar with respect to the distribution of effect modifiers such as severity of crowding at baseline, age, sample size, and study quality. Transitivity can be viewed as an extension of clinical and methodological homogeneity to comparisons across groups of studies comparing different treatments. The plausibility of the transitivity assumption requires clinical

judgment in order to decide whether differences in the distributions of the effect modifiers across studies are large enough to make network meta-analysis invalid.

One of the possible effect modifiers is the degree of bias in the included studies; for instance, blinding might be less feasible with some treatments compared to others and this may influence the results. To evaluate this assumption, we present the location and dispersions characteristics (e.g. medians and ranges) of effect modifiers across treatment comparisons. However, evaluation of the assumption can be complicated by failure to report effect modifiers in primary studies or due to limited numbers of studies per comparison to permit a reasonable judgment.

1.6 Estimation of direct, indirect and mixed evidence for the competing interventions

Using indirect evidence, it is possible to estimate the relative effectiveness of two treatments, despite the lack of studies directly comparing specific interventions. For example, we can compare Damon (D) with Smart-Clip (S) using conventional (C) as a common comparator (Figure 1.6.2).

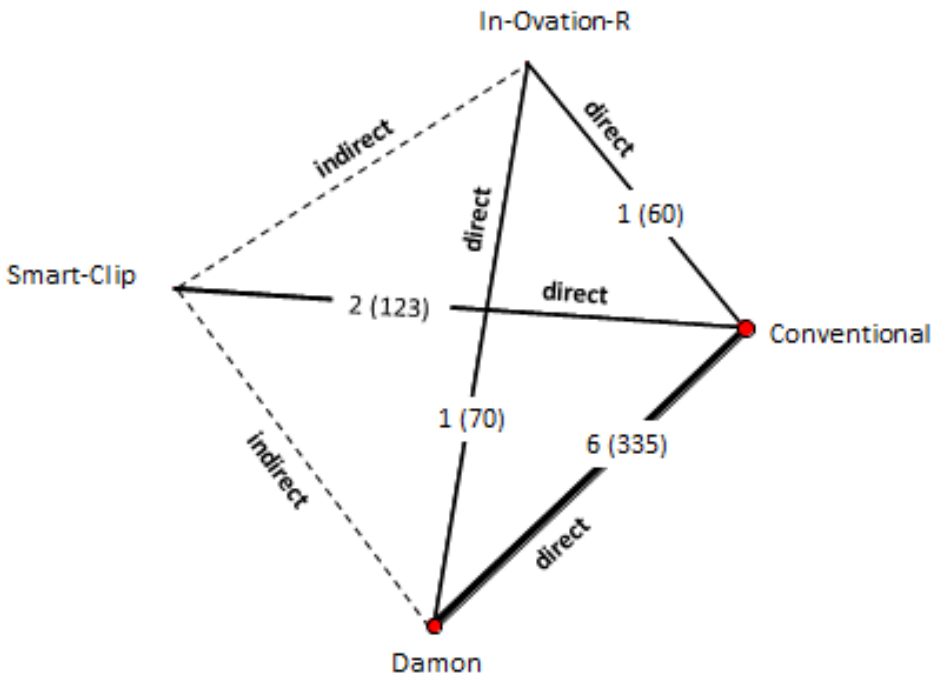

Figure 1.6.2

The meta-analytical effect estimates of Damon versus conventional and Smart-Clip versus conventional trials are denoted as $\hat{\mu}_{DC}^{D}$ and $\hat{\mu}_{SC}^{D}$, respectively, where the superscript refers to the direct estimate and the subscript refers to the pair-wise comparison (Damon vs. conventional and Smart-Clip vs. conventional). The variances of the direct meta-analytic estimates are denoted with $\hat{v}_{CS}^{D}$ and $\hat{v}_{CD}^{D}$. An *indirect estimate* $\hat{\mu}_{DS}^{I}$ for the comparison Damon versus Smart-Clip can be obtained by subtracting the meta-analytical estimate conventional versus Damon from the meta-analytical estimate conventional versus Smart-Clip:

$$\hat{\mu}_{DS}^{I} = \hat{\mu}_{CS}^{D} - \hat{\mu}_{CD}^{D} \quad (1)$$

The variance of the indirect estimate is simply the sum of the variances of the two direct estimates, as long as the direct estimates are independent:

$$\hat{v}_{DS}^{I} = \hat{v}_{CS}^{D} + \hat{v}_{CD}^{D}$$

Then a 95% confidence interval for the indirect estimate can be obtained as :

$$\hat{\mu}_{DS}^{I} \pm 1.96\sqrt{\hat{v}_{DS}^{I}}\,.$$

If both direct and indirect evidence are available for a comparison (as is the case of Damon versus In-Ovation-R), we can combine these two sources of evidence by taking the weighted average of the direct $\hat{\mu}_{DI}^{D}$ and the indirect estimates $\hat{\mu}_{DI}^{I}(= \hat{\mu}_{CI}^{D} - \hat{\mu}_{CD}^{D})$. This combination is known as *mixed evidence* and it will be denoted as $\hat{\mu}_{DI}^{M}$ [28]. Using the inverse-variance method, the mixed relative treatment effect for Damon versus In-Ovation-R (I) is calculated as follows

$$\hat{\mu}_{DI}^{M} = \frac{\hat{\mu}_{DI}^{D} \cdot \widehat{w}_{DI}^{D} + \hat{\mu}_{DI}^{I} \cdot \widehat{w}_{DI}^{I}}{\widehat{w}_{DI}^{D} + \widehat{w}_{DI}^{I}}$$

where $\widehat{w}_{DI}^{D} = 1/\hat{v}_{DI}^{D}$ and $\widehat{w}_{DI}^{I} = 1/\hat{v}_{DI}^{I}$ are the weights for the direct and indirect evidence, respectively.

The variance of $\hat{\mu}_{DI}^{M}$ is calculated by

$$\hat{v}_{DI}^{M} = \frac{1}{\widehat{w}_{DI}^{D} + \widehat{w}_{DI}^{I}}$$

and a 95% confidence interval for the mixed estimate is obtained as $\hat{\mu}_{DI}^{M} \pm 1.96\sqrt{\hat{v}_{DI}^{M}}$.

The method described above to estimate direct and indirect effects was first presented by Bucher et al [25].

1.7 Statistical evaluation of the consistency between direct and indirect evidence

Evaluation of the transitivity assumption in practice might be challenging due to limited amount of data or poor reporting of the possible effect modifiers in the studies. Even if transitivity is likely to hold, the collected data might not be comparable because of chance or unobserved differences across studies.

The statistical manifestation of transitivity is called consistency (or coherence or congruence) referring to the situation in which the direct and indirect effects are in agreement. Consistency can be evaluated only in a closed loop of comparable treatments where studies can provide both direct and indirect evidence. The terms transitivity and consistency can be confusing, but a distinction analogous to clinical or methodological heterogeneity and statistical heterogeneity in standard meta-analysis can be helpful. Statistical heterogeneity refers to the degree of disagreement between study-specific treatment effects beyond what chance can explain. In an analogous way, inconsistency refers to the degree of disagreement between sources (including direct and indirect evidence) beyond what chance and heterogeneity can explain [2, 5]. More detailed summary of the merits and caveats associated with the NMA approach are described in details elsewhere [3–5] . Where transitivity is questionable and/or there is statistical evidence of inconsistency, investigators should consider exploring the disagreement using subgroup analysis or meta-regression.

Damon, In-Ovation-R and conventional form a triangle often called *closed loop of evidence* (Figure 1.6.2). In this triangle direct and indirect evidence are available for all three possible pair-wise comparisons (DI, DC and IC). Indirect evidence is estimated using equation (1), known as the *consistency equation*. This equation assumes that both direct (left part of the equation) and indirect (right part of the

equation) evidence can contribute to inference on the comparison of interest. For example, the consistency equation $\hat{\mu}_{DI}^{I} = \hat{\mu}_{CI}^{D} - \hat{\mu}_{CD}^{D}$ states that both the direct $(\hat{\mu}_{CD}^{D})$ and indirect $(\hat{\mu}_{CD}^{I})$ evidence will provide on average the same effect estimate for the Damon versus In-Ovation-R comparison. In a triangle of treatments the consistency equation can be expressed in three different, yet equivalent, ways, such that

$\hat{\mu}_{DI}^{I} = \hat{\mu}_{CI}^{D} - \hat{\mu}_{CD}^{D}$;

$\hat{\mu}_{CI}^{I} = \hat{\mu}_{CD}^{D} + \hat{\mu}_{DI}^{D}$;

$\hat{\mu}_{CD}^{I} = \hat{\mu}_{CI}^{D} - \hat{\mu}_{DI}^{D}$

Hence, the reliability of the indirect and mixed evidence depends on the validity of the consistency [24]. Statistical evaluation of the assumption of consistency is presented in the following section.

In practice, it is plausible that the indirect evidence of a comparison may differ considerably from the direct evidence. The discrepancy between the direct and indirect estimates provides a measure of inconsistency known as the *inconsistency factor (IF)* [22, 26, 27]. The size of this discrepancy reflects potential inconsistency in the loop of interest (say Damon, In-Ovation-R, Conventional) and is estimated as

$$\widehat{IF}_{DIC} = |\hat{\mu}_{DI}^{D} - \hat{\mu}_{DI}^{I}|$$

Note that the direction of the inconsistency factor is not important for the evaluation of consistency, and hence, only absolute differences are taken. In a closed loop of evidence only one inconsistency factor can be estimated, because the same inconsistency factor will be obtained whichever edge of the triangle is chosen [22, 26, 27].

The variance of the inconsistency factor is

$$\widehat{var}(\widehat{IF}_{DIC}) = \widehat{var}(\hat{\mu}_{DI}^{D}) + \widehat{var}(\hat{\mu}_{DI}^{I})$$

Then, a 95% confidence interval for the inconsistency factor can be obtained as

$\widehat{IF}_{DIC} \pm 1.96\sqrt{\widehat{var}(\widehat{IF}_{DIC})}$ (8)

A $z - test$ can be used to test the null hypothesis of consistency $(IF_{DIC} = 0)$

$z = \frac{\widehat{IF}_{DIC}}{\sqrt{\widehat{var}(\widehat{IF}_{DIC})}}$.

The method described above is known also as 'the Bucher method [25].

Calculating inconsistency in a closed loop can be extended to all loops formed in a network allowing the inconsistency to be plotted together with 95% confidence intervals[15, 22, 28]. The results from this approach should be interpreted with caution as the tests applied in the loops are not independent and therefore have low power. Due to issues relating to power, investigators should not rely solely on the results of the inconsistency test to infer the plausibility of consistency but should also evaluate consistency clinically and epidemiologically, as appropriate [29].

1.8 Full Network meta-analysis model

Estimation of indirect and mixed relative treatment effects can be carried out as described above, when the network has few closed loops. For complicated networks, an NMA model is needed to synthesize the data efficiently. The simplest NMA model is an extension of meta-regression. This approach treats the different treatment comparisons as covariates in a meta-regression model[30].

The aim of NMA is to estimate all possible relative treatment effects between pairs of the T treatments, for instance, μ_{kc} the relative effect of treatment k versus treatment t. In our example, $T = 4$ and k, t are any of the treatments: Damon, conventional, Smart-Clip and In-Ovation-R. There is no need to estimate all μ_{kt}, but only a few of them, termed *basic parameters*. At the start of every NMA a set of basic parameters representing the summary treatment effects of $T - 1$ independent treatment comparisons are chosen. An easy way to define the basic parameters is to choose one of the T interventions as the reference (denoted with A) and each μ_{At} would represent the comparison of treatment t vs. A.

Each study i $(i = 1, \dots, S)$ reports at least one treatment effect y_i corresponding to the comparison of treatments c and k. This is denoted as y_i and its variance v_i^2; in our example, these are the observed mean differences and their variances. Each study is also assumed to estimate the underlying 'true' treatment effect with normally distributed random errors and random effects. Heterogeneity is represented by the variance τ^2 in the distribution of the random effects (assuming a

common heterogeneity for all comparisons). The NMA model can be then written as a meta-regression model without an intercept and covariate denoting the basic parameters or a function thereof. The estimated regression coefficients are interpreted as the NMA summary effects for comparisons representing the basic parameters.

The meta-regression model will estimate all $\mu_{At}s$; the network summary effects of all other functional comparisons can then be derived using the consistency equations

$$\hat{\mu}_{kc} = \hat{\mu}_{Ac} - \hat{\mu}_{Ak}$$

Any set of $T - 1$ basic parameters can be chosen with the only constraint that all the competing treatments in the network should be included in at least one basic contrast. In our example, conventional can be selected as the reference treatment so that the basic parameters are μ_{CD} for the conventional versus Damon, μ_{CI} for the conventional versus In-Ovation-R and μ_{CS} for the conventional versus Smart-Clip.

For studies comparing more than two treatments, known as *multi-arm studies*, application of the network meta-analysis model becomes complex for two reasons. Firstly, within a multi-arm study, effect estimates are correlated as they include a common treatment. Secondly, the random effects are also correlated. Therefore, the model must be modified to account for these correlations [6, 31]. In our dataset there is no multi-arm study.

For each treatment, the probability to be the first best, the second best and so on, can be calculated. These probabilities are known as *rank probabilities* and can be estimated using resampling techniques. Graphical presentation of the ranking probabilities can be generated for each treatment, with the rank probabilities plotted against all possible ranks in order to obtain a *rankogram*. Alternative guidance on graphical presentation of rank probabilities of treatments has been described [15, 32].

To provide an overall rank score for the effectiveness of each treatment, cumulative rank probabilities can be used. These are plotted against the ranks resulting in a step function and the surface under this curve is known as the *SUCRA (Surface Under the Cumulative Ranking)* value. A treatment obtaining higher rank probabilities for the

first ranks will get a high SUCRA value, while a treatment obtaining higher rank probabilities for the last ranks will get a low SUCRA value.The STATA command "*sucra"* was used to extract the rank probabilities and SUCRA values [15].

1.9 Exploring heterogeneity and inconsistency using subgroup analysis and network meta-regression

The aim of subgroup analysis or network meta-regression is to explore possible sources of heterogeneity and improve the likelihood of transitivity by accounting for characteristics that may vary across the comparisons [7]. These characteristics may relate to the design, conduct and the quality of the trials. Aggregated patient-level data, such as age, gender and extractions can also be considered, but caution is required in the interpretation of results because of the possibility of ecological bias [33]. In our example subgroup analysis will be considered in order to investigate the role of the following variables: use of extractions in the treatment plan, methodological design, and year of publication. To implement the NMA model as a meta-regression, the STATA command *metareg* was used. Network meta-regression has the same limitations as pairwise meta-regression. Confounding, lack of power when only a few trials are studied, high rate of false positive results and aggregation (or *ecological*) bias are the most common pitfalls of meta-regression [22]

1.10 Results

The PRISMA flow chart of the included studies and the assessment of the risk of bias of the included randomized clinical trials (RCTs) and non-randomized controlled clinical trials (CCTs) are given in Figures 1.10.3 and 1.10.4.

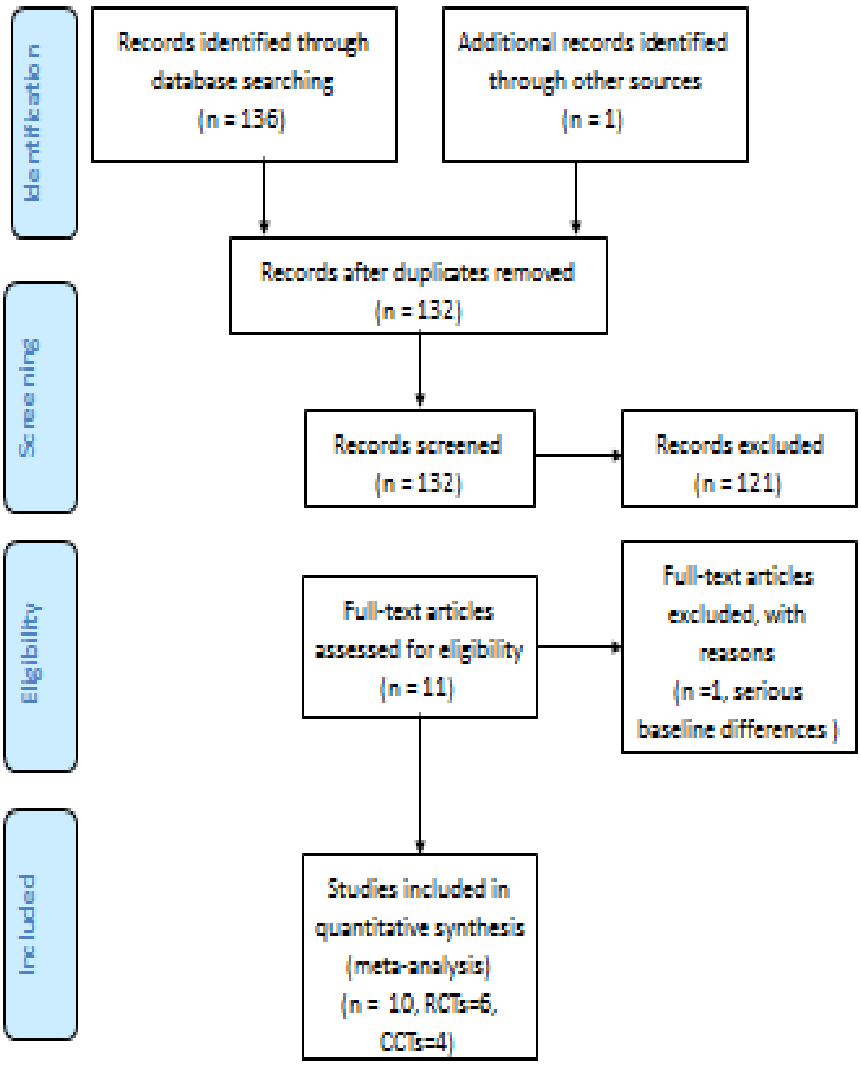
Identification
Records identified through database searching (n = 136)
Additional records identified through other sources (n = 1)
Records after duplicates removed (n = 132)
Screening
Records screened (n = 132)
Records excluded (n = 121)
Eligibility
Full-text articles assessed for eligibility (n = 11)
Full-text articles excluded, with reasons (n =1, serious baseline differences)
Included
Studies included in quantitative synthesis (meta-analysis) (n = 10, RCTs=6, CCTs=4)

Figure 1.10.3

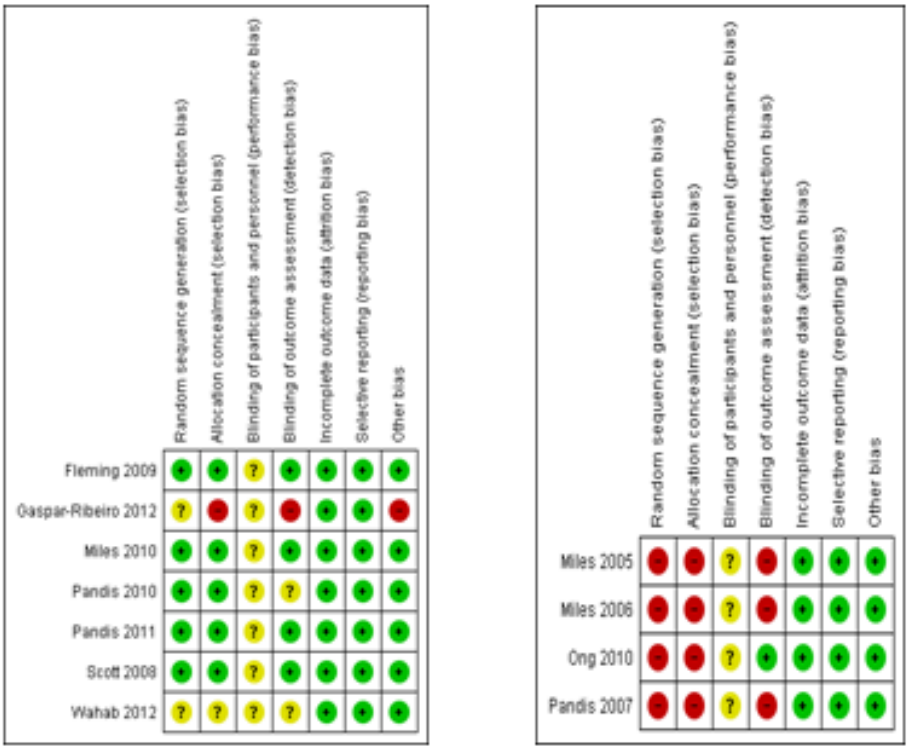

Figure 1.10.4

A total of 11 trials were identified with one trial omitted in view of significant between-group baseline differences [17–20, 34–40]. All trials involved comparisons of two bracket systems. In total, 588 participants were assigned to one of these four treatment modalities. In terms of study characteristics, six of the trials were randomized clinical trials, sample size in each group ranged from 14 to 35 with a mean of 27 patients per group. The trials were published between 2005 and 2012 with the majority published in 2010. Most of the studies investigated mandibular alignment in the anterior (inter-canine) segment. Participants ranged from 10 to 18 years in most trials, although one study had an age range of 14 to 30 years; 50% of the studies involved extraction-based treatment. After study selection and data extraction, a network of four dental bracket systems was generated (Figure 2). The size of the nodes corresponds to the number of trials relating to each particular bracket system. The larger the number of trials concerning a particular bracket system, the larger the node relating to this group. The (directly) comparable treatments are linked with a line. If there is a dotted line between two nodes, it means that there are no studies (i.e., no direct evidence) comparing the two bracket

systems. For example, there is direct evidence for all self-ligating systems versus conventional but no direct evidence for Smart-Clip versus Damon and In-Ovation-R (Figure 1.6.2). The distributions of the effect modifiers across the observed comparison are outlined in Table 1.10.1. The observed differences between the comparisons are small, although there are few studies on which to base judgement of transitivity in terms of study randomisation, extractions, age and gender of the participants and year of trial (Table 1.10.1).

