The impact of bias on the magnitude of treatment effect estimates in oral health randomized trials

by

Humam Hisham Saltaji

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Medical Sciences - Orthodontics
University of Alberta

Abstract

Background: There is emerging evidence that randomized trials are subject to biases. Flaws in the design of such trials can result in over- or underestimation of the treatment effect size.

Aim: To examine the empirical evidence for bias, to quantify the extent of bias associated with methodology, and to make recommendations for reducing bias in oral health randomized trials.

Methods and Analyses: The study was conducted in four interconnected phases. First, to develop a register of oral health systematic reviews, 1188 reviews published in the domain of oral health research were identified and described. In the second phase, a methodology study of 1114 therapeutic reviews (published between 1991 and 2014) was constructed using descriptive and multivariate logistic regression analyses to explore (a) how often and by what means risk of bias in trials had been assessed, and (b) factors associated with completed risk of bias assessments. In the third phase, from the register of reviews, all meta-analyses were selected that examined at least one continuous outcome and included a minimum of five oral health randomized trials (n = 64); this package consisted of 540 randomized trials that analyzed 137,957 patients. The risk of bias was examined with respect to the reporting and methodological characteristics of these trials in order to assess the state of oral health trials over time. In the fourth phase, using a two-level meta-meta-analytic approach with a random effects model to allow for intra- and intermeta-analysis heterogeneity,

the impact of (a) bias associated with 22 methodological characteristics, and (b) specific features (such as dental specialty and type of outcome) on the magnitude of treatment effect estimates in oral health randomized trials was evaluated.

Results: Risk of bias assessment of primary studies had not been made in a considerable portion of 1114 therapeutic reviews published between 1991 and 2014 (61.4%, n = 684). This occurred more often in reviews published after dissemination of the PRISMA statement (odds ratio = 1.55; 95% confidence interval [CI]: 1.17 to 2.06), and in reviews published in nondental journals (odds ratio = 0.28; 95% CI: 0.19 to 0.41). The results of the risk of bias and quality assessments were unfavorable in general, indicating substandard quality and high potential for bias in oral health trials. The proportion of trials judged as having a low risk of bias did not exceed 60% in the majority of the risk of bias domains, but this proportion has significantly increased over time. In the 540 oral health randomized trials examined, significantly larger treatment effect size estimates were identified in trials that had inadequate sequence generation (difference in treatment effect size estimates = 0.13; 95% CI: 0.01 to 0.25), inadequate allocation concealment (difference in treatment effect size estimates = 0.15; 95% CI: 0.02 to 0.27), lack of patient and assessor blinding (difference in treatment effect size estimates = 0.19; 95% CI: 0.06 to 0.32), inadequate reporting of the dropout rate (difference in treatment effect size estimates = 0.24; 95% CI: 0.05 to 0.43), and inappropriate influence of funders (difference in treatment effect size estimates = 0.10; 95% CI: 0.02 to 0.19). Although not statistically significant, a

tendency toward exaggeration of the treatment effect size was found in the presence of imbalance in cointervention, inadequate compliance to treatment, incomplete outcome data, or having dropout without performing the intention-to-treat approach. In contrast, baseline imbalance, caregiver blinding, an acceptable dropout rate (≤ 20%), selective outcome reporting, and analysis based on an intention-to-treat approach, were not associated with inflated or underestimated treatment effect size.

Conclusion: Bias was found to be, overall, associated with inflated treatment effect size in oral health randomized trials. Therefore, systematic reviewers would be advised to exclude trials conducted in the domains of dental, oral, and craniofacial research from meta-analyses or to perform sensitivity analyses based on the adequacy of these criteria. Because of the impact of bias on treatment effect size estimates, dental journal editors and reviewers should insist on adequate conduct and reporting of trial reports submitted for publication.

Dedication

This thesis is dedicated to my mother Faiha Alkhani, and to the memory of my father Hisham Saltaji (1931-2013).