Binary modifiers	Levels	Studies/Total			
		Damon vs. Conventional	In-Ovation-R vs. Conventional	Smart-Clip vs. Conventional	Damon vs. In-Ovation-R
Study type	RCT	3/6	1/1	1/2	1/1
	CCT	3/6	0/1	1/2	0/1
Tooth extraction	Yes	4/6	0/1	1/2	0/1
	No	2/6	1/1	1/2	1/1
Gender	Male	117/335	22/60	48/123	29/70
	Female	218/335	38/60	75/123	41/70
Continuous modifiers					
Age	Mean (Std.)	15.87 (2.68)	13.50 (-)	16.69 (0.58)	13.80 (-)
	Min. – max.	13.30 – 20.70	-	16.28 – 17.10	-
Year of publication	Mean (Std.)	2009 (2.37)	2010 (-)	2007 (2.83)	2010 (-)
	Min. – max.	2006 – 2012	-	2005 – 2009	-

Note: RCT: randomized-controlled trials; CCT: controlled-clinical trials; Std.: standard deviation; Min: minimum; Max: maximum

Table 1.10.1

The direct, indirect and mixed relative treatment effects from the network of the four bracket systems estimated using the formulae detailed in the methods above are presented in Table 1.10.2. The confidence interval for the mixed estimates is in general narrower than the confidence interval for the direct or indirect estimates (Table 1.10.2).

Comparisons		**Direct evidence**	**Indirect evidence**	**Mixed evidence**
Conventional versus	In-Ovation-R	-0.12 (*0.39*) [-0.69, 0.45]	0.18 (*0.37*) [-0.54, 0.90]	0.04 (*0.26*) [-0.48, 0.56]
	Damon	0.12 (*0.14*) [-0.16, 0.40]	-0.18 (*0.52*) [-1.19, 0.83]	0.10 (*0.13*) [-0.16, 0.36]
	Smart-Clip	0.16 (*0.20*) [-0.20, 0.52]	-	-
In-Ovation-R versus	Damon	-0.06 (*0.34*) [-0.51, 0.39]	0.24 (*0.41*) [-0.57, 1.05]	0.06 (*0.26*) [-0.45, 0.57]
	Smart-Clip	-	0.28 (*0.44*) [-0.58, 1.14]	-
Damon versus	Smart-Clip	-	0.04 (*0.24*) [-0.44, 0.52]	-

Note: Direct evidence has been extracted from head-to-head comparisons. Indirect evidence has been obtained using the consistency equation (1). Mixed evidence is the result of the weighting average of the direct and indirect evidence. In parenthesis we present the standard error of the estimate and in square brackets the 95% confidence interval.

Table 1.10.2

For the comparison Damon versus In-Ovation-R by combining direct and indirect evidence, the variance is reduced by 36%. The results of the conventional meta-analysis for the observed pairwise comparisons are illustrated in Figure 1.10.5.

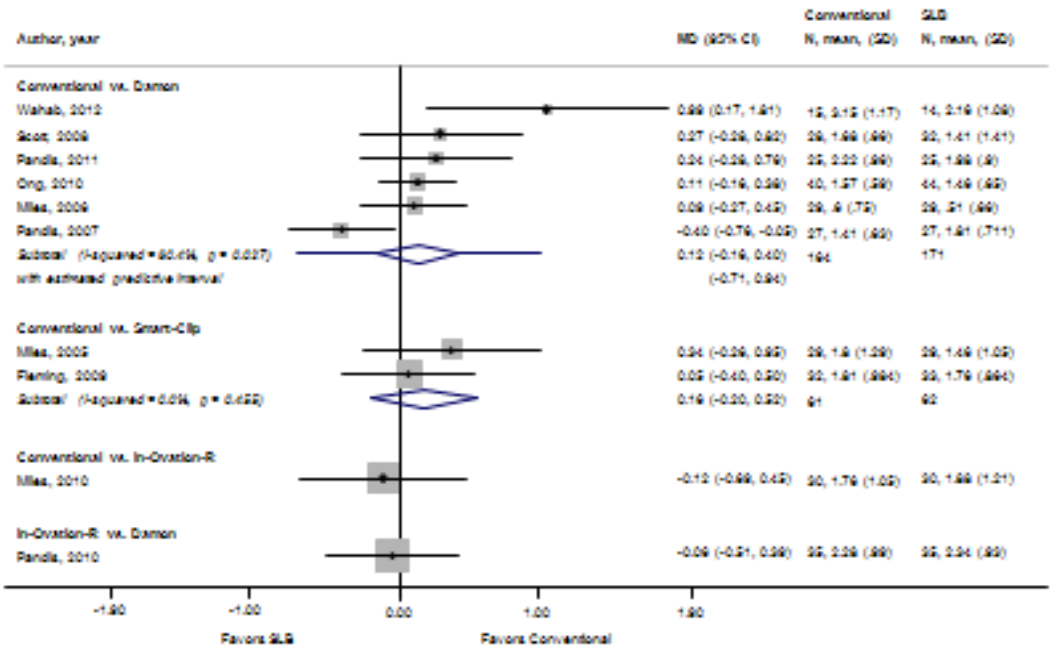

Figure 1.10.5

In the network shown in Figure 2 there is only one closed loop of evidence; Damon versus In-Ovation-R versus Conventional. To assess the inconsistency in this triangle, the comparison of Damon versus In-Ovation-R was randomly chosen from Table 1.10.1. The inconsistency factor is calculated as $IF_{DIC} = |-0.06 - 0.24| = 0.30$ with 95% confidence interval $(-0.74,\ 1.34)$. The confidence interval includes zero, and hence, there is no indication of statistically significant inconsistency between direct and indirect estimate. The interpretation of these results should be viewed with caution due to the small number of studies; the confidence intervals for IF_{DIC} are wide and include potentially large inconsistency values.

The estimated mean differences for the basic parameters are: for conventional versus Damon $\mu_{CD} = 0.08$ with $v_{CD} = 0.13$, for conventional versus I-Ovation-R $\mu_{CI} = 0.03$ with $v_{CI} = 0.26$ and for conventional versus Smart-Clip $\mu_{CS} = 0.17$ with $v_{CS} = 0.25$. These three basic contrasts can be combined to obtain all other estimates; for example, to obtain the In-Ovation-R versus Damon estimate formula (1) can be used as $\mu_{ID} = \mu_{CD} - \mu_{CI} = 0.08 - 0.03 = 0.05$. The results for all comparisons are given in Table 1.10.3.

Conventional	0.03 (-0.54, 0.48)	0.08 (-0.33, 0.17)	0.17 (-0.66, 0.32)
-0.12 (-0.69, 0.45)	In-Ovation-R	0.05 (-0.52, 0.62)	0.14 (-0.57, 0.85)
0.12 (-0.16, 0.40)	-0.06 (-0.51, 0.39)	Damon	0.09 (-0.82, 1.00)
0.16 (-0.20, 0.52)	-	-	Smart-Clip

Note: Network meta-analysis mean differences (mm/month) in tooth movement are presented above the diagonal, while direct meta-analysis results are presented below the diagonal. Interventions are ordered according to their ranking. Comparisons between systems are indicated by the column-defining bracket versus the rows-defining bracket system. Mean differences above 0 favours the bracket system in the column. To obtain mean differences for comparisons in the opposite direction, negative values should be converted into positive values, and vice versa. Heterogeneity variance (τ^2) from network meta-analysis was estimated equal to 0.05

Table 1.10.3

The effectiveness of the alignments suggests a rank according to the studied outcome. Conventional brackets were ranked as the most effective system in terms of alignment, followed by In-Ovation-R, Damon and Smart-Clip. As shown in the rankograms (Figure 1.10.6), conventional brackets and In-Ovation-R are more likely to be among the first two best options, since they have higher rank probabilities in the first two ranks compared to Damon and Smart-Clip.

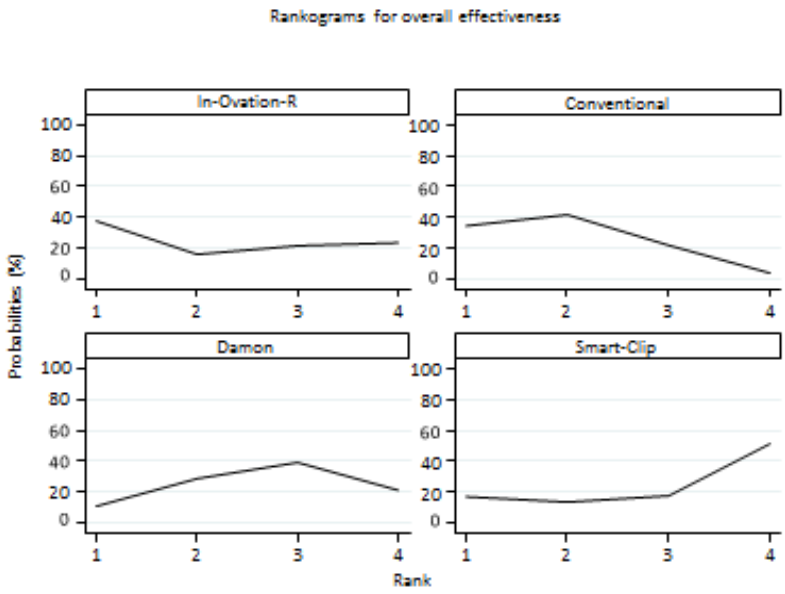

Figure 1.10.6

On the contrary, Damon and Smart-Clip are more likely to be the least good options, since they have higher rank probabilities within the last two ranks. Figure 1.10.7 illustrates the SUCRA value of each alignment in a bar plot.

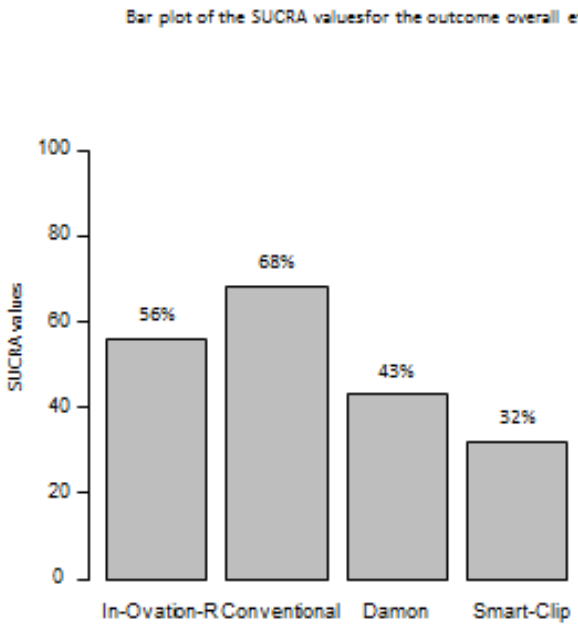

Figure 1.10.7

Conventional brackets appear to lead to the most effective alignment with a SUCRA value of 68%. These are followed by In-Ovation-R, Damon and Smart-Clip with SUCRA values 56%, 43% and 32%, respectively.

In our review, potential confounders may include use of extractions in the treatment plan, methodological design, and the year of the trial. Methodological design (RCT versus CCT) has been shown to have a possible impact on the results [41]. Year of trial may be a confounder due to improvements in orthodontic materials utilized, trial conduct, outcome assessments, and novelty bias. Due to the limited number of studies, it is difficult to evaluate whether the distribution of these confounders is comparable across the comparisons. As an example of the subgroup analysis, we present the relative effectiveness of the comparisons studied in the RCTs. The mean relative effectiveness of the conventional alignment studied in RCTs tend to be larger although the large uncertainty does not allow to draw firm conclusions; the relative effectiveness of conventional compared to In-Ovation-R is 0.17 (95% CI: -0.70, 1.04)

from NMA using only RCTs, while it is 0.03 (95% CI: -0.54, 0.48) irrespective of the methodological design. The same is observed for the relative effectiveness of conventional compared to Damon (NMA using RCTs: 0.33, 95% CI: -0.25, 1.00, NMA irrespective the methodological design: 0.08, 95% CI: -0.33, 0.17). However, the mean relative effectiveness of the Smart-Clip compared to conventional seems to be lower in the RCTs (NMA using RCTs: 0.05, 95% CI: -1.07, 1.17, NMA irrespective the methodological design: 0.17, 95% CI: -0.66, 0.32).

1.11 Clinical interpretation of the findings

The results from NMA and mixed treatment effects (Tables 1.10.2 and 1.10.3) display the mean difference in mm per month of alignment achieved for the different bracket comparisons. The NMA results indicate that the conventional appliances perform better in terms of alignment efficiency compared to all other systems with a greater mean improvement of 0.03, 0.08 and 0.17 mm/month with conventional compared to In-Ovation-R, Damon and Smart-Clip, respectively. The estimated differences are not statistically significant since the associated 95% confidence intervals include the value zero and more importantly the estimates are of little clinical importance. If we assume that an average duration for initial alignment is around 4 months the results suggest that conventional appliance will be more efficient on average anywhere from 0.12 to 0.68 mm over the 4-month period compared to the other three brackets. The expected 4 month differences were calculated by multiplying by 4 (number of months) the minimum and maximum NMA estimates for the comparisons of conventional against the other systems. The results should be interpreted with caution as the number of studies is small and the associated confidence intervals are relatively wide indicating imprecision of the estimated treatment effects. The comparisons among In-Ovation-R, Damon and Smart-Clip yielded differences of limited clinical relevance.

1. 12 Discussion

Network meta-analyses offer important advantages over conventional meta-analysis increasing the precision of the estimated effect sizes, allowing interventions that have not been compared in any trial to be compared, and permitting the creation of a hierarchical rank of interventions. NMA should, however, be used with caution with the underlying assumptions of the analysis considered carefully. In particular, the network should be consistent, with direct and indirect evidence on the same comparisons in agreement. Joint analysis of treatments can be misleading if the network is substantially inconsistent. Inconsistency may be attributed to an uneven distribution of effect modifiers across groups of trials comparing different treatments. Therefore, the distribution of clinical and methodological variables suspected to be potential sources of either heterogeneity or inconsistency in each comparison or specific group of trials should be investigated.

This is the first reported usage of multiple treatment meta-analysis in orthodontic research. Further application of this technique in orthodontics would be of great value, particularly as a range of mechanics and treatment alternatives may be deployed to address overall malocclusions e.g. increased overjet, or to effect specific tooth movements e.g. space closure. Orthodontic treatment also rarely involves binary decisions with a range of options possible in most situations [42–44]. Individual preferences continue to have an integral role in treatment planning decisions, with little uniformity in respect of a range of approaches; consequently, NMA is likely to have important application in the future of evidence-based orthodontics.

1.13 Conclusions

The results indicate that the conventional appliances perform better in terms of alignment efficiency compared to all other systems with a greater mean improvement of 0.03, 0.08 and 0.17 mm/month with conventional compared to In-Ovation-R, Damon and Smart-Clip, respectively. The estimated differences are not statistically significant.

The results should be interpreted with caution as the number of studies is small and the results are imprecise the estimated differences were small and of limited clinical relevance.

Chapter 1 is based on the publication: Pandis N, Fleming PS, Spineli LM, Salanti G. Initial orthodontic alignment effectiveness with self-ligating and conventional appliances: A network meta-analysis in practice. Am J Orthod Dentofacial Orthop. 2014 Apr;145(4 Suppl):S152-63. With permission for tables and figures [Elsevier License #**:** 3710771097383]

Chapter 2

Discrepancies in outcome reporting exist between protocols and published oral health Cochrane systematic reviews.

2. 1. Introduction

Systematic reviews (SRs) for interventions provide the basis for evaluating the body of evidence and by extension are uniquely influential in leading to healthcare recommendations [1, 45]. However, inappropriate practices such as selective inclusion of trials, selective reporting of outcomes and results-driven reporting can bias estimates of treatment effects culminating in compromised patient care, healthcare decisions and service configuration [46, 47] . Selective outcome reporting involving preferential reporting of specific data or outcomes within a study is a recognized problem and has been investigated in respect of randomized controlled trials (RCTs), in particular [48–52].

While research on selective reporting has focused both on RCTs and non-randomized studies [53, 54] similar issues may arise within SRs [55]. At the systematic review level, selective reporting may develop for a variety of reasons, for example, due to the use of multiple measurement scales, outcomes or time points [56, 57] and selective inclusion of specific outcomes. Arbitrary inclusion of outcomes based on *post hoc* results-driven decisions may bias the conclusions from subsequent syntheses [56, 58].

A recent systematic review [58] has highlighted among other issues the paucity of information on selective reporting based on comparisons between review protocols and final systematic review reports. The limited number of meta-epidemiological analyses identified were restricted to cystic fibrosis reviews on the Cochrane Database of Systematic Reviews (CDSR) prior to 2010 [59], or involved analysis of selected issues within the CDSR in three time periods between 2000 and 2008 [60–62].

Historically, with the notable exception of the CDSR, systematic reviews have lacked pre-published protocols. This has changed following the relatively recent introduction of the PROSPERO database, and publication of SR protocols in isolation in a dedicated journal (www.systematicreviewsjournal.com) [63], with reporting on the existence of a pre-published protocol also encouraged within the PRISMA guidelines [45] . To our knowledge, while the existence and impact of outcome reporting bias within dental SRs has been acknowledged [64], there are no reports addressing selective reporting and discrepancies between systematic review protocols and final reports within the oral health field. Therefore, the aim of the present study was to explore the prevalence and nature of selective reporting and factors associated with selective reporting within SRs published by the Cochrane Oral Health Group (COHG) in the CDSR by analyzing discrepancies between systematic review protocols and final reports.

2.2 Methods

All systematic reviews in the Cochrane Oral Health Group (COHG) published in the CDSR until November, 2014 were screened for final reports and pre-published protocols. The COHG was contacted in order to retrieve review protocols that were not available in the Cochrane library. Reviews where the protocol could not be found as well as those reviews published in duplicate were to be excluded from further analysis.

Data were independently extracted and entered on pre-piloted standardized forms for the eligible studies. Initial calibration was performed between the two researchers (NP, PSF) on 10 articles. Disagreements were resolved by discussion or, if necessary, with adjudication by a third reviewer (KD). Information obtained included the number and type of primary and secondary outcomes both within the protocol and in the final publication. If a review did not distinguish between primary and secondary outcomes [unlabelled outcomes], the first three outcomes listed were taken to be the primary outcomes and the rest considered as secondary outcomes. Specific details as to whether outcomes present in the protocol were omitted, upgraded or downgraded as well as inclusion of outcomes in the final report that

were absent from the protocol were recorded. Data relating to year of publication of the protocol and the final report, and any previous versions of the review, number of authors, geographical location of the corresponding author, whether the collaboration involved a single or multiple centers and subject area were also extracted. In addition, information concerning statistical significance was extracted for one primary outcome from the reviews with applicable meta-analyses. The primary review comparison was assumed for each review [62] according to the following hierarchy by selecting that which met the first of the following criteria: (1) an intervention comparison described in the protocol as the primary review comparison; (2) the first intervention comparison mentioned in the objectives of the protocol; (3) an intervention comparison described in the review as the primary review comparison; (4) the first intervention comparison mentioned in the objectives of the review; (5) the intervention comparison used in the first meta-analysis presented in the review.

2.3 Data analysis

Descriptive statistics were calculated for all included articles related to the following variables: geographical representation, number of research centers, number of review updates, number of authors and year of publication. The year of publication was converted to a binary variable [<2008, $\geq$ 2008, as from 2008 Cochrane included a requirement to declare changes between protocol and review]. Cross-tabulations were undertaken to investigate associations between the presence or absence of discrepancies and review characteristics. A logistic model was also fitted in order to assess and quantify any association of discrepancies between protocol and final review and review characteristics. Finally, upgrades, downgrades, new introductions and omissions between protocol entries and final reports for all outcomes listed were tabulated. Risk ratios were calculated in order to examine the relationship between possible discrepancies in the primary outcomes and statistical significance. All statistical analyses were conducted with Stata® version 13.1 software (Stata Corporation, College Station, Texas, USA).

2.4 Results

Initially 172 titles were identified in the COHG database of which 19 reviews were excluded as duplicates [same title and DOI] and one diagnostic review was excluded (Figure 2.4.1). From the remaining 152 reviews 116 (76.3%) had accessible protocols next to the review under the ‘‘Protocol and previous versions’’ section on the Cochrane Library. The COHG provided the protocols for 36 reviews for which the protocols were not available in the Cochrane library.

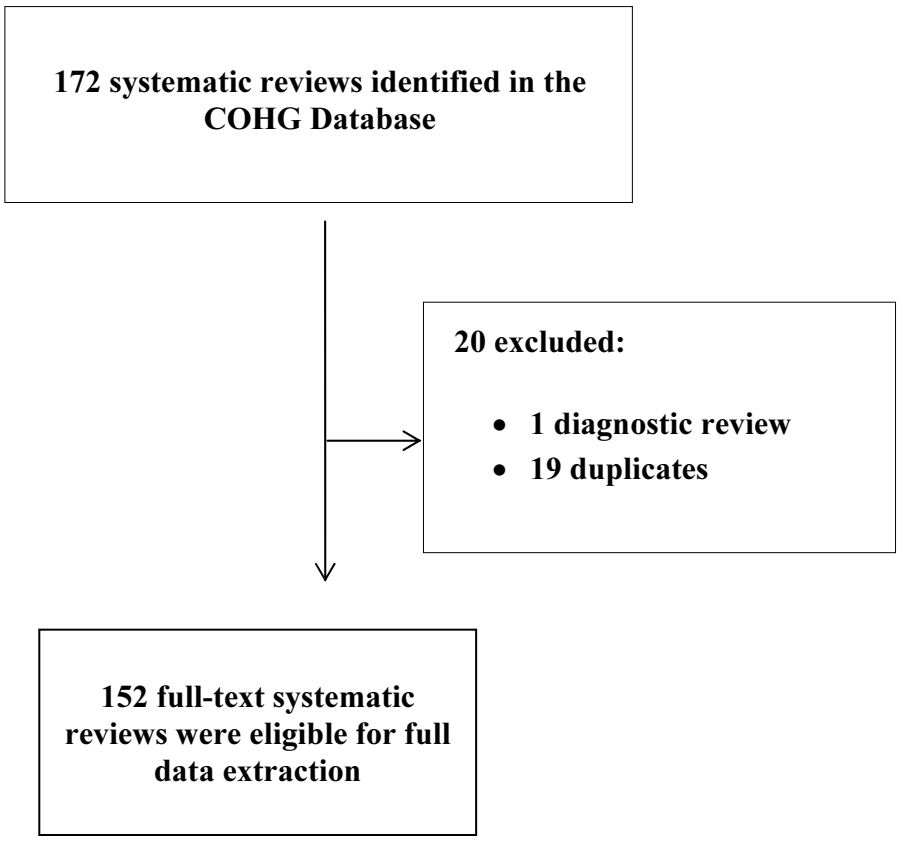

Figure 2.4.1

A variety of conditions, interventions and outcomes were considered with the highest number of reviews dealing with dental caries, orthodontics, periodontics and implantology (Table 2.4.1).

Classification		**Discrepancy**			
	Total	NO	NO*	YES	YES*
	No.	No.	%	No.	%
Anesthesia	1	1	100	0	0
Antibiotics	4	2	50	2	50
Burning mouth syndrome	1	1	100	0	0
Cancer prevention	1	0	0	1	100
Cancer treatment	4	1	25	3	75
Central giant cell granuloma	1	1	100	0	0
Cleft lip & palate	3	1	33	2	67
Complications of surgery	2	0	0	2	100
Cosmetic therapy	1	1	100	0	0
Dental anxiety	2	1	50	1	50
Dental caries prevention	19	14	74	5	26
Dental caries treatment	20	12	60	8	40
Dentistry practice and systems	1	1	100	0	0
Dry mouth	1	1	100	0	0
Gingivostomatitis	1	1	100	0	0
Halitosis	2	2	100	0	0
Impacted teeth	1	1	100	0	0
Implants and prosthesis	16	7	44	9	56
Maintenance	1	1	100	0	0
Oral care for cancer patients	5	0	0	5	100
Oral hygiene	3	2	67	1	33
Oral lesions	2	2	100	0	0
Oral leukoplakia	1	0	0	1	100
Oral lichen planus	2	0	0	2	100
Oral mucosotis	1	0	0	1	100
Oral pain	9	5	56	4	44
Oral submucous fibrosis	1	1	100	0	0
Oral ulcers	2	1	50	1	50
Orthodontic treatment	24	13	54	11	46
Periodontal disease associated	2	2	100	0	0
Periodontal disease prevention	3	0	0	3	100
Periodontal disease treatment	8	4	50	4	50
Removal of third molars	2	0	0	2	100
Traumatic injury	5	4	80	1	20
Total	152	83	55	69	45

Table 2.4.1

The largest proportion of reviews originated in the UK according to the details of the corresponding authors (Table 2.4.2).

Country of origin of corresponding author		discrepancy			
	Total	NO	NO*	YES	YES*
	No.	No.	%	No.	%
Argentina	1	1	100	0	0
Bahrain	6	4	67	2	33
Brazil	12	10	83	2	17
China	7	4	57	3	43
Croatia	1	1	100	0	0
Denmark	1	1	100	0	0
Finland	2	1	50	1	50
France	3	2	67	1	33
Germany	2	2	100	0	0
Holland	1	0	0	1	100
Iran	1	1	100	0	0
Ireland	3	2	67	1	33
Italy	6	2	33	4	67
Japan	1	1	100	0	0
Mexico	1	1	100	0	0
Nigeria	1	1	100	0	0
Oman	1	1	100	0	0
Singapore	1	1	100	0	0
South Africa	1	0	0	1	100
Switzerland	1	0	0	1	100
Syria	1	1	100	0	0
Thailand	1	0	0	1	100
UK	95	45	47	50	53
USA	2	1	50	1	50
Total	152	83	55	69	45

Table 2.4.2

The median number of labelled primary outcomes was 3 (range: 1-11) and secondary outcomes was 4 (range: 1-36) both in the protocols and published reviews. Overall, 51 protocols (33.6 %) and 34 final reviews (22.4 %) did not distinguish between primary and secondary outcomes. This was less common among reviews published after to 2008 compared to those published prior to 2008 (OR: 0.33, 95% CI: 0.14, 0.75, p=0.01) but no important improvement was observed

at protocol level for outcome distinction after 2008 compared to the previous period (OR: 0.54, 95% CI: 0.25, 1.16, p=0.11). The distribution of individual discrepancies between protocols and reviews per geographical area, single or multicenter collaboration, number of review updates, number of authors and publication period is presented in Table 2.4.3.

		Discrepancy				**p-value**
Continent of origin of corresponding author	Total	no	no	yes	yes	
	No.	No.	%	No.	%	
Middle East & Africa	10	7	70	3	30	0.09*
Americas	15	12	80	3	20	
Asia	11	7	64	4	36	
Europe	116	57	49	59	51	
	152	79	52	73	48	
Collaboration between centers						
No	4	3	75	1	25	0.41*
Yes	148	80	54	68	48	
Number of review updates						
No updates	104	69	66	35	34	<0.001**
1 update	29	12	41	17	59	
≥2 updates	19	2	11	17	89	
Number of authors						
2-3	21	11	52	10	48	0.98**
4-5	73	40	55	33	45	
≥ 6	58	32	55	26	45	
Publication period						
<2008	36	24	67	12	33	0.10**
≥2008	116	59	51	57	49	
Total	152	83	55	69	45	

**Fisher's exact test, ** Chi2 test*

Table 2.4.3

According to the multivariable logistic model (Table 2.4.4) the number of review updates appears to be associated with discrepancies between final review and protocol (OR: 3.29, 95% CI: 1.84, 5.88, p<0.001). Geographical area, single or multicenter collaboration and number of authors do not appear to be important predictors of discrepancies between protocols and final reports.

Characteristic	OR	95% CI	p-value
Number of authors [per unit]	1.12	0.92, 1.36	0.25
Geographical region			
Americas	reference	-	-
Middle East & Africa	2.02	0.30, 13.75	0.72
Asia	2.13	0.34, 13.23	0.42
Europe	2.99	0.76, 11.86	0.12
Number of updates [per unit]	3.29	1.84, 5.88	<0.001
Collaboration between centers			
No	reference	-	-
Yes	3.18	0.21, 48.16	0.40

Table 2.4.4

The detailed analysis of discrepancies (Table 2.4.5) revealed that the primary outcome was not stated in 3.9% (6/152) of protocols but stated in all final reports. In 69.1% (105/152) of the reviews the primary outcomes were consistent in the protocol and the report while secondary outcomes were the same in 56.6% of the reviews (86/152). In 11.2% (17/152) of the reviews the primary outcomes were downgraded to secondary, in 9.9% of the reviews (15/152) they were omitted and in 18.4% (28/152) new primary outcomes were introduced. In four reviews out of 152 (2.6%) a change in the definition was identified. In 2% (3/152) of the reviews secondary outcomes were upgraded to primary, 12.5% (19/152) were omitted and in 30.9% (47/152) new secondary outcomes were introduced. No outcome definition change was identified for secondary outcomes.

Type of discrepancy		
	n	%
Primary outcome variable not stated in the protocol	6/152	3.9
Primary outcome variable not stated in the published report	0/152	0.0
All primary outcome(s) stated in the protocol is/are not the same as in the published report	47/152	30.9
One or several primary outcome(s) stated in the protocol is downgraded to secondary in the published report	17/152	11.2
One or several primary outcome stated in the protocol is/are omitted from the published report	15/152	9.9
One or several new primary outcome(s) that was/were not stated in the protocol is included in the published report	28/152	18.4
The definition of one or several primary outcome(s) was different in the protocol compared to the published report	4/152	2.6
All secondary outcomes stated in the protocol are not the same as in the published report	66/152	43.4
One or several non-primary outcome(s) in the protocol is/are changed to primary in the published report	3/152	2.0
One or several secondary outcome(s) stated in the protocol is/are omitted from the published	19/152	12.5
One or several new secondary outcome(s) that was/were not stated in the protocol is/are included in the published report	47/152	30.9
The definition of one or several secondary outcome(s) was different in the protocol compared to the published report	0/152	0.0

Table 2.4.5

Overall, 45.4% (69/152) of the reviews presented at least one discrepancy but only in 14.5% (10/69) were discrepancies justified in the text (Table 2.4.6).

Author	**Explanation provided**
Guo[65]	We have revised the primary and secondary outcomes to make them more precise
Yengopal [66]	Other clinically important outcomes (e.g. gingival health and occlusion) have been added as secondary outcomes. Some included papers provided information on outcomes deemed to be clinically important to oral health professionals by the authors of these trials. These outcomes (e.g. gingival health and occlusion) did not appear in the original protocol and were added to provide a wider spectrum of outcomes that may be considered as useful to clinicians and consumers.
Rasines Alcaraz [67]	Only data on permanent posterior teeth were reported in this review In the protocol, survival rate was listed as the primary outcome but the review lists failure rate as primary outcome. Failure rate is reported in this review as a proxy for survival rate
Coulthart [68]	The primary outcome of 'altered sensation' in the protocol has been changed to 'patient-reported altered sensation' for the review
Daly [69]	We added the primary outcome for treatment of dry socket: time to heal.
Coulthart [70]	There have been some changes to the prespecified outcomes and prioritisation of outcomes
Furness[71]	The protocol for this review stated that quality of life would be a primary outcome for this review. Quality of life is an important outcome, for both patients with oral cavity and oropharyngeal cancers and their doctors. In this deadly and disfiguring disease, searching for treatments that offer an improvement in both quantity and quality of life for patients motivates the large body of research into the management of this disease. The search for effective chemotherapies is motivated at least in part by the desire to avoid patients having to undergo radical disfiguring surgery with resultant loss of function. However, as the review has progressed we have found the large quantity of research on chemotherapy focused on finding better treatments that prolong overall survival, disease free survival and progression free survival. Quality of life is inconsistently reported in trials which address a primary outcome of overall survival. Therefore we have opted to transfer this outcome to the list of secondary outcomes to be considered in future updates of this review as appropriate.
Glenny[72]	Types of outcomes: The protocol for this review stated that quality of life would be a primary outcome for this review. Quality of life is an important outcome, for both patients with oral cavity and oropharyngeal cancers and their doctors. However, quality of life is infrequently and inconsistently reported in trials which address a primary outcome of overall survival. Therefore we have opted to transfer this outcome to the list of secondary outcomes to be considered in future updates of this review as appropriate.

Cheng[73]	We did not collect data on adverse events from interventions for erosive lichen planus by running separate searches looking specifically for adverse effects of treatments used. This was an over-ambitious goal and an under-estimation of the extensive search required to fulfil this. We added in a secondary outcome (j) reduction in target/mean lesion size (for oral lesions), which was measured by two studies
Ashley[74]	Adverse events were added as an outcome

Table 2.4.6

Of the 152 reviews assessed, 89 did not have a meta-analysis or the meta-analysis did not relate to the primary outcomes. Within the remaining reviews the distribution of discrepancies for the primary outcome by statistical significance is shown in Table 2.4.7.

	Total	**Significant (<0.05)**	**Non-significant (>0.05)**	**Risk Ratio (95% CI)**	**p-value**
No discrepancy	24	16	8	Reference	-
Downgrade	9	4	5	0.52 (0.17, 1.58)	0.24
Upgrade/Inclusion	16	9	7	0.77 (0.36, 1.64)	0.50
	Total	**Significant (<0.05)**	**Non-significant (>0.05)**	**Risk Ratio (95% CI)**	**p-value**
Downgrade	9	4	5	Reference	-
Upgrade/Inclusion	16	9	7	1.19 (0.65, 2.16)	0.57

Table 2.4.7

The risk of reporting significant results was lower for downgraded outcomes [RR: 0.52, 95% CI: 0.17, 1.58, p=0.24] and upgraded or newly introduced outcomes [RR: 0.77, 95% CI: 0.36, 1.64, p=0.50]. The risk of a significant result was higher for upgraded/new outcomes (RR=1.19, 95% CI: 0.65, 2.16, p=0.57) compared to outcome downgrades. None of the comparisons reached statistical significance.

2.5. Discussion

This meta-epidemiological study attempted to identify discrepancies between 152 COHG systematic reviews and protocols in order to shed some light on selective reporting in oral health systematic reviews. The reviews covered a wide range of topics and were produced by authors located predominantly in the UK. Similar reviews have been undertaken within Cystic Fibrosis reviews on the CDSR [16] or generally within specific issues of the CDSR covering a range of review groups [17-19]. The most recent of the latter reviews focused on a period up to 2010 [16] identifying a discrepancy rate of up to 39%, while an earlier review alluded to a discrepancy rate of 22% [17]. Our findings spanning a 16-year period indicate that within SRs in the COHG discrepancies continue to be highly prevalent, with a prevalence of 45% for primary outcomes. The risk for discrepancy increased with increasing number of review updates. This is intuitive as additional outcomes can become more or less important over time; moreover, new reviewers may differ from previous reviewers in relation to the priority with which they assign to certain outcomes. The risk of statistically significant results was lower for both outcome downgrades and upgrades/new inclusion compared to outcomes with no discrepancy. The finding for upgrades/new inclusions were opposite to what was expected. The risk of statistically significant results, consistent with other studies, was higher for outcome upgrades/new inclusions compared to outcome downgrades. None of those finding reach statistical significance however, as data was thin.

Some of these problems with discrepancies between protocols and reviews will be addressed when the COHG completes research being undertaken on Core Outcome Measures in Effectiveness Trials (COMET; http://www.comet-initiative.org) which will provide a list of core outcomes to be included in reviews being undertaken in certain areas of oral health.

A large number of reviews were captured relative to published previous assessments, which have ranged from 46 to 288 SRs [16,17]. Articles were identified by just one author in the present study; however, as just a single database was searched with articles readily accessible, the risk of selection bias and inappropriate

omission of relevant articles is very low. We are, therefore, confident that the present findings are a true reflection of the discrepancy rate both for primary and secondary outcomes with the Oral Health Group of the CDSR. The Oral Health Group relates to reviews published within dentistry and oral and maxillofacial surgery; previous meta-epidemiological assessments have confirmed that dental SRs are equally susceptible to methodological and reporting weaknesses as other biomedical areas [75–78]. It is, therefore, likely that the findings from the present review are representative of the CDSR more broadly.

A large number of discrepancies were identified between protocols and final reports for both primary and secondary outcomes; these included introduction of new outcomes, upgrades, downgrades, omissions and definition changes. Since 2008, the Cochrane Collaboration requires that published reviews disclose discrepancies between protocols and reports. In this study a low percentage (14.5%) of outcome discrepancies were declared. This figure is in keeping with similar research with Kirkham *et al.* reporting acknowledgement of *post hoc* changes in just 6% of 64 reviews found to have discrepancies [17]. However, in a more recent review [16] discrepancies were mentioned in 39%, albeit based on a smaller sample of 46 reviews overall. These figures, however, may indicate a lack of awareness among review authors and failure of reviewers to cross-reference final submissions with pre-published protocols. While there is ample evidence that SRs published on the CDSR are of higher methodological quality than those published elsewhere [79–81] the stipulation of disclosure of discrepancies within the CDSR does not appear to have had the desired effect either within the COHG or more generally.

In the present subset, no statistically significant association between alteration in outcomes (upgrade from secondary to primary, introduction of new outcomes, downgrades) and statistically significant of the meta-analysis results was found. The risk for reporting significant results was higher for upgraded or newly introduced outcomes compared to downgraded outcomes, although a non-statistically significant finding. Kirkham *et al.* [62] found an increased risk of obtaining a significant result with discrepancies such as both new inclusion or upgrade but not with downgrades. This inconsistency with our study for only upgrades/new inclusion

of outcomes may be attributed to the relatively small number of included meta-analyses in the present study but may also relate to the type of relevant outcomes.

The continued stubbornly high outcome discrepancy rate within the CDSR suggests that the prerequisite of delineating changes between protocol and submitted review is ineffective. However, external screening by the Cochrane Editorial Unit from 2013 has helped to ensure these changes are documented and justified in new reviews published since September 2013.

An alternative may be to consider a streamlined approach incorporating checks ensuring that outcomes in the review correspond with those delineated in the protocols. This approach could be implemented at the editorial level with proprietary software. For example, outcomes from the protocol may be automatically transferred to the review. Where authors wish to make changes in the review those changes would only be permitted if a declaration for changes field was completed. The editor could then be alerted of the outcome changes between protocol and review prompting the need for approval and registering the change.

In terms of predictors, it appears differences between protocols and published reviews were more likely with increasing numbers of review updates. This finding is intuitive and may reflect the fact that reviews may span periods in excess of a decade and may, therefore, involve different author groups. Moreover, there are instances of multiple reviews being derived from the same protocol. The correct protocols for these "split" reviews were provided by the COHG group office and were clearly identified. It may, however, be sensible to consider more regular update of SR protocols within the CDSR after a defined time period to mitigate this.

2.6 Conclusions

Although the requirement to declare outcome changes between protocols and reviews which was implemented in 2008 is certainly a positive, there is further evidence that discrepancies still exist based on this analysis of SRs published within the Oral Health Group since its establishment. Alternative approaches to reduce the prevalence of this issue including the use of Core Outcome Sets need to be explored.

Chapter 2 is based on the following publication: Pandis N, Fleming PS, Worthington H, Dwan K, Salanti G (2015) Discrepancies in Outcome Reporting Exist Between Protocols and Published Oral Health Cochrane Systematic Reviews. PLoS ONE 10(9): e0137667. doi:10.1371/journal.pone.0137667. **Open access journal allows free use of published content.**

Chapter 3

The quality of the evidence according to GRADE is predominantly low or very low in oral health systematic reviews

3.1 Introduction

Systematic reviews aim to assimilate high-quality evidence on the area of interest in a systematic, transparent, and unbiased manner leading to qualitative or quantitative synthesis. Quantitative synthesis can produce a more precise estimate on the efficacy and safety of a therapy, and can reconcile misunderstandings and controversies, and form the basis future trials. Healthcare practitioners should seek high quality evidence when searching for answers to clinical questions in relation to the effectiveness or otherwise of a proposed intervention. Several meta-epidemiological studies have been published on the quality of systematic reviews in oral health, with evidence to suggest both reporting and methodological deficiencies common to systematic reviews within a range of specialty areas; these findings have also mirrored analogous research studies in other biomedical areas. [75–77, 79, 82, 83] However, the overall quality of the existing evidence in oral health has not yet been assessed. Moreover, there is increasing concern of a gulf between research evidence and its clinical applicability. Patients are not typically conversant in the scientific publications but are exercised by the implications of research findings on their lives and wellbeing. It has become evident that a system capable of simultaneously assessing the quality of the evidence, balancing benefits and harms, while accounting for patient preferences and aiding clear treatment recommendations is imperative. This approach would resonate both within medicine and in dentistry where an increasing number and quality of systematic reviews are published.[75–77, 79, 82, 83]

Several groups have proposed complex methods for evaluating and translating evidence into clinical practice; many of these have been somewhat confusing and impractical [84]. The GRADE (Grades of Recommendation, Assessment, Development, and Evaluation) initiative, however, has become an accepted approach for assessing the evidence and consequently making recommendations [85]. The GRADE approach is predicated on a precise clinical question, with consideration of all important outcomes within a systematic review prioritizing them based on their relative importance.[86] Subsequently, the existing quality of the evidence for an outcome from a systematic review is assessed based on a specific protocol and graded as high, moderate, low or very low (Table 3.1.1) [87].

Rank	
High **++++**	Further research is very unlikely to change our confidence in the estimate of effect
Moderate **+++**	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low **++**	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low **+**	Any estimate of effect is very uncertain

Table 3.1.1

While, precursors relied heavily on the overall study design (randomized vs. non-randomized studies) to evaluate a body of evidence, study design remains important within the GRADE assessment, although not the sole arbiter of the quality of evidence. Randomized studies provide a higher quality of evidence compared to observational studies; however, good quality observational studies are more valuable than uncontrolled cases series designs[88]. Randomized controlled trials (RCTs) are initially rated as high quality but may be downgraded after accounting for study limitations, indirectness, inconsistency, imprecision, and publication bias. Observational studies start at low quality but may be upgraded in the presence of a large magnitude of effect even after confounding has been considered and dose response effect. In the GRADE system, although expert opinion does not command a

quality of evidence score, it is considered critical in interpreting, combining and placing the available evidence in the correct context[88].

The GRADE approach has been endorsed by a large number of societies and institutions including the Cochrane Collaboration [89] and has been used in various fields of medicine such as allergies, heart disease, oncology, endocrinology and respiratory and critical care medicine [90–93] also being used to appraise the level of evidence to support medical interventions in two relatively small prior assessments [94] [95]. While GRADE is applicable within the oral health field [96, 97], there are no previous reports of an overall assessment of the quality of evidence in this area. Therefore, the aim of this study is to assess the quality of the evidence in contemporary oral health systematic reviews using GRADE.

3.2 Methods

Eligible SRs included at least one meta-analysis of at least 2 original studies; only one meta-analysis per systematic review was considered. The selected meta-analysis was that reporting on the primary outcome or the first or most important reported outcome if the primary outcome was not specifically outlined. If meta-analysis of more than one primary outcome was available, the meta-analysis including the largest number of trials was selected.

The following pre-specified exclusion criteria were applied:

- SRs with forest plots without a pooled estimate
- Duplicate publications, laboratory studies and reviews of animal studies
- Quantitative synthesis within single arms either due to absence of a control or due to analysis of before-after measurements
- Reviews including meta-analysis, which included the same studies multiple times without explanation on whether subgroups were mutually exclusive

Systematic reviews using network meta-analysis -Diagnostic test accuracy reviews or reviews on prognostic factors

Systematic reviews published from January 2008 [year of GRADE adoption by the Cochrane Collaboration] until the end of 2013 were retrieved from the Oral Health Group (OHG) of the Cochrane Database of Systematic Reviews (CDSR, IF: 5.785) and by hand searching of 14 general and specialty dental journals (January 2008-December 2013) with the highest impact factor (IF) in 2012 (Table 3.2.2).

		Impact factor	**N (%)**
Journal	American Journal of Orthodontics and Dentofacial Orthopedics (AJODO	1.458	5 (5.5%)
	Caries Research (CR)	2.514	1 (1.1%)
	COCHRANE Database of Systematic Reviews: Oral Health Group (CDSR-OHG)	5.785	41 (45.1%)
	Clinical Implant Dentistry and Related Research (CIDRR)	3.821	4 (4.4%)
	Clinical Oral Investigations (COI)	2.2	3 (3.3%)
	Clinical Oral Implant Research (COIR)	3.433	5 (5.5%)
	International Journal of Prosthetic Dentistry (IJPD)	1.625	1 (1.1%)
	Journal of Clinical periodontology (JCP)	3.688	7 (7.7%)
	Journal of Dentistry (JD)	3.2	4 (4.4%)
	Journal of Dental Research (JDR)	3.826	4 (4.4%)
	Journal of Endodontics (JOE)	2.929	3 (3.3%)
	Journal of Oral and maxillofacial Surgery (JOMS)	1.521	6 (6.6%)
	Journal of Periodontology (JP)	2.398	7 (7.7%)
Total			91 (100%)

Table 3.2.2

The titles and abstracts were initially read by one investigator (NP) and all full-text articles were retrieved and screened for inclusion. A second screening of the full reports was undertaken by one investigator (NP) resulting in exclusion of the reviews from further analyses based on the pre-specified exclusion criteria. Information was collected from all selected review articles at the review, meta-analysis and trial level (Appendix, Table 1).