Acknowledgements

The journey to earning a combined PhD–Residency in Orthodontics at the University of Alberta was not taken alone. It was supported with a collaborative effort of mentors, advisors, friends, and family.

Dr. Carlos Flores-Mir's guidance and direction and Dr. Greta G. Cummings' insight and encouragement were indelible. I thank Dr. Maryam Amin for consistent support and advice and Dr. Susan Armijo-Olivo for useful feedback and the ample time she invested in this work. Dr. Paul W. Major, the Chair of Dentistry at the University of Alberta, was a consistent supporter.

Many friends and colleagues helped along the way. A special thank you is due to Hussam Alfakir and Mostafa Altalibi, who were always there for me. Additional thanks go to Khaled Sharaf, Saleh Aldaghreer, Sabrina Lopresti, Michael Major, Ida Kornerup, Jose Roberto Pereira, Camila Pacheco-Pereira, Long Tieu, Genvieve Lemieux Hart, Natasha Nazarali, Connie Ling, Hazem Eimar, Kristopher Currie, Mohamed ElSalhy, Maryam Elyasi, Mohammed Al-Saleh, Aishah Alsumait, Ashraf Abd Elmeguid, Narjara Moreira, Samira Diar Bakirly, Nikki McGill, Susan Helwig, Heather Good, and Pat Lapointe.

The unconditional love of my wife Sarah, my lovely sisters Safa and Marwa, and my brothers, Anwar, Hashim, Ahmad, Abdallah, and Safwan made the journey all the more worthwhile.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AHRQ	Agency for Healthcare Research and Quality
ADA	American Dental Association
CCT	controlled clinical trial
CI	confidence interval
CONSORT	Consolidated Standards of Reporting Trials
CRs	Cochrane reviews
ES	effect size
ICMJE	International Committee of Medical Journal Editors
IQR	interquartile range
GRADE	Grading of Recommendations, Assessment, Development and Evaluation
MA	meta-analysis
NCRs	non-Cochrane reviews
OR	odds ratio
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT	randomized controlled trial
RoB	risk of bias
SR	systematic review
SD	standard deviation
WHO	World Health Organization

Chapter 1

Introduction

1.1. Literature Review

1.1.1. Assessing risk of bias of trials within systematic reviews

Clinical research continuously generates new scientific evidence that contributes to improved oral health care. Therapeutic The "evidence-based practice" approach is formed based on several levels of evidence, which range, for therapeutic interventions, from greatest to least, as follows: systematic reviews of randomized controlled clinical trials, randomized controlled clinical trials, systematic reviews of cohort studies, cohort studies, systematic reviews of case-control studies, case-control studies, case series, and consensus opinion of experts [1, 2]. Nevertheless, answers to research questions are sometimes better obtained by using evidence from well-conducted case-controlled or cohort studies than through biased randomized trials [3].

A systematic review serves to identify, appraise, and integrate the findings of studies of a specific topic using a systematic approach [4, 5]; systematic reviews have become the gold standard for decision-making by clinicians and policy makers, and is foundational to evidence-based practice [6]. With the growth of evidence-based practice in dentistry, the number of published systematic reviews conducted in dental fields has rapidly increased [7]. One of the valuable sources for systematic reviews is the Cochrane Collaboration, an international organization that aims to help health care professionals make well-informed decisions about treatment interventions by conducting high-quality systematic reviews. Published methodological studies have suggested that systematic reviews produced by this group are usually better than non-

Cochrane systematic reviews with respect to methodological characteristics and reporting quality [8-10].

Current scientific knowledge for clinical research should be based on randomized controlled trials that have been synthesized in systematic reviews with meta-analyses, which together comprise the "gold standard" of scientific evidence [11, 12]. Systematic reviews use a comprehensive search strategy to identify potentially relevant trials; they predefine eligibility criteria to minimize the impact of bias in study selection and synthesize the results based on the quality of the evidence from individual trials [12]. As with any research design, the value of a systematic review depends on how well it is conducted and reported. Guidelines such as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [13] has resulted in improvements in both reporting and methodological qualities of published reviews [14].