The conclusions of selected meta-analyses were assessed using the GRADE approach by the first author, who was familiar with GRADE. A second author (PSF) verified the ratings; any disagreements were reconciled after discussion. The selected meta-analyses were assessed in relation to the quality of the evidence scored in the 5 domains specified within GRADE: Limitations in study design and/or execution (Risk of Bias) [88], inconsistency of results [98], indirectness of evidence[99], imprecision of results [100], and publication bias [101]. Details on the assessment par domain are included in the appendix.

3.3 Data Analysis

The objectives of the statistical analyses were to tabulate frequency distributions of specific characteristics in the dental SR sample at the review, and meta-analysis levels together with the frequency distributions of the quality of the evidence in relation to GRADE for Cochrane and non-Cochrane SRs. Associations between SR/meta-analysis characteristics and GRADE assessment were also considered. The four-level GRADE rating (high, moderate, low, very low) was converted into a binary variable (high/moderate and low/very low) in order to fit a logistic regression model to assess potential associations between GRADE rating, impact factor and publication year. GRADE was the dependent variable and impact factor and publication year were the examined predictors. All the analyses were performed with Stata statistical software version 13.1 (Stata Corporation, College Station, Texas, USA).

3.4 Results

From the 510 SRs initially considered for inclusion in the study, 91 were included in the final assessment (Figure 3.4.1).

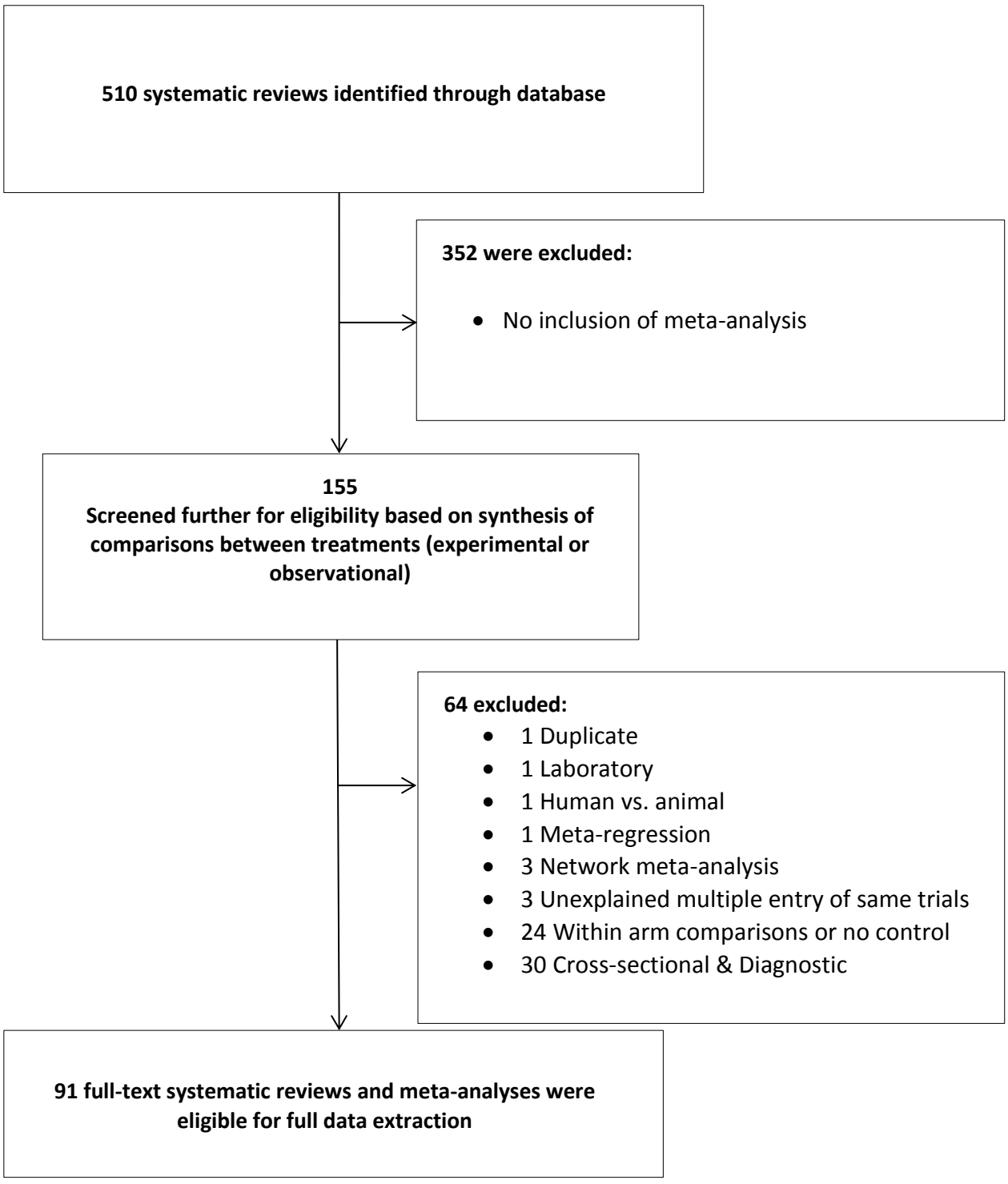
510 systematic reviews identified through database
352 were excluded:
No inclusion of meta-analysis
155
Screened further for eligibility based on synthesis of comparisons between treatments (experimental or observational)
64 excluded:
1 Duplicate
1 Laboratory
1 Human vs. animal
1 Meta-regression
3 Network meta-analysis
3 Unexplained multiple entry of same trials
24 Within arm comparisons or no control
30 Cross-sectional & Diagnostic
91 full-text systematic reviews and meta-analyses were eligible for full data extraction

Figure 3.4.1

The frequencies of SRs per journal were outlined in Table 3.2.2. Fifty SRs were included from the selected dental journals and 41 SRs from CDSR; a variety of conditions, interventions and outcomes were considered. The number of published SRs with meta-analyses increased over time (Figure 3.4.2).

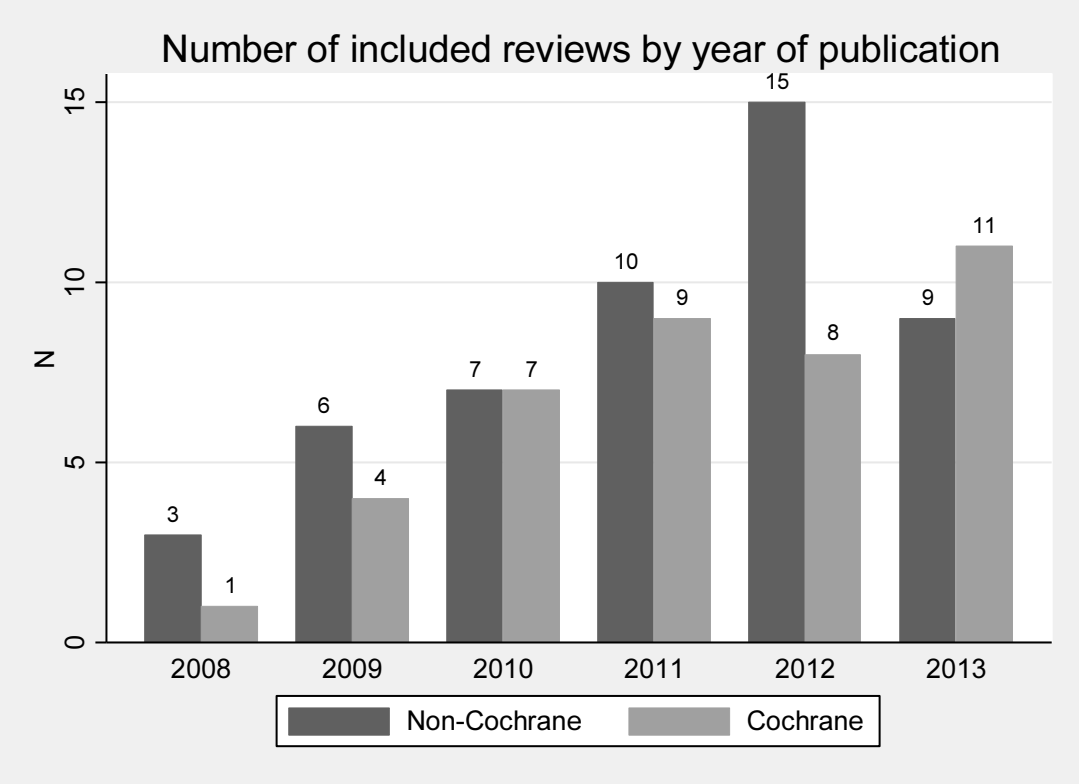

Figure 3.4.2

Significant differences were found between Cochrane and non-Cochrane reviews in respect of the region of authorship, involvement of a methodologist, number of collaborating centers, inclusion of GRADE assessment and assessment of harms in the outcome category. The Cochrane compared to non-Cochrane reviews are more likely to originate in Europe (OR: 14.62, 95% CI: 1.76, 121.19, p=0.01) or Asia (OR: 1.43, 95% CI: 0.11, 18.00, p=0.78) compared to Americas , to involve authors across multiple centers (OR: 5.26, 95% CI: 1.97, 14.09, p=0.001), to include a methodologist (p<0.001) and a GRADE assessment (p<0.001), and to consider at least one harm in the outcomes (OR= 2.66, 95% CI: 1.13, 6.27, p=0.03) (Table 3.4.3)."

Characteristic	Non- Cochrane Review	Cochrane Reviews	Total	Odds Ratio	95% CIs**	p-value#
	N(%)	N(%)	N(%)			
Continent of first author affiliation						
Americas	10 (20%)	1 (2%)	11 (12%)	Reference	-	
Europe	26 (52%)	38 (93%)	64(70%)	14.62	1.76, 121.19	0.01
Other	14 (28%)	2 (5%)	16(18%)	1.43	0.11, 18.00	0.78
Methodologist involvement						
No	33 (66%)	0 (0%)	33(36%)	Reference	-	
Yes	17 (34%)	41 (100%)	58(64%)	73.76	9.79, 555.93	<0.001
Collaboration between centers						
No	26 (52%)	7 (17%)	33(33%)	Reference	-	
Yes	24 (48%)	34(83%)	58(64%)	5.26	1.97, 14.09	0.001
GRADE assessment by SR original authors						
No	45 (90%)	17 (41%)	62(68%)	Reference	-	
Yes	5 (10%)	24(59%)	29(32%)	12.71	4.17, 38.69	<0.001
At least one harm (outcome) examined						
No	29 (58%)	14(34%)	43(47%)	Reference	-	0.03
Yes	21 (42%)	27(66%)	48(53%)	2.66	1.13, 6.27	
Total	50(55%)	41(45%)	91(100%)			

Table 3.4.3

“The Cochrane SRs included only randomized trials but the non-Cochrane reviews included both randomized and non-randomized studies and occasionally a combination of both designs. Authors of Cochrane reviews were more likely to account for clustering effects (OR: 15.62, 95% CI: 2.81, 86.76, p=0.002) often encountered in dentistry either in the analysis or discussion and to correctly analyze paired data (OR: 11.25, 95% CI: 11.05, 541.20, p=0.02) compared to non-Cochrane reviews (Table 3.4.4).

Characteristic	Non- Cochrane Review	Cochrane Reviews	Total	Odds Ratio	95% CIs[##]	p-value
	N (%)	N (%)	N (%)			
Study type						
Randomized	34(68%)	41 (100%)	75 (82%)	Not estimable		<0.001[#]
Non-randomized	5 (10%)	0(0%)	5 (6%)	Not estimable		
Mixed	9 (18%)	0 (0%)	9 (10%)	Not estimable		
Not-reported	2 (4%)	0 (0%)	2 (2%)	Not estimable		
Model						
Random	33 (66%)	25 (61%)	58 (66%)	Reference	-	-
Fixed	17 (34%)	16 (39%)	33 (34%)	1.24	0.53, 2.93	0.62
Outcome Type*						
Subjective	6 (12%)	10 (24%)	16 (18%)	Reference	-	-
Objective	44 (88%)	31 (76%)	75 (82%)	2.37	0.78, 7.20	0.13
Outcome Scale						
Binary	28 (56%)	22 (54%)	50 (56%)	Reference	-	
Continuous	19 (38%)	15 (37%)	34 (37%)	1	0.42, 2.42	0.99
Ordinal	3 (6%)	4 (10%)	7 (7%)	1.70	0.34, 8.39	0.52

Effect measure						
Hazard ratio (HR)	0 (0%)	2 (5%)	2 (2%)	Not estimable		0.83[#]
Mean difference (MD)	18 (36%)	10 (24%)	28 (31%)	Not estimable		
Odds ratio (OR)	9 (18%)	4 (10%)	13 (14%)	Not estimable		
Preventive fraction(PF)	0 (0%)	2 (5%)	2 (2%)	Not estimable		
Risk difference (RD)	1 (2%)	0 (0%)	1 (1%)	Not estimable		
Risk ratio (RR)	19 (38%)	15 (37%)	34 (37%)	Not estimable		
Standardized mean difference (SMD)	3 (6%)	8 (19%)	11 (12%)	Not estimable		
Total	50(100%)	41(100%)	91(100%)			
Clustering accounted/discussed**						
No	25 (76%)	2 (17%)	27 (60%)	Reference	-	-
Yes	8 (24%)	10 (83%)	18 (40%)	15.62	2.81, 86.76	0.002
Total	33 (100%)	12 (100%)	45 (100%)			
Paired design accounted/discussed**						
No	9 (64%)	0 (0%)	8 (35%)	Reference		
Yes	5 (36%)	10 (100%)	15 (65%)	11.25	11.05, 541.20	0.02
Total	14 (100%)	10 (100%)	23 (100%)			

ns = non-significant
*** Significant at 0.01*
#Pearson X^2 test or Fisher's exact test
P-value based on High/Moderate vs Low/Very Low for GRADE and on No vs Serious/Very serious for the GRADE domain comparisons

Table 3.4.4

A variety of approaches were used for the assessment of the methodological quality with the Cochrane risk of bias tool used in 62 out of the 91 SRs. In some SRs, a combination of approaches was used and in some instances incorrectly reporting guidelines were utilized.

The overall GRADE rating results based on 4 levels (high, moderate, low and very low) are shown in Table 3.4. 5. A more detailed description of the GRADE process is shown in the Appendix table 2. The GRADE judgment for Cochrane reviews was re-assessed independent of the findings of the original review authors. The GRADE assessment indicated that only around 20% of the evidence belongs to the moderate/high quality category with the remaining SRs at the low/very low level (Table 3.4.5). The distribution of the overall GRADE ratings and ratings for each GRADE domain were similar for Cochrane and non-Cochrane reviews (Table3.4. 5). The most commonly downgraded domains were those for study limitations and imprecision. The frequency of downgrading by 2 levels (very serious) was higher in the non-Cochrane reviews (46%) versus the Cochrane reviews (22%); the former include also non-randomized studies (Table 3.4.5).

	Non- Cochrane Review	Cochrane Review	Total	p-value #, ##
	N (%)	N (%)	N (%)	
GRADE rating				
High	2 (4%)	0 (0%)	2 (2%)	ns
Moderate	9 (18%)	11 (27%)	20 (22%)	
Low	24(48%)	19 (47%)	43 (47%)	
Very Low	15(30%)	11 (27%)	26 (29%)	
Domains				
Study limitations/Risk of Bias				
No	2 (4%)	2 (5%)	4 (4%)	ns
Serious	25(50%)	30 (73%)	55 (60%)	
Very Serious	23(46%)	9 (22%)	32 (35%)	
Inconsistency				
No	34(72%)	31(76%)	65 (71%)	ns
Serious	15(26%)	10(24%)	25 (27%)	
Very Serious	1(2%)	0(0%)	1 (1%)	
Indirectness				
No	50(100%)	38 (93%)	88 (97%)	ns
Serious	0 (0%)	3 (7%)	3 (3%)	
Very serious	0 (0%)	0 (0%)	0 (0%)	
Imprecision				
No	19(38%)	19 (46%)	38 (42%)	ns
Serious	23(46%)	18 (44%)	41 (45%)	
Very Serious	8 (16%)	4 (10%)	12 (13%)	
Publication bias				
Suspected	8 (16%)	0 (0%)	8 (9%)	**
Unsuspected	42(84%)	41 (100%)	83 (91%)	
GRADE rating as reported in the review				
High	2 (4%)	0 (0%)	2 (2%)	
Moderate	1 (2%)	9 (22%)	10 (11%)	
Low	0 (0%)	10 (24%)	10 (11%)	
Very Low	2 (4%)	3 (8%)	5 (6%)	
Not reported	45(10%)	19 (46%)	64 (70%)	
Total	50 (100%)	41 (100%)	91(100%)	

Table 3.4.5

The inconsistency domain received similar ratings for non-Cochrane and Cochrane reviews in agreement with the reported statistical heterogeneity represented by I^2 (Figure 3.4.3).

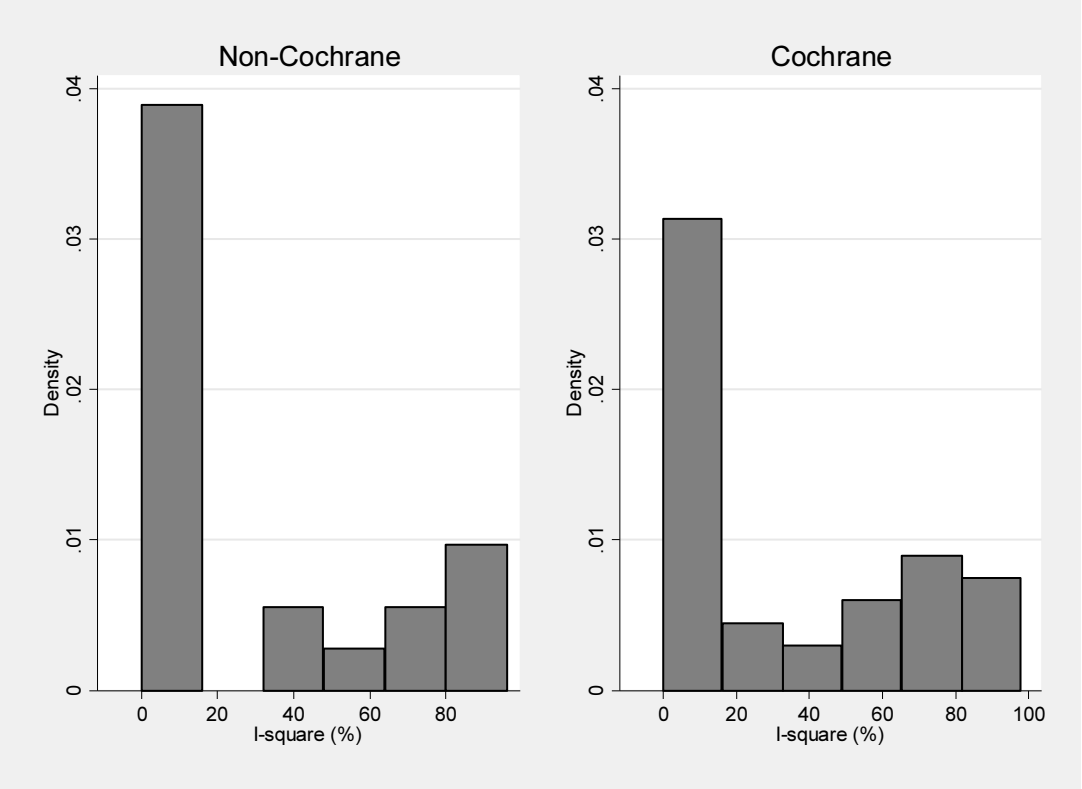

Figure 3.4.3

A GRADE assessment was included in 5/50 (10%) of the non-Cochrane and in 22/41 (54%) of the Cochrane reviews, respectively. Only three discrepancies were observed among the 27 included in the SRs between the evidence rating found by the initial review authors and our re-analysis. One review published [102] in the JCP was based a previous CDSR systematic review which did not include a GRADE assessment[103] . Logistic regression analysis indicated that neither the journal impact factor (OR=0.92, 95% CI: 0.68, 1.25, p=0.61) nor the year of publication (OR: 1.17, 95% CI: 0.82, 1.67, p=0.40) are important predictors for the quality of the evidence (Reference group: low/very low evidence).

3.5 Discussion

This cross-sectional meta-epidemiologic study is the first attempt to evaluate the evidence across the oral health field and has exposed that only 20% of the existing evidence derived from SRs in relation to primary outcomes is at the moderate or high level with no difference between Cochrane and non-Cochrane reviews. This finding indicates that 80% of the evidence considered, much of which will form the basis of clinical recommendations, is at the low or very low level. This finding is alarming indicating the need to undertake high quality clinical trials in order to reduce the uncertainty around therapies in the oral health field. Empirical studies in oral health and medicine in general using GRADE to assess the quality of evidence are sparse [94, 95]. Our study included 91 SRs and covered the entire field of oral health systematic reviews; it is likely that the quality of the evidence may vary across, between and within healthcare specialties. We were able to identify only two studies across the range of biomedical fields where GRADE was applied to assess the quality of the evidence. The scope of those studies was limited to hyperbaric oxygen therapy indications [95] and nonsurgical treatment of stress urinary incontinence [94]. In the analysis of hyperbaric oxygen treatment[95], the quality of the evidence ranged from very low to high depending on the indication with key modifiers identified as risk of bias, imprecision and large effect for observational studies. Similarities with our study include reasons for downgrading the quality of the evidence based on risk of bias and imprecision. The review on nonsurgical treatment of stress urinary incontinence involved assessment of 13 reviews finding that the quality of the evidence ranged from low to high depending on the intervention [94].

Cochrane reviews are considered to be particularly rigorous and are conducted according to strict criteria and usually include only randomized studies[79] . In the present study, the non-Cochrane reviews had a relatively small proportion of non-randomized studies; however, the quality of evidence did not differ among the groups suggesting the quality of studies populating both Cochrane and non-Cochrane reviews to be lacking. The two most frequently downgraded domains were study limitations and imprecision. This finding was consistent with a recent study evaluating the quality of evidence in hyperbaric oxygen therapy [95]. The latter

review, however, was also based on a relatively limited subset of 17 reviews, with the majority of primary studies (75%) being non-randomized studies[14].