In the area of oral health, approximately 50 dentistry-related trials are published per month, and the numbers are increasing every year [15]. Similarly, the number of systematic reviews published in oral health and within dental specialties has steadily grown over the last two decades [7]. However, it is suggested that the reporting and methodological aspects of systematic reviews in oral health do not meet acceptable levels of scientific research, and risk of bias assessments for primary studies are a particular weakness [16-18].

The interpretation and use of findings from a systematic review of an intervention rely heavily on the scope and internal validity of included studies; the internal validity is largely determined by the extent to which the design, conduct, and analysis of the trials follow the highest possible standards to minimize multiple biases and thus ensure that the findings can be attributed to the intervention [12, 19].

For this reason, it is essential to critically appraise the risk of bias—a critical component of overall methodological quality—of trials included in systematic reviews of therapeutic interventions [20]. Numerous tools exist to assess risk of bias of randomized clinical trials; however, few have undergone extensive testing for reliability or validity [21, 22]. While some are labelled as "scales" where quality items are summed, suggesting a total quality score (either by giving similar weights to each quality item or by putting more emphasis on specific items), others are labelled as "checklists" where quality items are scored separately [21]. Because the majority of these instruments have not been assessed for their measurement properties, whether and to what extent they tap the construct of risk of bias in ways that discriminate between trials with biased and trials with unbiased results, is unknown. For example, Herbison et al., argued that using a quality score in meta-analysis adjustments does not always adequately discriminate between high- and low-quality trials, and that each quality tool assessment may support a different conclusion, even though the primary evidence is the same [23].

The Jadad scale, developed to assess pharmacological trials, is one of the most popular quality assessment tools. It has been psychometrically evaluated and uses three items—randomization, double-blinding, and a description of dropouts and withdrawals—to assess the internal validity of randomized trials [24]. However, double-blinding accounts for 40% of the total score. The high value placed on double-blinding makes the Jadad scale less useful in oral health trials involving surgical or device interventions where patient blinding is not feasible. Additionally, allocation concealment is not included in this tool, despite its known relationship with the internal validity of randomized clinical trials [25-27]. Therefore, the Jadad scale has been reported to have inadequate responsiveness to differentiate between various quality grades [28].

The Consolidated Standards of Reporting Trials (CONSORT) statement, published in 1996, and updated in 2010 [29], is an accepted and widely used approach to assess the reporting quality of randomized trials in medical and dental research. The tool was developed to advance the transparency and the quality of reporting by creating reporting criteria [4, 46], which consists of 22-items that cover fundamental aspects of a trial's reporting quality. In the last 10 years, the CONSORT statement has been endorsed by several medical journals worldwide, including the majority of the high-impact oral health journals [29-31].

To overcome the limitations of the existing quality assessment tools, the Cochrane Collaboration developed and introduced the Cochrane risk of bias tool in 2008 [12]. This tool was designed to appraise risk of bias based on six quality domains related to the internal validity of a trial [19], namely: "sequence generation," "allocation concealment," "blinding of outcome assessors," "blinding of participants," "incomplete outcome data," "selective outcome reporting," and "other sources of bias." Since its inception, the Cochrane risk of bias tool has consistantly evolved and, while it has played a key role in leading improvements in risk of bias assessment in health research, it requires further development and improvement [32]. The risk of bias assessment of individual studies in a systematic review is a fundamental component of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach [33] which is used to evaluate the strength of a body of evidence. Importantly, research into the synthesis of dental knowledge would benefit from a clear definition of terms pertaining to the evaluation of the risk of bias (i.e., a determination of the extent to which the study results are close to the truth), the methodological quality (i.e., the extent to which the conduct of the research reaches the highest possible standards), and the quality of the evidence (i.e., the extent to which there is confidence that an estimate of the treatment effect size is near to the true value of an outcome across many trials) [12].