In relation to the characteristics of the SRs, Cochrane reviews were more likely to be published by European authors, to involve authors across multiple centers, to involve a methodologist, and to consider at least a single harm in the outcomes than was the case for non-Cochrane reviews. However, the strict methodological and reporting criteria and the involvement of a methodologist did not result in higher quality of evidence. This is logical as, while Cochrane reviews are likely to be conducted and reported to a higher standard, published SRs are based on studies from the same pool, regardless of the effect of methodological quality of the review itself, review authorship and associated expertise.

The SRs included a median of 5 meta-analyses, each comprising of a median of 4 studies. These figures are in keeping with a large survey of the CDSR involving analysis of 2,321 SRs which highlighted a median of 6 meta-analyses per review and a median of 3 studies per meta-analysis[104]. A similar previous review[105] of dental SRs reported a median of 9 studies were included in the largest meta-analysis within dental SRs involving 9 dental specialties, with the largest meta-analysis having no more than 4 studies in 19% of reviews. Similarly, the largest meta-analysis involved a median number of just 2 randomized studies, although that review referred back to SRs from as long ago as 1991 when randomized studies were considerably less prevalent in oral health than is now the case. In oral health, due to the fact that multiple matched or unmatched sites can receive the intervention of interest, clustering effects are common. If this is handled improperly during the analysis, significant results may arise which are not genuine [106–108]. Additionally, a common design which uses matching called split-mouth studies requires consideration of the within patient correlations during the synthesis. In split-mouth designs, clustering effects can also exist when both interventions are applied within the same patient and in the presence of multiple sites per treatment arm as is the case with teeth [108, 109]. Studies with clustering effects, paired data or a mixture of these were handled more appropriately in the Cochrane SRs.

The Cochrane Collaboration adopted GRADE in 2008; this may explain the inclusion of a GRADE assessment more frequently in the CDSR with 22/41 (82%) Cochrane SRs including a GRADE assessment versus 5/50 (10%) in the non-Cochrane reviews. We conducted our own assessment of GRADE highlighting a discrepancy in only 11% of occasions with the rating of the SR authors confirming the reliability of the GRADE approach when assessing the evidence [110] . Other investigators found larger inconsistencies during the application of GRADE [111], although the level of familiarity of these investigators with GRADE is unclear. In the present study one of the investigators who assessed the quality of the evidence had attended a GRADE course and both GRADE assessors have implemented GRADE previously in systematic reviews.

Older SRs, time from search to publication and delays in SR updates may lead to important differences in included studies compared to more recent reviews [112]. Publication year, however, was not found to have a bearing on the quality of the evidence in the present study. Similarly, journal impact factor showed no association with the quality of the evidence. There is some evidence that SRs published in higher impact medical journals are of higher methodological quality [113]; however, this does not necessarily translate into higher quality of evidence. SR methodological quality should not be confused with the quality of the evidence, although more detailed and broader searches in particular may lead to the identification of more eligible studies and potentially lead to higher quality of evidence.

The present study did not include a cross-section of all oral health SRs but focused on the CDSR and dental journals with the highest impact factor, which may have potentially influenced the results, although the pool of potential candidate studies for inclusion is similar. Nevertheless, the observed findings might represent a best-case scenario. The several steps of the GRADE assessment may introduce an element of inconsistency in the ratings; it is, therefore, important that the complete picture is considered along with the judgment made. Rating of study limitations for non-randomized studies can be problematic especially when the information provided in the review is limited. A particularly challenging area to rate is publication bias as it is difficult to detect exclusion of eligible studies. Both the description of the literature

search, reference to grey literature, trial registries and reference lists of included studies during the assessment of publication bias were considered to reach this decision. Statistical assessment of publication bias was typically impossible due to the small number of included studies in the meta-analyses [114]. It is of interest to note that the Agency for Healthcare Research and Quality Evidence based Practice Center Program considers publication bias to be an optional domain [115].

In the present study, one author made the GRADE rating with scores verified by a second author. In a previous study [111] implementation agreement ranged from low to high depending on the domain and concluded that both training in GRADE and clinical expertise are instrumental in improving consistency in its use. GRADE permits judgment in a methodical and transparent way [116] with inconsistencies relating to the type and number of outcomes considered during the assessment [111]. The approach favors an overall rating which may be considered as a continuum hinging on expertise and judgment resulting in a judgment where the overall may not be the sum of all parts [116]. Nevertheless, as an incidental finding, significant agreement was observed between the ratings made in the constituent reviews and in the present cross-sectional study. For the purposes of assessing the association between GRADE rating and publication year and impact factor, the 4-level scale was converted in a 2-level scale. We understand that GRADE uses 4 levels for a purpose and presenting the conclusions in a 2-level scale may have limitations. The 2-level presentation can help in communicating the results especially when applying GRADE for recommendations.

This study dealt with only one outcome for which meta-analyses were available. A more comprehensive approach could have involved consideration of several or all outcomes and could have included evidence from qualitative synthesis. However, this would have been very difficult to implement given the large number of SRs. Furthermore, we feel that inclusion of only the primary or first outcome over a wide range of oral health SRs is likely to be a good proxy of the quality of the evidence in the field of oral health research.

3.6. Conclusions

Only a small proportion (20%) of the studies assessing interventions in oral health was of moderate or high quality according to GRADE. The most common domains provoking downgrading of the evidence were study limitations and imprecision indicating the need for larger and higher quality trials to inform clinical decisions. The lack of robust evidence underpinning dental procedures should prompt a concerted drive to conduct funded, high quality clinical trials in order to improve the quality of the evidence for accepted but often unproven procedures.

Chapter 3 is based on the following publication: Pandis N, Fleming PS, Worthington H, Salanti G (2015) The Quality of the Evidence According to GRADE Is Predominantly Low or Very Low in Oral Health Systematic Reviews. PLoS ONE 10(7): e0131644. doi:10.1371/journal.pone.0131644. **Open access journal allows free use of published content**

Conclusions

The NMA results indicate that the conventional appliances perform better in terms of alignment efficiency compared to all other systems with a greater mean improvement of 0.03, 0.08 and 0.17 mm/month with conventional compared to In-Ovation-R, Damon and Smart-Clip, respectively. The estimated differences are not statistically significant and more importantly the estimates are of little clinical importance. The results should be interpreted with caution as the number of studies is small and the associated confidence intervals are relatively wide indicating imprecision of the estimated treatment effects. The comparisons among In-Ovation-R, Damon and Smart-Clip yielded differences of limited clinical relevance.

Discrepancies between protocols and final reviews were evident for both primary and secondary outcomes. Overall, 45.4% of reviews had at least one discrepancy with their protocol; these were justified in 14.5% reviews. The risk of reporting significant results were lower for both downgraded outcomes and upgraded or newly introduced outcomes compared to outcomes with no discrepancies. The risk for reporting significant results was higher for upgraded or newly introduced outcomes compared to downgraded outcomes. None of the comparisons reached statistical significance. Alternative solutions to reduce the prevalence of this issue may need to be explored.

The quality of evidence according to GRADE was high/moderate in only 20% of meta-analyses with no difference between Cochrane and non-Cochrane reviews, journal impact factor or year of publication. The most common domains prompting downgrading of the evidence were study limitations and imprecision. Those findings suggest that there is a pressing need for more studies of higher quality in order to inform clinical decisions thereby reducing the risk of instituting potentially ineffective and/or harmful therapies.

Summary in English

1. Initial orthodontic alignment effectiveness with self-ligating and conventional appliances: a systematic review and network Meta-Analysis.

Systematic reviews of well-designed trials constitute a high level of scientific evidence and are important for medical decision-making. Meta-analysis facilitates integration of the existing evidence using a transparent and systematic approach leading to a broader interpretation of treatment effectiveness and safety than can be attained from individual studies. Traditional meta-analyses are limited to comparing just two interventions concurrently and are unable to combine evidence concerning multiple treatments. A relatively recent extension of the traditional meta-analytical approach is network meta-analysis (NMA) allowing, under certain assumptions, the quantitative synthesis of all evidence under a unified framework and across a network of all eligible trials. NMA combines evidence from direct and indirect information via common comparators; interventions can therefore be ranked in terms of the analyzed outcome. In the first part the efficiency in orthodontic alignment is compared among different bracket systems under the network meta-analysis framework. The NMA results indicate that the conventional appliances perform better in terms of alignment efficiency compared to all other systems with a greater mean improvement of 0.03, 0.08 and 0.17 mm/month with conventional compared to In-Ovation-R, Damon and Smart-Clip, respectively. The estimated differences are not statistically significant since the associated 95% confidence intervals include the value zero and more importantly the estimates are of little clinical importance. If we assume that an average duration for initial alignment is around 4 months the results suggest that conventional appliance will be more efficient on average anywhere from 0.12 to 0.68 mm over the 4-month period compared to the other three brackets. The expected 4 month differences were calculated by multiplying by 4 (number of months) the minimum and maximum NMA estimates for the comparisons of conventional against the other systems. The results should be interpreted with caution as the number of studies is small and the

associated confidence intervals are relatively wide indicating imprecision of the estimated treatment effects. The comparisons among In-Ovation-R, Damon and Smart-Clip yielded differences of limited clinical relevance.

2. Discrepancies in outcome reporting exist between protocols and published oral health Cochrane systematic reviews.

Inappropriate practices such as selective inclusion of trials, selective reporting of outcomes and results-driven reporting can bias estimates of treatment effects culminating in compromised patient care, healthcare decisions and service configuration. Selective outcome reporting involving preferential reporting of specific data or outcomes within a study is a recognized problem and has been investigated in respect of randomized controlled trials (RCTs), in particular.

While research on selective reporting has focused both on RCT s and non-randomized studies similar issues may arise within SRs. At the systematic review level, selective reporting may develop for a variety of reasons, for example, due to the use of multiple measurement scales, outcomes or time points and selective inclusion of specific outcomes. Arbitrary inclusion of outcomes based on post hoc results-driven decisions may bias the conclusions from subsequent syntheses.

In the second part discrepancies on the analyzed outcomes between protocols and published reviews in oral health systematic reviews (COHG) on the Cochrane Database of Systematic Reviews (CDSR) were assessed.

All Systematic Reviews of COHG in CDSR and the corresponding protocols were retrieved and information on the reported outcomes was recorded. Data was collected at the systematic review level by two reviewers independently.

One hundred and fifty two reviews were included. In relation to primary outcomes, 11.2% were downgraded to secondary outcomes, 9.9% were omitted altogether in the final publication and new primary outcomes were identified in 18.4% of publications. For secondary outcomes, 2.0% were upgraded to primary, 12.5% were omitted and 30.9% were newly introduced in the publication. Overall, 45.4% of reviews had at least one discrepancy with their protocol; these were justified in 14.5% reviews. The risk of reporting significant results were lower for both

downgraded outcomes [RR: 0.52, 95% CI: 0.17, 1.58, p=0.24] and upgraded or newly introduced outcomes [RR: 0.77, 95% CI: 0.36, 1.64, p=0.50] compared to outcomes with no discrepancies. The risk for reporting significant results was higher for upgraded or newly introduced outcomes compared to downgraded outcomes (RR=1.19, 95% CI: 0.65, 2.16, p=0.57). None of the comparisons reached statistical significance

There is evidence that discrepancies between outcomes on pre-published protocols and final reviews continue to be common based on this analysis of SRs published within the COHG. Alternative solutions to reduce the prevalence of this issue may need to be explored.

3. The quality of the evidence according to GRADE is predominantly low or very low in oral health systematic reviews

There is increasing concern of a gulf between research evidence and its clinical applicability. Patients are not typically conversant in the scientific publications but are exercised by the implications of research findings on their lives and wellbeing. It has become evident that a system capable of simultaneously assessing the quality of the evidence, balancing benefits and harms, while accounting for patient preferences and aiding clear treatment recommendations is imperative. This approach would resonate both within medicine and in dentistry where an increasing number and quality of systematic reviews are published.

Several groups have proposed complex methods for evaluating and translating evidence into clinical practice; many of these have been somewhat confusing and impractical. The GRADE (Grades of Recommendation, Assessment, Development, and Evaluation) initiative, however, has become an accepted approach for assessing the evidence and consequently making recommendations. The GRADE approach is predicated on a precise clinical question, with consideration of all important outcomes within a systematic review prioritizing them based on their relative importance. Subsequently, the existing quality of the evidence for an outcome from a systematic review is assessed based on a specific protocol and graded as high, moderate, low or very low.

In the final part the main objective was to assess the credibility of the evidence using Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) in oral health systematic reviews on the Cochrane Database of Systematic Reviews (CDSR) and elsewhere.

Systematic Reviews (January 2008-December 2013) from 14 high impact general dental and specialty dental journals and the CDSR were screened for meta-analyses. Data was collected at the systematic review, meta-analysis and trial level. Two reviewers applied and agreed on the GRADE rating for the selected meta-analyses.

From the 510 systematic reviews initially identified 91 reviews (41 Cochrane and 50 non-Cochrane) were eligible for inclusion. The quality of evidence was high/moderate in only 20% of meta-analyses with no difference between Cochrane and non-Cochrane reviews, journal impact factor or year of publication. The most common domains prompting downgrading of the evidence were study limitations and imprecision.

The quality of the evidence in oral health, which likely to be representative of other biomedical fields, assessed using GRADE is predominantly low or very low suggesting a pressing need for more studies of higher quality in order to inform clinical decisions thereby reducing the risk of instituting potentially ineffective and/or harmful therapies.

Περίληψη στα Ελληνικά

1. Αποτελεσματικότητα στην μετακίνηση των οδόντων με αυτόδετα και συμβατικά ορθοδοντικά αγκύλια: συστηματική ανασκόπηση και μετα-ανάλυση δικτύων (NMA).

Συστηματικές ανασκοπήσεις που συμπεριλαμβάνουν υψηλού επιπέδου κλινικές μελέτες αποτελούν τη βάση της επιστημονικά τεκμηριωμένης οδοντιατρικής και ιατρικής και βοηθούν σημαντικά στην σωστή λήψη αποφάσεων στην κλινική πράξη. Η συστηματική ανασκόπηση και Μέτα-ανάλυση διευκολύνει την ενσωμάτωση των υφιστάμενων αποδεικτικών στοιχείων χρησιμοποιώντας μια διαφανή και συστηματική προσέγγιση που οδηγεί σε μια εγκυρότερη ερμηνεία της αποτελεσματικότητας και της ασφάλειας μιας θεραπείας από ό, τι μπορεί να επιτευχθεί από μεμονωμένες μελέτες. Παραδοσιακά στις Μέτα-αναλύσεις συγκρίνονται μόνο δύο παρεμβάσεις ταυτόχρονα και δεν είναι δυνατός ο συνδυασμός πολλαπλών παρεμβάσεων. Μια σχετικά πρόσφατη εξέλι

ξη στη μετά-ανάλυση επιτρέπει, υπό ορισμένες συνθήκες, το συνδυασμό της άμεσης και έμμεσης σύγκρισης των διαφόρων παρεμβάσεων από κλινικές δοκιμές που χρησιμοποιούν μια κοινή ομάδα σύγκρισης, οδηγώντας σε μείωση της απώλειας πληροφοριών κατά τον υπολογισμό των συγκεντρωτικών εκτιμήσεων. Αυτό το είδος της μετά-ανάλυσης έχει ονομαστεί μετά-ανάλυση δικτύων (NMA). Η εφαρμογή της NMA επιτρέπει την ιεράρχηση των διαφόρων παρεμβάσεων, ακόμη και αν δεν υπάρχουν άμεσες συγκρίσεις μεταξύ των παρεμβάσεων, αξιοποιώντας την μεταβατικότητα (transitivity) των θεραπειών, εφόσον ικανοποιούνται οι προϋποθέσεις που απαιτούνται.

Στο πρώτο μέρος μελετήθηκε η αποτελεσματικότητα στην μετακίνηση των οδόντων με αυτόδετα και συμβατικά ορθοδοντικά αγκύλια ακολουθώντας μία NMA προσέγγιση. Δέκα κλινικές δοκιμές συμπεριελήφθησαν στην και τα αποτελέσματα NMA δείχνουν ότι οι συμβατικές συσκευές έχουν καλύτερες επιδόσεις σε σύγκριση με όλα τα άλλα συστήματα με μεγαλύτερη μέση βελτίωση κατά 0.03, 0.08 και 0.17 χιλιοστά / μήνα με τα συμβατικά αγκύλια σε σχέση με τα In-Ovation-R, Damon και

Smart-clip αγκύλια, αντίστοιχα. Οι εκτιμώμενες διαφορές δεν είναι στατιστικά σημαντικές και το πιο σημαντικό, οι εκτιμήσεις είναι μικρής κλινική σημασίας. Τα αποτελέσματα πρέπει να ερμηνευτούν με προσοχή, δεδομένου ότι ο αριθμός των μελετών είναι μικρός και τα διαστήματα εμπιστοσύνης είναι σχετικά μεγάλα.

2. Διαφορές στην αναφορά αποτελέσματος μεταξύ των πρωτοκόλλων και δημοσιευμένων συστηματικών ανασκοπήσεων της Cochrane Oral Health Group (COHG).

Πρακτικές όπως η επιλεκτική επιλογή κλινικών δοκιμών και επιλεκτική αναφορά αποτελεσμάτων μπορεί να οδηγήσει σε λανθασμένες εκτιμήσεις των αποτελεσμάτων μιας θεραπείας με πιθανές βλαβερές συνέπειες για τους ασθενείς.

Ενώ η έρευνα για την επιλεκτική αναφορά έχει επικεντρωθεί τόσο σε τυχαιοποιημένες και μη κλινικές μελέτες παρόμοια προβλήματα μπορεί να προκύψουν και με τις συστηματικές ανασκοπήσεις.

Στο δεύτερο μέρος αναλύονται αποκλίσεις στις εκβάσεις μεταξύ των πρωτοκόλλων και δημοσιευμένων συστηματικών ανασκοπήσεων της Cochrane Oral Health Group.

Όλες οι συστηματικές ανασκοπήσεις της COHG και τα αντίστοιχα πρωτόκολλα ανακτήθηκαν και πληροφορίες σχετικά με τις εκβάσεις που αναφέρθηκαν καταγράφηκαν. Τα δεδομένα ανεξάρτητα συγκεντρώθηκαν από δύο κριτές.

Εκατόν πενήντα δύο συστηματικές ανασκοπήσεις συμπεριελήφθησαν. Σε σχέση με τις κύριες εκβάσεις, το 11.2% είχαν υποβαθμιστεί σε μη κύριες εκβάσεις, το 9.9% είχαν παραλειφθεί εντελώς στην τελική δημοσίευση και εισαγωγή νέων κυρίων εκβάσεων εντοπίστηκε στο 18.4% των δημοσιεύσεων. Για τις μη κύριες εκβάσεις, 2.0% αναβαθμίστηκαν σε κύριες, 12.5% είχαν παραλειφθεί και 30.9% ευρέθησαν μόνο στη τελική δημοσίευση. Συνολικά, 45.4% των τελικών ανασκοπήσεων είχαν τουλάχιστον μία διαφορά σε σχέση με τους πρωτόκολλο – αναφορά της διαφοράς δόθηκε μόνον στο 14.5% εξ αυτών. Ο σχετικός κίνδυνος αναφοράς στατιστικά σημαντικών αποτελεσμάτων ήταν χαμηλότερος για υποβαθμισμένες [RR: 0.52, 95% CI: 0.17, 1.58, p = 0.24] και αναβαθμισμένες ή νεοεισαχθείσες εκβάσεις [RR: 0,77, 95% CI: 0,36, 1,64, p = 0.50] σε σχέση με εκβάσεις χωρίς αποκλίσεις. Ο σχετικός κίνδυνος αναφοράς στατιστικά σημαντικών αποτελεσμάτων ήταν υψηλότερος για

και αναβαθμισμένες ή νεοεισαχθείσες εκβάσεις σε σχέση με υποβαθμισμένες εκβάσεις (RR = 1,19, 95% CI: 0,65, 2,16, p = 0,57). Καμία από τις συσχετίσεις δεν ήταν στατιστικά σημαντική.

3. Η ποιότητα της επιστημονικής τεκμηρίωσης σύμφωνα με την προσέγγιση GRADE είναι κυρίως χαμηλή ή πολύ χαμηλή στις συστηματικές ανασκοπήσεις στο χώρο της Στοματικής Υγείας

Υπάρχει αυξανόμενη ανησυχία για το χάσμα μεταξύ έρευνας και κλινικής εφαρμογής. Έχει καταστεί προφανές ότι ένα σύστημα ικανό να επιτρέψει ταυτόχρονα την αξιολόγηση της ποιότητας των ερευνητικών αποτελεσμάτων, εξισορροπώντας οφέλη και κινδύνους, και λαμβάνοντας υπό όψιν τις προτιμήσεις του ασθενούς είναι αναγκαίο.

Αρκετές ομάδες έχουν προτείνει πολύπλοκες μεθόδους για την αξιολόγηση και μετάφραση των ερευνητικών αποτελεσμάτων στην κλινική πρακτική. Η πρωτοβουλία GRADE (Grades of Recommendation, Assessment, Development, and Evaluation), έχει καταστεί αποδεκτή προσέγγιση για την εκτίμηση των ερευνητικών αποτελεσμάτων και τη διατύπωση συστάσεων κλινικών οδηγιών.

Στο τελευταίο μέρος ο κύριος στόχος ήταν να εκτιμηθεί η αξιοπιστία των ερευνητικών αποτελεσμάτων σύμφωνα με το GRADE σε συστηματικές ανασκοπήσεις στη βάση δεδομένων Cochrane (CDSR-COHG) και αλλού.

Δεδομένα από συστηματικές ανασκοπήσεις (Ιανουάριος 2008 - Δεκέμβριος 2013) με Μέτα-αναλύσεις από τα 14 σημαντικότερα οδοντιατρικά περιοδικά και CDSR - COHG συλλέχθηκαν από δύο αξιολογητές.

Από τις διαθέσιμες 510 συστηματικές ανασκοπήσεις επελέγησαν 91 (41 Cochrane και 50 μη-Cochrane) για περαιτέρω εξαγωγή δεδομένων. Η ποιότητα των αποδεικτικών στοιχείων ήταν υψηλή / μέτρια μόνο στο 20% των Μέτα-αναλύσεων με καμία διαφορά μεταξύ Cochrane και μη Cochrane ανασκοπήσεων, impact factor περιοδικού ή έτος δημοσίευσης. Οι πιο συχνοί λόγοι υποβάθμιση των αποτελεσμάτων ήταν οι περιορισμοί στο σχεδιασμό της μελέτης (study limitations) και η χαμηλή ακρίβεια των αποτελεσμάτων (imprecision).