1.1.2. Methodologic quality and risk of bias in dental randomized trials

The value and significance of a randomized trial depend on how potential biases are controlled and how well the trial is conducted and reported. The endorsement and implemention of reporting guidelines, such as the CONSORT statement [34, 35] and other recent initiatives such as the GRADE approach [36] and the International Committee of Medical Journal Editors' (ICMJE) statement on clinical trial registration [37, 38], have led to an improvement in the "quality of evidence," including the methodological and reporting quality of medical randomized clinical trials [39-41]. This is critical to oral health research and practice because high-quality randomized controlled trials contribute largely to the body of evidence measured in systematic reviews and meta-analyses (see **Appendices 1.A** and **1.B** for further details on trial registration in dentistry).

While randomized controlled clinical trials are often referred to as the optimal type of research to examine the effectiveness of treatment interventions in health sciences [19], emerging evidence from methodological reports published in various fields of oral health over the last decade (periodontics [42], prosthodontics [43], implantology [44], orthodontics [45], dentistry [46]) suggests that the reporting quality and the methodological quality of oral heath randomized trials are below acceptable levels to adequately inform clinical decision making. This raises questions about the validity of evidence stemming from oral health randomized trials; such evidence is broadly used by policy makers when developing clinical practice guidelines, and by dental practitioners when making clinical decisions in dental practice. Similar results have been reported in medicine by numerous investigators, who found that the quality of medical randomized trials was not optimal [29, 47].

Numerous investigations across dental specialties have found that randomized controlled trials were too poorly reported to support clinical decision making.

Montenegro et al. [42] assessed the methodologic quality of 177 periodontology randomized trials using four methodological characteristics: randomization, concealment allocation, blinding, and follow-up of patients. The results showed that the quality of the randomized trials was suboptimal. For example, adequacy of randomization, allocation concealment, and blinding were identified, respectively, in only 17%, 7%, and 17% of the 177 trials.

Similar results were found when assessing the quality of randomized trials in the field of orthodontics. Harrison (2003) [45] examined the reporting quality of 155 orthodontic randomized trials published from 1989–1998 using four characteristics: randomization, concealment allocation, double-blinding, and description of dropouts. The results indicated that 54.8% of the trials were reported to be randomized, 2.6% were reported to be allocation concealed, 6.5% reported blinding of patients and assessors, and 28.4% reported the dropout rate.

In prosthodontics and implantology, Nieri et al. [43] assessed reporting and statistical characteristics of 45 randomized controlled clinical trials in the area of implant therapy. The authors concluded that the methodological quality and the statistical quality of the selected randomized clinical trials was inadequate to establish conclusions that could inform dental clinicians making treatment decisions in dental practices.

Pandis et al. (2010) [46] assessed the reporting quality of 95 randomized clinical trials published in six dental specialty journals using the CONSORT statement. They concluded that the reporting quality of selected randomized trials was poor and not sufficient to guide dental clinicians in their practices. In addition, they identified a positive association between the quality of oral health randomized trials and the number of authors, as well as the involvement of a data analyst in the trial.

Vere and Joshi (2010) [44] reported similar findings when examining the methodologic quality of 38 randomized trials in the field of dental implantology (2004–2008) using 14 methodologic items. The results showed that 42% of randomized clinical trials were adequately randomized, 31% reported blinding of the outcome assessor, 18% were adequately concealed, and 18% assessed inter-rater and intra-rater reliability.

While no study in the field of dentistry has assessed the quality, the risk of bias improvement, and the variation in oral health randomized trials over time, a recent report by Reveiz et al. [48] examined the risk of bias improvement in a sample of medical randomized trials identified from a cohort of Cochrane reviews. The report stated that the number of trials with a low risk of bias had consistently increased with time. However Dechartres et al. [49] recently evaluated a cohort of methodological reports that assessed the reporting and methodological quality of randomized trials and indicated that the reviews employed different approaches to assess the methodologic quality and the risk of bias.