Reference List

1. Higgins JPT, Altman DG, Sterne JAC (editors): **Assessing risk of bias in included studies. Chapter 8**. In *In: 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.*; .

2. Cipriani A, Higgins JPT, Geddes JR, Salanti G: **Conceptual and technical challenges in network meta-analysis**. *Ann Intern Med* 2013, **159**:130–137.

3. Caldwell DM, Ades AE, Higgins JPT: **Simultaneous comparison of multiple treatments: combining direct and indirect evidence**. *BMJ* 2005, **331**:897–900.

4. Salanti G, Higgins JPT, Ades AE, Ioannidis JPA: **Evaluation of networks of randomized trials**. *Stat Methods Med Res* 2008, **17**:279–301.

5. Salanti G.: **Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool.** *Res Synth Method 2012;3:80-97* .

6. Lu G, Ades AE: **Combination of direct and indirect evidence in mixed treatment comparisons**. *Stat Med* 2004, **23**:3105–3124.

7. Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ: **Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation**. *Stat Med* 2009, **28**:1861–1881.

8. Song F, Altman DG, Glenny A-M, Deeks JJ: **Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses**. *BMJ* 2003, **326**:472.

9. Tu Y-K, Needleman I, Chambrone L, Lu H-K, Faggion CM: **A Bayesian network meta-analysis on comparisons of enamel matrix derivatives, guided tissue regeneration and their combination therapies**. *J Clin Periodontol* 2012, **39**:303–314.

10. Faggion CM, Chambrone L, Listl S, Tu Y-K: **Network meta-analysis for evaluating interventions in implant dentistry: the case of peri-implantitis treatment**. *Clin Implant Dent Relat Res* 2013, **15**:576–588.

11. Walsh T, Worthington HV, Glenny A-M, Appelbe P, Marinho VC, Shi X: **Fluoride toothpastes of different concentrations for preventing dental caries in children and adolescents**. *Cochrane Database Syst Rev* 2010:CD007868.

12. White IR: **Multivariate random-effects meta-analysis.** *The Stata Journal* 2009:40–56.

13. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR.: **Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies.** *Res Synth Method* 2012, **3**:98–110.

14. White IR, Barrett JK, Jackson D, Higgins JPT.: **Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression.** *Res Synth Method* 2012:111–25.

15. Chaimani A, Higgins JPT, Mavridis D, Spyridonos P, Salanti G: **Graphical tools for network meta-analysis in STATA**. *PloS One* 2013, **8**:e76654.

16. Fleming PS, Johal A: **Self-ligating brackets in orthodontics. A systematic review**. *Angle Orthod* 2010, **80**:575–584.

17. Miles PG: **SmartClip versus conventional twin brackets for initial alignment: is there a difference?**. *Aust Orthod J* 2005, **21**:123–127.

18. Miles PG, Weyant RJ, Rustveld L: **A clinical trial of Damon 2 vs conventional twin brackets during initial alignment**. *Angle Orthod* 2006, **76**:480–485.

19. Pandis N, Polychronopoulou A, Eliades T: **Self-ligating vs conventional brackets in the treatment of mandibular crowding: a prospective clinical trial of treatment duration and dental effects**. *Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod* 2007, **132**:208–215.

20. Ong E, McCallum H, Griffin MP, Ho C: **Efficiency of self-ligating vs conventionally ligated brackets during initial alignment**. *Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod* 2010, **138**:138.e1–7; discussion 138–139.

21. Baker SG, Kramer BS: **The transitive fallacy for randomized trials: if A bests B and B bests C in separate trials, is A better than C?**. *BMC Med Res Methodol* 2002, **2**:13.

22. Donegan S, Williamson P, D'Alessandro U, Smith CT: **Assessing key assumptions of network meta-analysis: a review of methods.** *Res Synth Method* 2013:291–323.

23. Jansen JP, Naci H: **Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers**. *BMC Med* 2013, **11**:159.

24. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE: **Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials**. *Med Decis Mak Int J Soc Med Decis Mak* 2013, **33**:641–656.

25. Bucher HC, Guyatt GH, Griffith LE, Walter SD: **The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials**. *J Clin Epidemiol* 1997, **50**:683–691.

26. Lu G, Ades AE.: **Assessing evidence inconsistency in mixed treatment comparisons.** *Journal of the Am Stat Assoc* 2006:447–459.

27. Dias S, Welton NJ, Caldwell DM, Ades AE: **Checking consistency in mixed treatment comparison meta-analysis**. *Stat Med* 2010, **29**:932–944.

28. Veroniki AA, Vasiliadis HS, Higgins JPT, Salanti G: **Evaluation of inconsistency in networks of interventions**. *Int J Epidemiol* 2013, **42**:332–345.

29. Donegan S, Williamson P, Gamble C, Tudur-Smith C: **Indirect comparisons: a review of reporting and methodological quality**. *PloS One* 2010, **5**:e11054.

30. Lumley T: **Network meta-analysis for indirect treatment comparisons**. *Stat Med* 2002, **21**:2313–2324.

31. Higgins JP, Whitehead A: **Borrowing strength from external trials in a meta-analysis**. *Stat Med* 1996, **15**:2733–2749.

32. Salanti G, Ades AE, Ioannidis JPA: **Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial**. *J Clin Epidemiol* 2011, **64**:163–171.

33. Higgins JPT, Thompson SG: **Controlling the risk of spurious findings from meta-regression**. *Stat Med* 2004, **23**:1663–1682.

34. Scott P, DiBiase AT, Sherriff M, Cobourne MT: **Alignment efficiency of Damon3 self-ligating and conventional orthodontic bracket systems: a randomized clinical trial**. *Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod* 2008, **134**:470.e1–8.

35. Fleming PS, DiBiase AT, Sarri G, Lee RT: **Efficiency of mandibular arch alignment with 2 preadjusted edgewise appliances**. *Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod* 2009, **135**:597–602.

36. Pandis N, Polychronopoulou A, Eliades T: **Active or passive self-ligating brackets? A randomized controlled trial of comparative efficiency in resolving maxillary anterior crowding in adolescents**. *Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod* 2010, **137**:12.e1–6; discussion 12–13.

37. Miles P, Weyant R: **Porcelain brackets during initial alignment: are self-ligating cosmetic brackets more efficient?**. *Aust Orthod J* 2010, **26**:21–26.

38. Pandis N, Polychronopoulou A, Katsaros C, Eliades T: **Comparative assessment of conventional and self-ligating appliances on the effect of mandibular intermolar distance in adolescent nonextraction patients: a single-center randomized controlled trial**. *Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod* 2011, **140**:e99–e105.

39. Wahab RMA, Idris H, Yacob H, Ariffin SHZ: **Comparison of self- and conventional-ligating brackets in the alignment stage**. *Eur J Orthod* 2012, **34**:176–181.

40. Gaspar-Ribeiro DA, deAlmeida MR, Conti AC, Navarro R, Oltramari-, Navarro P, Almeida R, Fernandes T.: **Efficiency of mandibular arch alignment with self-ligating and conventional edgewise appliances: a dental cast study**. *Dentistry* :128.

41. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F,, et al.,: **International Stroke Trial Collaborative Group; European Carotid Surgery Trial Collaborative Group. Evaluating nonrandomised intervention studies.** In *Health Technol Assess*; 2003:7: iii–iix,1–173.

42. Cobourne MT, Fleming PS, DiBiase AT, Ahmad S.: **Clinical cases in orthodontics.** Chichester, UK: Wiley-Blackwell; 2012.

43. Lee R, MacFarlane T, O'Brien K: **Consistency of orthodontic treatment planning decisions**. *Clin Orthod Res* 1999, **2**:79–84.

44. Ribarevski R, Vig P, Vig KD, Weyant R, O'Brien K: **Consistency of orthodontic extraction decisions**. *Eur J Orthod* 1996, **18**:77–80.

45. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group: **Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement**. *J Clin Epidemiol* 2009, **62**:1006–1012.

46. **Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. Applied Statistics 2000;49(3):359–70**. .

47. **Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al.Dissemination and publication of research findings: an updated review of related biases. Health Technology Assessment 2010;14(8):iii, ix-xi, 1-193.** .

48. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Ghersi D, Ioannidis JPA, Simes J, Williamson PR: **Systematic review of the empirical evidence of study publication bias and outcome reporting bias**. *PloS One* 2008, **3**:e3081.

49. Dwan K, Altman DG, Clarke M, Gamble C, Higgins JPT, Sterne JAC, Williamson PR, Kirkham JJ: **Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials**. *PLoS Med* 2014, **11**:e1001666.

50. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR: **The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews**. *BMJ* 2010, **340**:c365.

51. Chan A-W, Altman DG: **Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors**. *BMJ* 2005, **330**:753.

52. Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG: **Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles**. *JAMA J Am Med Assoc* 2004, **291**:2457–2465.

53. 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**. *Syst Rev* 2012, **1**:2.

54. **Norris SL, Moher D, Reeves BC, Shea B, Loke Y, Garner S, et al.Issues relating to selective reporting when including non-randomized studies in systematic reviews on the effects of healthcare interventions. Research Synthesis Methods 2013;4(1):36–47.** .

55. Page MJ, McKenzie JE, Forbes A: **Many scenarios exist for selective inclusion and reporting of results in randomized trials and systematic reviews**. *J Clin Epidemiol* 2013, **66**:524–537.

56. Tendal B, Higgins JPT, Jüni P, Hróbjartsson A, Trelle S, Nüesch E, Wandel S, Jørgensen AW, Gesser K, Ilsøe-Kristensen S, Gøtzsche PC: **Disagreements in meta-**

analyses using outcomes measured on continuous or rating scales: observer agreement study. *BMJ* 2009, **339**:b3128.

57. Bender R, Bunce C, Clarke M, Gates S, Lange S, Pace NL, Thorlund K: **Attention should be given to multiplicity issues in systematic reviews**. *J Clin Epidemiol* 2008, **61**:857–865.

58. Page MJ, McKenzie JE, Kirkham J, Dwan K, Kramer S, Green S, Forbes A: **Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions**. *Cochrane Database Syst Rev* 2014, **10**:MR000035.

59. Dwan K, Kirkham JJ, Williamson PR, Gamble C: **Selective reporting of outcomes in randomised controlled trials in systematic reviews of cystic fibrosis**. *BMJ Open* 2013, **3**.

60. Silagy CA, Middleton P, Hopewell S: **Publishing protocols of systematic reviews: comparing what was done to what was planned**. *JAMA* 2002, **287**:2831–2834.

61. **Parmelli E, Liberati A, D'Amico R. Reporting of outcomes in systematic reviews: comparison of protocols and published systematic reviews (abstract). XV Cochrane Colloquium; 2007 Oct 23-27; São Paulo, Brazil. 2007: 118–9.** .

62. Kirkham JJ, Altman DG, Williamson PR: **Bias due to changes in specified outcomes during the systematic review process**. *PloS One* 2010, **5**:e9810.

63. **www.systematicreviewsjournal.com**. .

64. Millett D: **Bias in systematic reviews?**. *J Orthod* 2011, **38**:158–160.

65. Guo J, Li C, Zhang Q, Wu G, Deacon SA, Chen J, Hu H, Zou S, Ye Q: **Secondary bone grafting for alveolar cleft in children with cleft lip or cleft lip and palate**. In *Cochrane Database of Systematic Reviews*. Edited by The Cochrane Collaboration. Chichester, UK: John Wiley & Sons, Ltd; 2011.

66. Yengopal V, Harnekar SY, Patel N, Siegfried N: **Dental fillings for the treatment of caries in the primary dentition**. In *Cochrane Database of Systematic Reviews*. Edited by The Cochrane Collaboration. Chichester, UK: John Wiley & Sons, Ltd; 2009.

67. Rasines Alcaraz MG, Veitz-Keenan A, Sahrmann P, Schmidlin PR, Davis D, Iheozor-Ejiofor Z: **Direct composite resin fillings versus amalgam fillings for permanent or adult posterior teeth**. In *Cochrane Database of Systematic Reviews*.

Edited by The Cochrane Collaboration. Chichester, UK: John Wiley & Sons, Ltd; 2014.

68. Coulthard P, Kushnerev E, Yates JM, Walsh T, Patel N, Bailey E, Renton TF: **Interventions for iatrogenic inferior alveolar and lingual nerve injury**. In *Cochrane Database of Systematic Reviews*. Edited by The Cochrane Collaboration. Chichester, UK: John Wiley & Sons, Ltd; 2014.

69. Daly B, Sharif MO, Newton T, Jones K, Worthington HV: **Local interventions for the management of alveolar osteitis (dry socket)**. *Cochrane Database Syst Rev* 2012, **12**:CD006968.

70. Coulthard P, Bailey E, Esposito M, Furness S, Renton TF, Worthington HV: **Surgical techniques for the removal of mandibular wisdom teeth**. In *Cochrane Database of Systematic Reviews*. Edited by The Cochrane Collaboration. Chichester, UK: John Wiley & Sons, Ltd; 2014.

71. Furness S, Glenny A-M, Worthington HV, Pavitt S, Oliver R, Clarkson JE, Macluskey M, Chan KK, Conway DI: **Interventions for the treatment of oral cavity and oropharyngeal cancer: chemotherapy**. *Cochrane Database Syst Rev* 2011:CD006386.

72. Glenny A-M, Furness S, Worthington HV, Conway DI, Oliver R, Clarkson JE, Macluskey M, Pavitt S, Chan KK, Brocklehurst P, CSROC Expert Panel: **Interventions for the treatment of oral cavity and oropharyngeal cancer: radiotherapy**. *Cochrane Database Syst Rev* 2010:CD006387.

73. Cheng S, Kirtschig G, Cooper S, Thornhill M, Leonardi-Bee J, Murphy R, others: **Interventions for erosive lichen planus affecting mucosal sites**. *Cochrane Database Syst Rev* 2012, **2**.

74. Ashley PF, Parekh S, Moles DR, Anand P, Behbehani A: **Preoperative analgesics for additional pain relief in children and adolescents having dental treatment**. *Cochrane Database Syst Rev* 2012, **9**:CD008392.

75. Seehra J, Fleming PS, Polychronopoulou A, Pandis N: **Reporting completeness of abstracts of systematic reviews published in leading dental specialty journals**. *Eur J Oral Sci* 2013, **121**:57–62.

76. Kiriakou J, Pandis N, Fleming PS, Madianos P, Polychronopoulou A: **Reporting quality of systematic review abstracts in leading oral implantology journals**. *J Dent* 2013, **41**:1181–1187.

77. Fleming PS, Seehra J, Polychronopoulou A, Fedorowicz Z, Pandis N: **A PRISMA assessment of the reporting quality of systematic reviews in orthodontics**. *Angle Orthod* 2013, **83**:158–163.

78. Papageorgiou SN, Papadopoulos MA, Athanasiou AE: **Reporting characteristics of meta-analyses in orthodontics: methodological assessment and statistical recommendations**. *Eur J Orthod* 2014, **36**:74–85.

79. Fleming PS, Seehra J, Polychronopoulou A, Fedorowicz Z, Pandis N: **Cochrane and non-Cochrane systematic reviews in leading orthodontic journals: a quality paradigm?**. *Eur J Orthod* 2013, **35**:244–248.

80. Remschmidt C, Wichmann O, Harder T: **Methodological quality of systematic reviews on influenza vaccination**. *Vaccine* 2014, **32**:1678–1684.

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

82. Martel G, Duhaime S, Barkun JS, Boushey RP, Ramsay CR, Fergusson DA: **The quality of research synthesis in surgery: the case of laparoscopic surgery for colorectal cancer**. *Syst Rev* 2012, **1**:14.

83. Brito JP, Tsapas A, Griebeler ML, Wang Z, Prutsky GJ, Domecq JP, Murad MH, Montori VM: **Systematic reviews supporting practice guideline recommendations lack protection against bias**. *J Clin Epidemiol* 2013, **66**:633–638.

84. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, Liberati A, O'Connell D, Oxman AD, Phillips B, Schünemann H, Edejer TT-T, Vist GE, Williams JW Jr, GRADE Working Group: **Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group**. *BMC Health Serv Res* 2004, **4**:38.

85. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schunemann HJ, for the GRADE Working Group: **GRADE: an emerging consensus on rating quality of evidence and strength of recommendations**. *BMJ* 2008, **336**:924–926.

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

87. Guyatt G, Oxman AD, Akl EA, 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.

88. Balshem H, Helfand M, Schünemann 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**. *J Clin Epidemiol* 2011, **64**:401–406.

89. **http://www.gradeworkinggroup.org/society/index.htm**. .

90. De Palma R, Liberati A, Ciccone G, Bandieri E, Belfiglio M, Ceccarelli M, Leoni M, Longo G, Magrini N, Marangolo M, Roila F, Programma Ricerca e Innovazione Emilia Romagna Oncology Research Group: **Developing clinical recommendations for breast, colorectal, and lung cancer adjuvant treatments using the GRADE system: a study from the Programma Ricerca e Innovazione Emilia Romagna Oncology Research Group**. *J Clin Oncol Off J Am Soc Clin Oncol* 2008, **26**:1033–1039.

91. Swiglo BA, Murad MH, Schünemann HJ, Kunz R, Vigersky RA, Guyatt GH, Montori VM: **A case for clarity, consistency, and helpfulness: state-of-the-art clinical practice guidelines in endocrinology using the grading of recommendations, assessment, development, and evaluation system**. *J Clin Endocrinol Metab* 2008, **93**:666–673.

92. Guyatt G, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B, Raskob G, Lewis SZ, Schünemann H: **Grading strength of recommendations and quality of evidence in clinical guidelines: report from an american college of chest physicians task force**. *Chest* 2006, **129**:174–181.

93. Brozek JL, Akl EA, Alonso-Coello P, Lang D, Jaeschke R, Williams JW, Phillips B, Lelgemann M, Lethaby A, Bousquet J, Guyatt GH, Schünemann HJ, GRADE Working Group: **Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions**. *Allergy* 2009, **64**:669–677.

94. Latthe P, Foon R, Khan K: **Nonsurgical treatment of stress urinary incontinence (SUI): grading of evidence in systematic reviews**. *BJOG Int J Obstet Gynaecol* 2008, **115**:435–444.

95. Murad MH, Altayar O, Bennett M, Wei JC, Claus PL, Asi N, Prokop LJ, Montori VM, Guyatt GH: **Using GRADE for evaluating the quality of evidence in hyperbaric oxygen therapy clarifies evidence limitations**. *J Clin Epidemiol* 2014, **67**:65–72.

96. Faggion CM Jr: **The shortened dental arch revisited: from evidence to recommendations by the use of the GRADE approach**. *J Oral Rehabil* 2011, **38**:940–949.

97. Faggion CM Jr: **Grading the quality of evidence and the strength of recommendations in clinical dentistry: a critical review of 2 prominent approaches**. *J Evid-Based Dent Pract* 2010, **10**:78–85.

98. 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, GRADE Working Group: **GRADE guidelines: 7. Rating the quality of evidence--inconsistency**. *J Clin Epidemiol* 2011, **64**:1294–1302.

99. 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, GRADE Working Group: **GRADE guidelines: 8. Rating the quality of evidence--indirectness**. *J Clin Epidemiol* 2011, **64**:1303–1310.

100. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux PJ, 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.

101. 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.

102. Eberhard J, Jervøe-Storm P-M, Needleman I, Worthington H, Jepsen S: **Full-mouth treatment concepts for chronic periodontitis: a systematic review**. *J Clin Periodontol* 2008, **35**:591–604.

103. Eberhard J, Jepsen S, Jervøe-Storm P-M, Needleman I, Worthington HV: **Full-mouth disinfection for the treatment of adult chronic periodontitis**. *Cochrane Database Syst Rev* 2008:CD004622.

104. Davey J, Turner RM, Clarke MJ, Higgins JPT: **Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: a cross-sectional, descriptive analysis**. *BMC Med Res Methodol* 2011, **11**:160.

105. Saltaji H, Cummings GG, Armijo-Olivo S, Major MP, Amin M, Major PW, Hartling L, Flores-Mir C: **A descriptive analysis of oral health systematic reviews published 1991-2012: cross sectional study**. *PloS One* 2013, **8**:e74545.

106. Fleming PS, Koletsi D, Polychronopoulou A, Eliades T, Pandis N: **Are clustering effects accounted for in statistical analysis in leading dental specialty journals?**. *J Dent* 2013, **41**:265–270.

107. Koletsi D, Pandis N, Polychronopoulou A, Eliades T: **Does published orthodontic research account for clustering effects during statistical data analysis?**. *Eur J Orthod* 2012, **34**:287–292.

108. Pandis N, Walsh T, Polychronopoulou A, Eliades T: **Cluster randomized clinical trials in orthodontics: design, analysis and reporting issues**. *Eur J Orthod* 2013, **35**:669–675.

109. Pandis N, Walsh T, Polychronopoulou A, Katsaros C, Eliades T: **Split-mouth designs in orthodontics: an overview with applications to orthodontic clinical trials**. *Eur J Orthod* 2013, **35**:783–789.

110. Mustafa RA, Santesso N, Brozek J, Akl EA, Walter SD, Norman G, Kulasegaram M, Christensen R, Guyatt GH, Falck-Ytter Y, Chang S, Murad MH, Vist GE, Lasserson T, Gartlehner G, Shukla V, Sun X, Whittington C, Post PN, Lang E, Thaler K, Kunnamo I, Alenius H, Meerpohl JJ, Alba AC, Nevis IF, Gentles S, Ethier M-C, Carrasco-Labra A, Khatib R, et al.: **The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses**. *J Clin Epidemiol* 2013, **66**:736–742.e5.

111. Hartling L, Fernandes RM, Seida J, Vandermeer B, Dryden DM: **From the Trenches: A Cross-Sectional Study Applying the GRADE Tool in Systematic Reviews of Healthcare Interventions**. *PLoS ONE* 2012, **7**:e34697.

112. Beller EM, Chen JK-H, Wang UL-H, Glasziou PP: **Are systematic reviews up-to-date at the time of publication?**. *Syst Rev* 2013, **2**:36.

113. Fleming PS, Koletsi D, Seehra J, Pandis N: **Systematic reviews published in higher impact clinical journals were of higher quality**. *J Clin Epidemiol* 2014.

114. *Sterne JAC, Egger M, Moher D (editors). Chapter 10: Addressing Reporting Biases. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1 [updated September 2008]. The Cochrane Collaboration, 2008. Available from Www.cochrane-Handbook.org. .*

115. Owens DK, Lohr KN, Atkins D, Treadwell JR, Reston JT, Bass EB, Chang S, Helfand M: **AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions--agency for healthcare research and quality and the effective health-care program**. *J Clin Epidemiol* 2010, **63**:513–523.

116. 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.

Appendix

	Collected Information per Systematic Review
Review level	**Journal title**
	Name of first author
	SR title
	Year of SR publication
	Specialty of journal
	doi
	Country of first author
	Continent of first author: Europe; Americas; Asia & other
	Number of authors
	Involvement of methodologist from affiliations, titles, methods-all Cochrane received a yes
	Involvement of one or more universities: SR conducted within a single university/setting or with authors from multiple universities
	Inclusion of meta-analysis in the SR or not
	Exclusion reasons of SR from full data extraction
	Interventional or not interventional SR
	GRADE table provided by SR authors or not
	List of SR outcomes includes treatment **benefits** or not
	List of SR outcomes includes treatment **harms** or not
	Number of non-randomized trials included in the SR
	Number of cases-series included in the SR
	Number of observational studies (cohort, case-control) included in the SR-cross-sectional excluded
	Number of parallel RCTs included in the SR
	Number of cross-over RCTs included in the SR
	Number of cluster RCTs included in the SR
	Number of split-mouth RCTs included in the SR
	Total number of studies included in SR
	Total number of meta-analyses included in SR

Meta-analysis level	**Number of trials** in selected meta-analysis
	Identification of meta-analysis selected for full data extraction
	Applied intervention in the selected meta-analysis
	Category of applied intervention in the selected meta-analysis
	Control therapy applied in the selected meta-analysis
	Category of applied control in the selected meta-analysis
	Outcome for selected meta-analysis
	Type of outcome: subjective or objective
	Outcome scale: continuous or binary
	Approach implemented by SR authors to assess methodological quality: i.e. Cochrane ROB; Jadad scale
	Overall risk of bias assessment at meta-analysis level: high; unclear; low: if one or more domains at high or unclear ROB declare as high or unclear ROB and low if all trials at low ROB
	Total sample size when sample size per group break down was not available
	Intervention mean effect per trial included in the selected meta-analysis
	Intervention standard deviation (sd) of the mean effect per trial included in the selected meta-analysis
	Intervention sample size per trial included in the selected meta-analysis
	Control mean effect per trial included in the selected meta-analysis
	Control standard deviation (sd) of the mean effect per trial included in the selected meta-analysis
	Control sample size per trial included in the selected meta-analysis
	Intervention events per trial included in the selected meta-analysis
	Intervention sample size per trial included in the selected meta-analysis
	Control events per trial included in the selected meta-analysis
	Control sample size per trial included in the selected meta-analysis
	Model used for meta-analysis: fixed; random
	Effect size value per trial included in the selected meta-analysis and pooled
	Effect size type: OR; RR ;HR; MD ; SMD
	Lower CI bound of effect size value per trial included in the selected meta-analysis and pooled
	Upper CI bound of effect size value per trial included in the selected meta-analysis and pooled

	I^2 **value** for statistical heterogeneity
	Accounting for paired data in the case of inclusion of split-mouth designs: not discussed or accounted for in the analysis; discussed and/or accounted for in the analysis
	Accounting for clustering effects in the case of inclusion of clustered designs: not discussed or accounted for in the analysis; discussed and/or accounted for in the analysis in the trials
Trial level	**Study design** of trials included in the selected meta-analysis: parallel; split-mouth; cross-over;
	Study type included in the meta-analysis: RCT; CCT ; case-series; cohort; case-control
	Author's name of trials included in the selected meta-analysis
	Allocation concealment risk of bias (ROB) assessment: high; unclear; low
	Blinding risk of bias assessment: high; unclear; low
	Attrition risk of bias assessment: high; unclear; low
	Selective outcome reporting risk of bias assessment: high; unclear; low
	Other limitations risk of bias assessment [stopping early for benefit; non- validated outcome measures; carry-over effect; recruitment bias in clustered trials] : high; unclear; low
	Overall risk of bias assessment at trial level: high; unclear; low: if one or more domains at unclear or high ROB declare as unclear or high ROB and low otherwise
GRADE assessment	**Study limitations:** If most trials at low ROB=no downgrade for limitations of design; If most trials at unclear ROB= downgrade for limitations of design by 1 level; If most trials at high ROB= downgrade for limitations of design by 2 levels.
	Inconsistency: 1 or 2 levels of downgrade depending on clinical & methodological heterogeneity (PICO), statistical heterogeneity, Confidence interval overlap
	Indirectness: 1 or 2 levels of downgrade depending on whether or not head-to-head comparisons were used
	Imprecision: 1 or 2 levels of downgrade depending on statistical heterogeneity, CI overlap and inclusion of benefits and harms in the CI
	Publication bias: suspected or unsuspected based on completeness of search strategy, formal statistical assessment and SR authors' comments
	Outcome Importance: not important; important but not critical; critical. Refers to selected outcome and based on importance relative to intervention. i.e. implant failure critical but clinical attachment loss may be important but not
	Effect size: if effect is strong applicable for observational studies RR>2 or <0.5; RR>5 or RR<0.2
	GRADE rating in this review: High, moderate, low, very low

	GRADE reported by SR authors (if provided): High, moderate, low, very low
RCTs=Randomized controlled trials; CCT=Controlled clinical trial implying non-randomized; SR=Systematic review; ROB=Risk of bias; CI=Confidence interval; OR=Odds ratio; Risk ratio ;Hazard ratio; Mean difference ; Standardized mean difference	

author	Study type	Study limitations	inconsistency	indirectness	Imprecision	Publication bias	GRADE	GRADE in SR
Chen[1]	Mixed	serious	no	no	serious	undetected	low	nr
Fleming[2]	Randomized	serious	no	no	no	undetected	high	high
Fleming[3]	Randomized	no	no	no	no	undetected	high	high
Kaklamanos[4]	Randomized	serious	no	no	serious	undetected	low	nr
Marsico[5]	Randomized	serious	no	no	no	undetected	moderate	nr
Santos[6]	Randomized	very serious	no	no	no	undetected	low	nr
Alsabeeha[7]	Mixed	very serious	no	no	very serious	undetected	very low	nr
Del Fabbro[8]	Mixed	very serious	no	no	no	undetected	moderate	nr
Del Fabbro [9]	Non-randomized	very serious	no	no	very serious	undetected	low	nr
Safii[10]	Non-randomized	very serious	no	no	serious	undetected	low	nr
Marathiotou[11]	Non-randomized	very serious	no	no	no	undetected	low	nr
Mickenautsch[12]	Non-randomized	very serious	no	no	serious	undetected	low	nr
Shabazfar[13]	Mixed	very serious	serious	no	serious	detected	very low	nr
Atieh[14]	Mixed	very serious	no	no	no	undetected	moderate	nr
Emami[15]	Randomized	serious	serious	no	serious	undetected	very low	nr
Rocuzzo[16]	Randomized	serious	no	no	serious	detected	low	nr
Sanz [17]	Mixed	very serious	no	no	very serious	undetected	very low	nr
Thoma[18]	Randomized	serious	serious	no	very serious	detected	very low	nr
Agarwal[19]	Randomized	serious	no	no	serious	undetected	low	nr
Ashley[20]	Randomized	serious	no	no	serious	undetected	low	moderate
Bessel[21]	Randomized	serious	no	no	serious	undetected	low	nr
Chambrone[22]	Randomized	very serious	serious	no	serious	undetected	very low	nr
Cooper[23]	Randomized	serious	no	no	serious	undetected	low	low
Daly[24]	Randomized	serious	no	no	no	undetected	moderate	moderate
Deacon[25]	Randomized	serious	no	no	no	undetected	moderate	nr
Eberhard[26]	Randomized	serious	no	no	no	undetected	moderate	nr
Esposito1[27]	Randomized	serious	no	no	very serious	undetected	low	nr
Esposito2[28]	Randomized	serious	serious	no	no	undetected	low	low
Esposito3[29]	Randomized	serious	no	no	serious	undetected	low	moderate
Esposito4[30]	Randomized	serious	no	no	serious	undetected	low	low
Esposito5[31]	Randomized	serious	no	no	serious	undetected	very low	nr
Esposito6[32]	Randomized	serious	no	no	very serious	undetected	very low	nr
Esposito7[33]	Randomized	very serious	no	no	very serious	undetected	very low	nr
Esposito8[34]	Randomized	serious	no	no	no	undetected	moderate	moderate
Furness1[35]	Randomized	serious	no	serious	no	undetected	very low	very low
Furness2[36]	Randomized	serious	no	no	serious	undetected	low	low
Furness3[37]	Randomized	very serious	no	serious	no	undetected	low	nr
Glenny1[38]	Randomized	serious	no	no	no	undetected	moderate	nr
Glenny2[39]	Randomized	very serious	serious	serious	no	undetected	low	low

Hu[40]	Randomized	serious	no	no	serious	undetected	low	low
Jambi[41]	Randomized	serious	serious	no	serious	undetected	very low	very low
Tubert-Jeannin[42]	Randomized	serious	no	no	no	undetected	moderate	nr
Lodi[43]	Randomized	very serious	no	no	no	undetected	moderate	moderate
Marihno[44]	Randomized	serious	no	no	no	undetected	moderate	moderate
Martharou[45]	Randomized	serious	serious	no	no	undetected	low	nr
Millett[46]	Randomized	no	serious	no	serious	undetected	low	nr
Nasser[47]	Randomized	serious	no	no	very serious	undetected	low	nr
Ricketts[48]	Randomized	serious	no	no	no	undetected	moderate	moderate
Rigon[49]	Randomized	very serious	no	no	serious	undetected	very low	nr
Ahovuo-Saloranta[50]	Randomized	no	no	no	no	undetected	moderate	moderate
Sambunjak[51]	Randomized	serious	serious	no	serious	undetected	very low	very low
Shi[52]	Randomized	serious	no	no	no	undetected	moderate	moderate
Simpson[53]	Randomized	serious	no	no	serious	undetected	low	low
Thiruvenkatachari[54]	Randomized	serious	serious	no	no	undetected	low	low
Thongprasom[55]	Randomized	very serious	no	no	serious	undetected	low	nr
Watkinson[56]	Randomized	serious	serious	no	no	undetected	low	low
Worthington[57]	Randomized	very serious	serious	no	no	undetected	very low	nr
Worthington[58]	Randomized	very serious	no	no	serious	undetected	very low	nr
Worthington[59]	Randomized	serious	no	no	serious	undetected	low	low
Atieh[60]	Randomized	serious	no	no	serious	undetected	low	nr
Annibali[61]	Randomized	very serious	serious	no	no	undetected	low	nr
Bouziane[62]	Randomized	very serious	serious	no	no	detected	very low	nr
Cairo[63]	Randomized	very serious	no	no	no	undetected	low	nr
Chambrone[64]	Randomized	serious	no	no	serious	detected	low	nr
Eberhard[65]	Randomized	serious	no	no	very serious	undetected	low	nr
Kunnen[66]	Randomized	serious	no	no	serious	undetected	low	nr
Sgolastra[67]	Randomized	serious	serious	no	no	undetected	low	nr
He[68]	Randomized	serious	no	no	serious	undetected	low	nr
Jung[69]	Randomized	serious	no	no	serious	undetected	low	nr
Katyal[70]	Randomized	serious	no	no	no	undetected	moderate	nr
Schwendicke[71]	Randomized	very serious	serious	no	very serious	undetected	very low	very low
Cruz[72]	Randomized	very serious	very serious	no	serious	undetected	low	nr
Long[73]	Randomized	very serious	no	no	no	undetected	low	nr
Schwendicke[74]	Randomized	very serious	no	no	no	undetected	moderate	moderate
Stoecklin-Wasmer[75]	Randomized	very serious	serious	no	no	detected	very low	nr
Gillen[76]	Not reported	serious	serious	no	no	undetected	moderate	nr
Su[77]	Randomized	serious	no	no	serious	undetected	low	nr
Tsesis[78]	Randomized	serious	no	no	very serious	detected	very low	nr
Dan[79]	Mixed	serious	no	no	no	undetected	moderate	nr

Katsnelson[80]	Non-randomized	very serious	no	no	serious	undetected	very low	nr
Carrasco-labra[81]	Randomized	serious	serious	no	serious	undetected	very low	very low
Li[82]	Randomized	serious	serious	no	serious	undetected	low	nr
Brignardello-Petersen[83]	Randomized	very serious	no	no	serious	undetected	very low	nr
Yu[84]	Randomized	serious	no	no	serious	undetected	low	nr
atieh[85]	Mixed	very serious	no	no	very serious	detected	very low	nr
darby[86]	Mixed	no	no	no	serious	undetected	moderate	nr
koop[87]	Randomized	very serious	serious	no	serious	undetected	very low	nr
kotsovilis[88]	Not reported	serious	serious	no	no	undetected	low	nr
lin[89]	Randomized	serious	no	no	no	undetected	moderate	nr
sgolastra[90]	Randomized	very serious	no	no	serious	undetected	very low	nr
sohrabi[91]	Randomized	serious	no	no	serious	undetected	low	nr

References GRADE

1. Chen SS-H, Greenlee GM, Kim J-E, Smith CL, Huang GJ. Systematic review of self-ligating brackets. Am J Orthod Dentofacial Orthop. 2010;137: 726.e1–726.e18. doi:10.1016/j.ajodo.2009.11.009

2. Fleming PS, Johal A, Pandis N. Self-etch primers and conventional acid-etch technique for orthodontic bonding: A systematic review and meta-analysis. Am J Orthod Dentofacial Orthop. 2012;142: 83–94. doi:10.1016/j.ajodo.2012.02.023

3. Padhraig S. Fleming, Eliades T, Katsaros C, Pandis N. Curing lights for orthodontic bonding: A systematic review and meta-analysis. Am J Orthod Dentofacial Orthop. 2013;143: S92–S103. doi:10.1016/j.ajodo.2012.07.018

4. Kaklamanos EG, Kalfas S. Meta-analysis on the effectiveness of powered toothbrushes for orthodontic patients. Am J Orthod Dentofacial Orthop. 2008;133: 187.e1–187.e14. doi:10.1016/j.ajodo.2007.07.015

5. Marsico E, Gatto E, Burrascano M, Matarese G, Cordasco G. Effectiveness of orthodontic treatment with functional appliances on mandibular growth in the short term. Am J Orthod Dentofacial Orthop. 2011;139: 24–36. doi:10.1016/j.ajodo.2010.04.028

6. Santos APP, Oliveira BH, Nadanovsky P. Effects of Low and Standard Fluoride Toothpastes on Caries and Fluorosis: Systematic Review and Meta-Analysis. Caries Res. 2013;47: 382–390. doi:10.1159/000348492

7. Alsabeeha N, Atieh M, Payne AGT. Loading Protocols for Mandibular Implant Overdentures: A Systematic Review with Meta-Analysis: Mandibular Overdenture Review and Meta-Analysis. Clin Implant Dent Relat Res. 2009;12: e28–e38. doi:10.1111/j.1708-8208.2009.00152.x

8. Del Fabbro M, Ceresoli V, Taschieri S, Ceci C, Testori T. Immediate Loading of Postextraction Implants in the Esthetic Area: Systematic Review of the Literature: Immediate Implant Placement and Restoration. Clin Implant Dent Relat Res. 2013; n/a–n/a. doi:10.1111/cid.12074

9. Del Fabbro M, Bellini CM, Romeo D, Francetti L. Tilted Implants for the Rehabilitation of Edentulous Jaws: A Systematic Review: Literature Review of Tilted Implants. Clin Implant Dent Relat Res. 2012;14: 612–621. doi:10.1111/j.1708-8208.2010.00288.x

10. Safii SH, Palmer RM, Wilson RF. Risk of Implant Failure and Marginal Bone Loss in Subjects with a History of Periodontitis: A Systematic Review and Meta-Analysis. Clin Implant Dent Relat Res. 2009; doi:10.1111/j.1708-8208.2009.00162.x

11. Ioannidou-Marathiotou I, Zafeiriadis AA, Papadopoulos MA. Root resorption of endodontically treated teeth following orthodontic treatment: a meta-analysis. Clin Oral Investig. 2013;17: 1733–1744. doi:10.1007/s00784-012-0860-8

12. Mickenautsch S, Yengopal V, Banerjee A. Retention of orthodontic brackets bonded with resin-modified GIC versus composite resin adhesives—a quantitative systematic review of clinical trials. Clin Oral Investig. 2012;16: 1–14. doi:10.1007/s00784-011-0626-8

13. Shabazfar N, Daubländer M, Al-Nawas B, Kämmerer PW. Periodontal intraligament injection as alternative to inferior alveolar nerve block—meta-analysis of the literature from 1979 to 2012. Clin Oral Investig. 2014;18: 351–358. doi:10.1007/s00784-013-1113-1

14. Atieh MA, Payne AGT, Duncan WJ, Cullinan MP. Immediate restoration/loading of immediately placed single implants: is it an effective bimodal approach? Clin Oral Implants Res. 2009;20: 645–659. doi:10.1111/j.1600-0501.2009.01725.x

15. Emami E, Heydecke G, Rompré PH, de Grandmont P, Feine JS. Impact of implant support for mandibular dentures on satisfaction, oral and general health-related quality of life: a meta-analysis of randomized-controlled trials. Clin Oral Implants Res. 2009; doi:10.1111/j.1600-0501.2008.01693.x

16. Roccuzzo M, Bonino F, Gaudioso L, Zwahlen M, Meijer HJA. What is the optimal number of implants for removable reconstructions? A systematic review on implant-supported overdentures. Clin Oral Implants Res. 2012;23: 229–237. doi:10.1111/j.1600-0501.2012.02544.x

17. Sanz I, Garcia-Gargallo M, Herrera D, Martin C, Figuero E, Sanz M. Surgical protocols for early implant placement in post-extraction sockets: a systematic review. Clin Oral Implants Res. 2012;23: 67–79. doi:10.1111/j.1600-0501.2011.02339.x

18. Thoma DS, Benić GI, Zwahlen M, Hämmerle CHF, Jung RE. A systematic review assessing soft tissue augmentation techniques. Clin Oral Implants Res. 2009;20: 146–165. doi:10.1111/j.1600-0501.2009.01784.x

19. Aggarwal VR, Lovell K, Peters S, Javidi H, Joughin A, Goldthorpe J. Psychosocial interventions for the management of chronic orofacial pain. Cochrane Database Syst Rev. 2011; CD008456. doi:10.1002/14651858.CD008456.pub2

20. Ashley PF, Parekh S, Moles DR, Anand P, Behbehani A. Preoperative analgesics for additional pain relief in children and adolescents having dental treatment. Cochrane Database Syst Rev. 2012;9: CD008392. doi:10.1002/14651858.CD008392.pub2

21. Bessell A, Hooper L, Shaw WC, Reilly S, Reid J, Glenny A-M. Feeding interventions for growth and development in infants with cleft lip, cleft palate or cleft lip and palate. Cochrane Database Syst Rev. 2011; CD003315. doi:10.1002/14651858.CD003315.pub3

22. Chambrone L, Sukekava F, Araújo MG, Pustiglioni FE, Chambrone LA, Lima LA. Root coverage procedures for the treatment of localised recession-type defects. Cochrane Database Syst Rev. 2009; CD007161. doi:10.1002/14651858.CD007161.pub2

23. Cooper AM, O'Malley LA, Elison SN, Armstrong R, Burnside G, Adair P, et al. Primary school-based behavioural interventions for preventing caries. Cochrane Database Syst Rev. 2013;5: CD009378. doi:10.1002/14651858.CD009378.pub2

24. Daly B, Sharif MO, Newton T, Jones K, Worthington HV. Local interventions for the management of alveolar osteitis (dry socket). Cochrane Database Syst Rev. 2012;12: CD006968. doi:10.1002/14651858.CD006968.pub2

25. Deacon SA, Glenny A-M, Deery C, Robinson PG, Heanue M, Walmsley AD, et al. Different powered toothbrushes for plaque control and gingival health. Cochrane Database Syst Rev. 2010; CD004971. doi:10.1002/14651858.CD004971.pub2

26. Eberhard J, Jepsen S, Jervøe-Storm P-M, Needleman I, Worthington HV. Full-mouth disinfection for the treatment of adult chronic periodontitis. Cochrane Database Syst Rev. 2008; CD004622. doi:10.1002/14651858.CD004622.pub2

27. Esposito M, Grusovin MG, Felice P, Karatzopoulos G, Worthington HV, Coulthard P. Interventions for replacing missing teeth: horizontal and vertical bone augmentation techniques for dental implant treatment. Cochrane Database Syst Rev. 2009; CD003607. doi:10.1002/14651858.CD003607.pub4

28. Esposito M, Grusovin MG, Papanikolaou N, Coulthard P, Worthington HV. Enamel matrix derivative (Emdogain(R)) for periodontal tissue regeneration in intrabony defects. Cochrane Database Syst Rev. 2009; CD003875. doi:10.1002/14651858.CD003875.pub3

29. Esposito M, Grusovin MG, Chew YS, Coulthard P, Worthington HV. Interventions for replacing missing teeth: 1- versus 2-stage implant placement. Cochrane Database Syst Rev. 2009; CD006698. doi:10.1002/14651858.CD006698.pub2

30. Esposito M, Grusovin MG, Rees J, Karasoulos D, Felice P, Alissa R, et al. Interventions for replacing missing teeth: augmentation procedures of the maxillary sinus. Cochrane Database Syst Rev. 2010; CD008397. doi:10.1002/14651858.CD008397

31. Esposito M, Grusovin MG, Polyzos IP, Felice P, Worthington HV. Interventions for replacing missing teeth: dental implants in fresh extraction sockets (immediate, immediate-delayed and delayed implants). Cochrane Database Syst Rev. 2010; CD005968. doi:10.1002/14651858.CD005968.pub3

32. Esposito M, Maghaireh H, Grusovin MG, Ziounas I, Worthington HV. Interventions for replacing missing teeth: management of soft tissues for dental implants. Cochrane Database Syst Rev. 2012;2: CD006697. doi:10.1002/14651858.CD006697.pub2

33. Esposito M, Grusovin MG, Worthington HV. Interventions for replacing missing teeth: treatment of peri-implantitis. Cochrane Database Syst Rev. 2012;1: CD004970. doi:10.1002/14651858.CD004970.pub5