1.1.3. Associations between trial quality and treatment effect size estimate

The importance of randomized trials as potential building blocks for policies and clinical decisions makes it imperative to monitor the quality of such trials [29]. Unfortunately, drawbacks in the conduct and/or reporting in randomized trials have been shown to affect the estimation of treatment effect size [50].

Individual concepts of methodological conduct and reporting quality of oral health clinical trials have been assessed in several investigations [51-54]. Assessments of the methodological conduct and reporting quality of oral health clinical trials have examined sample size justification [55], reported statistical findings [56], clustering effects [57], and the randomization process [58, 59], among other factors. Despite these efforts, gaps in the assessment process, such as the failure to identify design

flaws that can impact treatment effect size estimates, still exist within the domain of oral health research [60-62].

Methodological research published in the last decade that quantified bias in treatment effect size estimates in a series of meta-analyses, found associations between the adequacy of methodological characteristics and the bias in randomized trials. Specific methodologic characteristics associated with bias in estimates of treatment effect size have been identified by examining meta-analyses gathered in "meta-epidemiological studies" [63]. Such studies indicate that the conduct of randomized trials may impact the reported treatment effect size estimate [64], and suggest that trials with low methodological quality tend to exaggerate estimates of treatment effect size [65]. For example, the treatment effect size was overestimated 11–51% [66] in cases of inadequate randomization, and 10–52% [66] in cases of inadequate allocation concealment. Other investigations have indicated that randomized trials with lack of blinding exaggerate estimates of treatment effect size compared with adequately blinded randomized trials [4]. This is concerning because of the potential impact of bias on patient care and clinical decisions. In contrast, these associations were not confirmed in two studies [67, 68].

Moher et al. [69] examined the methodological quality of 127 randomized trials included in 11 meta-analyses with binary outcomes (conducted in the fields of mental, digestive, and circulatory diseases) using the Jadad scale and Schulz's allocation concealment tool. They reported a 34% increase in the treatment effect size estimate in randomized trials with low methodological quality compared to randomized trials with high methodological quality. In addition, inadequate allocation concealment was associated with an increase in the treatment effect size estimate of 37% compared to adequately concealed allocation in randomized trials [57].

Kjaergard et al. [66] assessed methodological quality in 190 randomized trials included in 14 meta-analyses with binary outcomes using four quality items: randomization, allocation concealment, double-blinding, and adequacy of follow-up. The study reported that randomized trials with low methodological quality, including inadequate sequence generation, inadequate allocation concealment, and inadequate double-blinding were associated with exaggerated treatment effect size estimates.

In contrast, Balk et al. [67] used 24 methodologic items to examine 256 randomized clinical trials with binary outcomes included in 26 meta-analyses in numerous medical domains (surgery, pediatrics, cardiovascular disease, and infectious diseases). The study found no significant differences in treatment effect size estimates across the examined medical domains.

Egger et al. [4] reviewed the impact of bias associated with allocation concealment and double-blinding in 304 randomized clinical trials included in 39 meta-analyses with binary outcomes in several medical fields (infectious diseases, neurology, among others). The study reported significant associations between adequate allocation concealment and smaller treatment effect size estimates.

Kunz et al. [70] used a list of methodological criteria to compare treatment effect size estimates between high-quality and low-quality conduct of randomized trials. The study showed that low-quality conduct was associated with an increase of 35–50% in the treatment effect size estimate compared to high-quality conduct, and that inadequate allocation concealment was associated with a 35–40% increase in the treatment effect size estimate compared to adequately concealed allocation in the randomized clinical trials.

Pildal et al. [71] assessed the adequacy of allocation concealment and compared findings from 38 meta-analyses after including only randomized clinical trials that had adequate allocation concealment. The study concluded that if only

randomized trials with adequate allocation concealment were included in the meta-analyses conducted with the reviews, about two-thirds of the conclusions lost statistical significance.