34. Esposito M, Grusovin MG, Worthington HV. Interventions for replacing missing teeth: antibiotics at dental implant placement to prevent complications. Cochrane Database Syst Rev. 2013;7: CD004152. doi:10.1002/14651858.CD004152.pub4

35. Furness S, Glenny A-M, Worthington HV, Pavitt S, Oliver R, Clarkson JE, et al. Interventions for the treatment of oral cavity and oropharyngeal cancer: chemotherapy. Cochrane Database Syst Rev. 2011; CD006386. doi:10.1002/14651858.CD006386.pub3

36. Furness S, Bryan G, McMillan R, Birchenough S, Worthington HV. Interventions for the management of dry mouth: non-pharmacological interventions. Cochrane Database Syst Rev. 2013;9: CD009603. doi:10.1002/14651858.CD009603.pub3

37. Furness S, Worthington HV, Bryan G, Birchenough S, McMillan R. Interventions for the management of dry mouth: topical therapies. Cochrane Database Syst Rev. 2011; CD008934. doi:10.1002/14651858.CD008934.pub2

38. Glenny A-M, Fernandez Mauleffinch LM, Pavitt S, Walsh T. Interventions for the prevention and treatment of herpes simplex virus in patients being treated for cancer. Cochrane Database Syst Rev. 2009; CD006706. doi:10.1002/14651858.CD006706.pub2

39. Glenny A-M, Furness S, Worthington HV, Conway DI, Oliver R, Clarkson JE, et al. Interventions for the treatment of oral cavity and oropharyngeal cancer:

radiotherapy. Cochrane Database Syst Rev. 2010; CD006387. doi:10.1002/14651858.CD006387.pub2

40. Hu H, Li C, Li F, Chen J, Sun J, Zou S, et al. Enamel etching for bonding fixed orthodontic braces. Cochrane Database Syst Rev. 2013;11: CD005516. doi:10.1002/14651858.CD005516.pub2

41. Jambi S, Thiruvenkatachari B, O'Brien KD, Walsh T. Orthodontic treatment for distalising upper first molars in children and adolescents. Cochrane Database Syst Rev. 2013;10: CD008375. doi:10.1002/14651858.CD008375.pub2

42. Tubert-Jeannin S, Auclair C, Amsallem E, Tramini P, Gerbaud L, Ruffieux C, et al. Fluoride supplements (tablets, drops, lozenges or chewing gums) for preventing dental caries in children. Cochrane Database Syst Rev. 2011; CD007592. doi:10.1002/14651858.CD007592.pub2

43. Lodi G, Figini L, Sardella A, Carrassi A, Del Fabbro M, Furness S. Antibiotics to prevent complications following tooth extractions. Cochrane Database Syst Rev. 2012;11: CD003811. doi:10.1002/14651858.CD003811.pub2

44. Marinho VCC, Worthington HV, Walsh T, Clarkson JE. Fluoride varnishes for preventing dental caries in children and adolescents. Cochrane Database Syst Rev. 2013;7: CD002279. doi:10.1002/14651858.CD002279.pub2

45. Lourenço-Matharu L, Ashley PF, Furness S. Sedation of children undergoing dental treatment. Cochrane Database Syst Rev. 2012;3: CD003877. doi:10.1002/14651858.CD003877.pub4

46. Millett DT, Mandall NA, Mattick RC, Hickman J, Glenny A-M. Adhesives for bonded molar tubes during fixed brace treatment. Cochrane Database Syst Rev. 2011; CD008236. doi:10.1002/14651858.CD008236.pub2

47. Nasser M, Pandis N, Fleming PS, Fedorowicz Z, Ellis E, Ali K. Interventions for the management of mandibular fractures. Cochrane Database Syst Rev. 2013;7: CD006087. doi:10.1002/14651858.CD006087.pub3

48. Ricketts D, Lamont T, Innes NPT, Kidd E, Clarkson JE. Operative caries management in adults and children. Cochrane Database Syst Rev. 2013;3: CD003808. doi:10.1002/14651858.CD003808.pub3

49. Rigon M, Pereira LM, Bortoluzzi MC, Loguercio AD, Ramos AL, Cardoso JR. Arthroscopy for temporomandibular disorders. Cochrane Database Syst Rev. 2011; CD006385. doi:10.1002/14651858.CD006385.pub2

50. Ahovuo-Saloranta A, Forss H, Walsh T, Hiiri A, Nordblad A, Mäkelä M, et al. Sealants for preventing dental decay in the permanent teeth. Cochrane Database Syst Rev. 2013;3: CD001830. doi:10.1002/14651858.CD001830.pub4

51. Sambunjak D, Nickerson JW, Poklepovic T, Johnson TM, Imai P, Tugwell P, et al. Flossing for the management of periodontal diseases and dental caries in adults. Cochrane Database Syst Rev. 2011; CD008829. doi:10.1002/14651858.CD008829.pub2

52. Shi Z, Xie H, Wang P, Zhang Q, Wu Y, Chen E, et al. Oral hygiene care for critically ill patients to prevent ventilator-associated pneumonia. Cochrane Database Syst Rev. 2013;8: CD008367. doi:10.1002/14651858.CD008367.pub2

53. Simpson TC, Needleman I, Wild SH, Moles DR, Mills EJ. Treatment of periodontal disease for glycaemic control in people with diabetes. Cochrane Database Syst Rev. 2010; CD004714. doi:10.1002/14651858.CD004714.pub2

54. Thiruvenkatachari B, Harrison JE, Worthington HV, O'Brien KD. Orthodontic treatment for prominent upper front teeth (Class II malocclusion) in children. Status Date Ed No Change Conclus Publ In. 2013; Available: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD003452.pub3/pdf/standard

55. Thongprasom K, Carrozzo M, Furness S, Lodi G. Interventions for treating oral lichen planus. Cochrane Database Syst Rev. 2011; CD001168. doi:10.1002/14651858.CD001168.pub2

56. Watkinson S, Harrison JE, Furness S, Worthington HV. Orthodontic treatment for prominent lower front teeth (Class III malocclusion) in children. Cochrane Database Syst Rev. 2013;9: CD003451. doi:10.1002/14651858.CD003451.pub2

57. Worthington HV, Clarkson JE, Khalid T, Meyer S, McCabe M. Interventions for treating oral candidiasis for patients with cancer receiving treatment. Cochrane Database Syst Rev. 2010; CD001972. doi:10.1002/14651858.CD001972.pub4

58. Worthington HV, Clarkson JE, Bryan G, Furness S, Glenny A-M, Littlewood A, et al. Interventions for preventing oral mucositis for patients with cancer receiving treatment. Cochrane Database Syst Rev. 2011; CD000978. doi:10.1002/14651858.CD000978.pub5

59. Worthington HV, Clarkson JE, Bryan G, Beirne PV. Routine scale and polish for periodontal health in adults. Cochrane Database Syst Rev. 2013;11: CD004625. doi:10.1002/14651858.CD004625.pub4

60. Atieh MA, Atieh AH, Payne AGT, Duncan WJ. Immediate loading with single implant crowns: a systematic review and meta-analysis. Int J Prosthodont. 2009;22: 378–387.

61. Annibali S, Bignozzi I, Cristalli MP, Graziani F, La Monaca G, Polimeni A. Peri-implant marginal bone level: a systematic review and meta-analysis of studies comparing platform switching *versus* conventionally restored implants. J Clin Periodontol. 2012;39: 1097–1113. doi:10.1111/j.1600-051X.2012.01930.x

62. Bouziane A, Ahid S, Abouqal R, Ennibi O. Effect of periodontal therapy on prevention of gastric *Helicobacter pylori* recurrence: a systematic review and meta-analysis. J Clin Periodontol. 2012;39: 1166–1173. doi:10.1111/jcpe.12015

63. Cairo F, Pagliaro U, Nieri M. Treatment of gingival recession with coronally advanced flap procedures: a systematic review. J Clin Periodontol. 2008;35: 136–162. doi:10.1111/j.1600-051X.2008.01267.x

64. Chambrone L, Pannuti CM, Guglielmetti MR, Chambrone LA. Evidence grade associating periodontitis with preterm birth and/or low birth weight: II. A systematic review of randomized trials evaluating the effects of periodontal treatment: Periodontitis and adverse pregnancy outcomes. J Clin Periodontol. 2011;38: 902–914. doi:10.1111/j.1600-051X.2011.01761.x

65. Eberhard J, Jervøe-Storm P-M, Needleman I, Worthington H, Jepsen S. Full-mouth treatment concepts for chronic periodontitis: a systematic review. J Clin Periodontol. 2008;35: 591–604. doi:10.1111/j.1600-051X.2008.01239.x

66. Kunnen A, Van Doormaal JJ, Abbas F, Aarnoudse JG, Van Pampus MG, Faas MM. Review Article: Periodontal disease and pre-eclampsia: a systematic review: Periodontal disease and pre-eclampsia. J Clin Periodontol. 2010;37: 1075–1087. doi:10.1111/j.1600-051X.2010.01636.x

67. Sgolastra F, Petrucci A, Severino M, Graziani F, Gatto R, Monaco A. Adjunctive photodynamic therapy to non-surgical treatment of chronic periodontitis: a systematic review and meta-analysis. J Clin Periodontol. 2013;40: 514–526. doi:10.1111/jcpe.12094

68. He L-B, Shao M-Y, Tan K, Xu X, Li J-Y. The effects of light on bleaching and tooth sensitivity during in-office vital bleaching: A systematic review and meta-analysis. J Dent. 2012;40: 644–653. doi:10.1016/j.jdent.2012.04.010

69. Jung A, Shin B-C, Lee MS, Sim H, Ernst E. Acupuncture for treating temporomandibular joint disorders: A systematic review and meta-analysis

of randomized, sham-controlled trials. J Dent. 2011;39: 341–350. doi:10.1016/j.jdent.2011.02.006

70. Katyal V. The efficacy and safety of articaine versus lignocaine in dental treatments: A meta-analysis. J Dent. 2010;38: 307–317. doi:10.1016/j.jdent.2009.12.003

71. Schwendicke F, Meyer-Lueckel H, Dörfer C, Paris S. Failure of incompletely excavated teeth—A systematic review. J Dent. 2013;41: 569–580. doi:10.1016/j.jdent.2013.05.004

72. Cunha-Cruz J, Stout JR, Heaton LJ, Wataha JC, for Northwest PRECEDENT. Dentin Hypersensitivity and Oxalates: a Systematic Review. J Dent Res. 2011;90: 304–310. doi:10.1177/0022034510389179

73. Long H, Zhou Y, Liao L, Pyakurel U, Wang Y, Lai W. Coronectomy vs. Total Removal for Third Molar Extraction: A Systematic Review. J Dent Res. 2012;91: 659–665. doi:10.1177/0022034512449346

74. Schwendicke F, Dorfer CE, Paris S. Incomplete Caries Removal: A Systematic Review and Meta-analysis. J Dent Res. 2013;92: 306–314. doi:10.1177/0022034513477425

75. Stoecklin-Wasmer C, Rutjes AWS, da Costa BR, Salvi GE, Juni P, Sculean A. Absorbable Collagen Membranes for Periodontal Regeneration: A Systematic Review. J Dent Res. 2013;92: 773–781. doi:10.1177/0022034513496428

76. Gillen BM, Looney SW, Gu L-S, Loushine BA, Weller RN, Loushine RJ, et al. Impact of the Quality of Coronal Restoration versus the Quality of Root Canal Fillings on Success of Root Canal Treatment: A Systematic Review and Meta-analysis. J Endod. 2011;37: 895–902. doi:10.1016/j.joen.2011.04.002

77. Su Y, Wang C, Ye L. Healing Rate and Post-obturation Pain of Single- versus Multiple-visit Endodontic Treatment for Infected Root Canals: A Systematic Review. J Endod. 2011;37: 125–132. doi:10.1016/j.joen.2010.09.005

78. Tsesis I, Rosen E, Tamse A, Taschieri S, Del Fabbro M. Effect of Guided Tissue Regeneration on the Outcome of Surgical Endodontic Treatment: A Systematic Review and Meta-analysis. J Endod. 2011;37: 1039–1045. doi:10.1016/j.joen.2011.05.016

79. Dan AEB, Thygesen TH, Pinholt EM. Corticosteroid Administration in Oral and Orthognathic Surgery: A Systematic Review of the Literature and Meta-Analysis. J Oral Maxillofac Surg. 2010;68: 2207–2220. doi:10.1016/j.joms.2010.04.019

80. Katsnelson A, Markiewicz MR, Keith DA, Dodson TB. Operative Management of Temporomandibular Joint Ankylosis: A Systematic Review and Meta-Analysis. J Oral Maxillofac Surg. 2012;70: 531–536. doi:10.1016/j.joms.2011.10.003

81. Carrasco-Labra A, Brignardello-Petersen R, Yanine N, Araya I, Guyatt G. Secondary Versus Primary Closure Techniques for the Prevention of Postoperative Complications Following Removal of Impacted Mandibular Third Molars: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. J Oral Maxillofac Surg. 2012;70: e441–e457. doi:10.1016/j.joms.2012.03.017

82. Li C, Zhang Y, Lv J, Shi Z. Inferior or Double Joint Spaces Injection Versus Superior Joint Space Injection for Temporomandibular Disorders: A Systematic Review and Meta-Analysis. J Oral Maxillofac Surg. 2012;70: 37–44. doi:10.1016/j.joms.2011.04.009

83. Brignardello-Petersen R, Carrasco-Labra A, Araya I, Yanine N, Beyene J, Shah PS. Is Adjuvant Laser Therapy Effective for Preventing Pain, Swelling, and Trismus After Surgical Removal of Impacted Mandibular Third Molars? A Systematic Review and Meta-Analysis. J Oral Maxillofac Surg. 2012;70: 1789–1801. doi:10.1016/j.joms.2012.01.008

84. Yu SH, Beirne OR. Laryngeal Mask Airways Have a Lower Risk of Airway Complications Compared With Endotracheal Intubation: A Systematic Review. J Oral Maxillofac Surg. 2010;68: 2359–2376. doi:10.1016/j.joms.2010.04.017

85. Atieh MA, Alsabeeha NHM, Faggion CM, Duncan WJ. The Frequency of Peri-Implant Diseases: A Systematic Review and Meta-Analysis. J Periodontol. 2012; 1–15. doi:10.1902/jop.2012.120592

86. Darby IB, Morris KH. A Systematic Review of the Use of Growth Factors in Human Periodontal Regeneration. J Periodontol. 2013;84: 465–476. doi:10.1902/jop.2012.120145

87. Koop R, Merheb J, Quirynen M. Periodontal Regeneration With Enamel Matrix Derivative in Reconstructive Periodontal Therapy: A Systematic Review. J Periodontol. 2012;83: 707–720. doi:10.1902/jop.2011.110266

88. Kotsovilis S, Fourmousis I, Karoussis IK, Bamia C. A Systematic Review and Meta-Analysis on the Effect of Implant Length on the Survival of Rough-Surface Dental Implants. J Periodontol. 2009;80: 1700–1718. doi:10.1902/jop.2009.090107

89. Lin G-H, Chan H-L, Bashutski JD, Oh T-J, Wang H-L. The Effect of Flapless Surgery on Implant Survival and Marginal Bone Level: A Systematic Review and Meta-Analysis. J Periodontol. 2014;85: e91–e103. doi:10.1902/jop.2013.130481

90. Sgolastra F, Petrucci A, Gatto R, Giannoni M, Monaco A. Long-Term Efficacy of Subantimicrobial-Dose Doxycycline as an Adjunctive Treatment to Scaling and Root Planing: A Systematic Review and Meta-Analysis. J Periodontol. 2011;82: 1570–1581. doi:10.1902/jop.2011.110026

91. Sohrabi K, Saraiya V, Laage TA, Harris M, Blieden M, Karimbux N. An Evaluation of Bioactive Glass in the Treatment of Periodontal Defects: A Meta-Analysis of Randomized Controlled Clinical Trials. J Periodontol. 2012;83: 453–464. doi:10.1902/jop.2011.110347

92. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008;336: 601–605. doi:10.1136/bmj.39465.451748.AD

Detailed explanation for assessing the level of evidence using the GRADE approach [1–5]

Limitations in study design (Risk of bias) was assessed in the following domains: allocation concealment, blinding (participants and assessors), especially in the presence of subjective outcomes, attrition, selective outcome reporting and other limitations (Stopping early for benefit, use of non-validated outcome measures such as patient reported outcomes, carry-over effect in cross-over designs and recruitment bias in cluster randomized designs). The assessments for the study limitations in the above domains was be based on the Cochrane Collaborations' risk of bias tool recommendations)[6] . A rating of high, unclear or low risk of bias was assigned per domain, per trial overall and across all trials included in the selected met-analysis. Overall trial rating as high, unclear or low was done as follows: if one or more domains at unclear or high ROB declare as unclear or high ROB respectively, otherwise rating is low risk of bias. Across studies, a summary assessment is rated as low risk of bias when most information is from studies at low risk of bias, unclear risk of bias when most information (amount of information: sample size, number of events etc and not the number of studies) is from studies at low or unclear risk of bias, and high risk of bias when the proportion of information is from studies at high risk of bias sufficient to affect the interpretation of the results. Consequently, for low overall risk of bias no downgrade was implemented and for serious or very serious a 1 and 2 levels of downgrading was applied respectively.

Inconsistency: In the presence of heterogeneity or variability of results across included studies in the assessed meta-analysis, failure of plausible explanation for the observed heterogeneity is a reason for downgrading for inconsistency. Study methods and associated bias, variability of point estimates, confidence interval overlap and statistical criteria such as heterogeneity tests (p-value for heterogeneity and I^2) were considered during the assessment. A rough guideline, however not restrictive, given the effect of sample size on the value of I^2, would be: I^2<40% is low, 30-60% may be moderate, 50-90% may be substantial, and 75-100% is regarded as considerable[7]. Minor difference in direction with overlapping confidence intervals

will not result in a downgrade. A 3-point rating was utilized corresponding to no inconsistency, serious inconsistency, very serious inconsistency.

Indirectness: Reasons for a change in the rating of indirectness included no head-to-head comparisons and evidence derived from studies with different participants, interventions and outcomes to the question being addressed by the SR. A 3-point rating was be utilized corresponding to no indirectness, serious indirectness, very serious indirectness. Surrogate outcomes generally result in a 1-point downgrade, whereas additional issues in the other areas (participants, interventions, indirect comparisons) resulted in a 2- point downgrade.

Imprecision: Studies of small size and limited number of events result in imprecise estimates with wide confidence intervals. The optimal information size (OIS) criterion suggest that if the total number of patients included in the systematic review is less than the number of patients generated by a conventional sample calculation for a single adequately powered trial, a possible downgrade due to the likelihood of imprecision should be considered. The OIS guidelines according to GRADE for potentially downgrading the quality of the evidence are discussed below for binary and continuous outcomes:

<u>Dichotomous outcomes</u>

-total sample size is lower than the calculated optimal information size (OIS), unless sample size is very large (at least 2000 or perhaps 4000 patients).

-If the 95% CI excludes a relative risk of 1 and the total number of events exceeds the OIS, precision criterion is considered adequate. If the 95% confidence intervals ranges from unimportant to important benefits and from unimportant to important harms, GRADE recommends downgrading the quality of evidence when appreciable benefit or harm are of 25% relative risk reduction or relative risk increase, respectively.

-An exception exists if event rates are very low. In this case 95% CIs around relative effects can be very wide; however, 95% CIs around absolute effects may be narrow. In this scenario the quality of evidence may not be downgraded for precision. Additionally, if the number of patients is sufficiently large, it is probable that

prognostic balance has been achieved and downgrading based on imprecision is not appropriate. The inference of unimportance should be considered over the appropriate follow-up duration, since a small number of events over a short follow-up may be misleading.

-If the sample size is large enough and there is an apparent treatment effect with a satisfactory CI there is no need to apply the OIS.

-SR quality of the evidence was not downgraded based on the balance between desirable and undesirable effects; this decision belongs to the guideline developers. Therefore, precision judgment should not be focused on the threshold indicating the clinical importance of the effect but on OIS. If OIS is not met, a downgrade is warranted unless the sample size is very large. If the criterion is met and the 95% CI around the effect excludes 1 there is no need to downgrade for imprecision

Continuous outcomes

-95% CI includes no effect and upper and lower CI bounds cross the minimal important clinical difference (MID) for harm or benefit

-if the MID is not known or use of different outcome measures is required for calculation of an effect size, GRADE recommends downgrading if the upper or lower confidence limit crosses an effect size of 0.5 in either direction.

-use OIS

Again a 3-point scale was utilized corresponding to no imprecision, serious imprecision, and very serious imprecision. Grading for imprecision was outlined earlier in detail.

Publication bias: Inclusion of small studies only in combination with industry sponsorship resulted in a downgrade. Publication bias was assessed by examining search methods and funnel plot asymmetry (if available). Industry sponsorship or suspicion of industry sponsorship and/or other conflicts of interest may also increase the likelihood of downgrade. Additional parameters that may influence the decision to downgrade include inclusion of small studies only, a relatively recent RCT or a number of RCTs assessing a new treatment and a less than comprehensive search for

studies; for example, failing to search for unpublished trials. A 2-point rating was assigned representing either undetected or strongly suspected publication bias. Absence of publication bias is difficult to confirm; it is similarly difficult to decide on a threshold at which to downgrade in view of publication bias. In view of these challenges, GRADE suggests use of the terms ‘‘undetected’’ and ‘‘strongly suspected’’ to describe the risk of publication bias. Given the uncertainty in the presence or absence of publication bias, GRADE suggests rating down a maximum of one level (rather than two) for suspicion of publication bias. For Cochrane reviews there is a trial search coordinator and therefore we assumed that the search was sufficient. We examined the publication bias section and downgraded if the authors stated suspicion on publication bias. For non-Cochrane reviews we considered searched databases, gray literature, reference hand search and author’s statements in the review in order to decide on downgrading on this **domain.**

References

[1] Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64:401–6. doi:10.1016/j.jclinepi.2010.07.015.

[2] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. J Clin Epidemiol 2011;64:1294–302. doi:10.1016/j.jclinepi.2011.03.017.

[3] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol 2011;64:1303–10. doi:10.1016/j.jclinepi.2011.04.014.

[4] Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. J Clin Epidemiol 2011;64:1277–82. doi:10.1016/j.jclinepi.2011.01.011.

[5] Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol 2011;64:1283–93. doi:10.1016/j.jclinepi.2011.01.012.

[6] Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.

[7] Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 9: Analysing data and undertaking meta-analyses. In: 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, n.d.