Tulder et al. [65] measured the associations between 11 methodologic items (recommended by the Cochrane Back Review Group) and treatment effect size estimates in 216 randomized trials included in 15 Cochrane reviews in the field of back pain. The study found significantly smaller treatment effect size estimates in randomized trials with high quality methodology compared to randomized trials with low quality methodology.

Hartling et al. [72] examined the risk of bias in 287 randomized trials included in 17 meta-analyses conducted in the field of child health, using the Cochrane risk of bias assessment tool. The study reported no significant association in any of the domains between risk of bias and estimated treatment effect size. The study concluded that the small sample size of included meta-analyses and randomized trials assessed in the investigation were responsible for the insignificant association between risk of bias and treatment effect size estimates.

Two studies in the field of dentistry examined the influence of bias on treatment effect size estimates in randomized trials using the trial as the level of analysis. Fenwick et al. [61] conducted a "pilot study" to review the bias associated with inadequate allocation concealment and inadequate examiner blinding in 34 randomized trials in the field of periodontology. They found no significant association between the treatment effect size estimate and inadequate allocation concealment or inadequate blinding. Koletsi et al. [73] found that inadequate sequence generation and incomplete outcome reporting (for binary outcomes) were associated with inflated treatment effect size estimates in 101 orthodontic randomized trials.

1.2. Rationale

Ideally, randomized clinical trials should be conducted with thoughtful methodology and accurate reporting in order to reach well-supported conclusions for decision making that are both valid and generalizable to patients who will receive the interventions in clinical practice [4, 15]. However, evidence is emerging that some randomized trials are biased due to flaws in their design and/or reported study characteristics and thus overestimate the magnitude of the treatment effect [15, 16]. These studies can skew the overall conclusions of meta-analyses once pooled, and thus lead to faulty treatment decisions [2, 17]. Generally, poor methodological quality in randomized clinical trials leads to an overestimation of the treatment effect size [18]. Flawed characteristics observed in previously published reports that were found to have an impact on the estimated treatment effect size were inadequate randomization, inadequate allocation concealment [19, 20], inadequate blinding [16, 20, 21], and industrial funding [22], although not all the studies confirmed these associations [23, 24]. Inflated treatment effect size estimates can lead dental clinicians to implement inappropriate or ineffective treatments and patient care will suffer in response [3].

The assumed association between trial quality and treatment effect size is derived from published 'meta-epidemiological' studies, which are investigations that quantify the extent of bias in treatment effect size estimates related to trial quality in a group of meta-analyses [25]. Meta-epidemiological investigations have been conducted in the field of medicine [16, 18, 23–26], but the transfer to other health care disciplines of conclusions reached in the medical investigations is limited by numerous factors; for example: the examination of quality items that were related to reporting quality and not to methodological quality or bias [23, 25, 26], the failure to examine continuous outcomes (which occur in the majority of studies) [16, 23, 25] based on a preference for evaluating dichotomous outcomes (which can limit the application of conclusions to randomized trials with continuous outcomes); and the

presentation of inconsistent findings regarding the methodological criteria associated with treatment effect size estimates [16, 23].

More importantly, meta-epidemiological studies have shown that bias in treatment effect size estimates related to methodological characteristics or the type of intervention employed [20, 25] can vary between medical fields and among different areas of health research [27]. To our knowledge, a meta-epidemiological study that examined the extent of bias related to the quality of oral health randomized trials, using meta-analysis as the level of analysis, has not been conducted in any of the nine specialized dental fields. It is unclear to what extent this holds true in oral health trials, which have some unique design characteristics compared with randomized trials in other health fields—for example: difficulty in applying blinding, use of a broad range of interventions (surgical, nonsurgical, drug, nondrug) [28], use of multiple outcomes, common use of split-mouth and crossover designs, and clustering effects [9, 29]—all of which make the evaluation of oral health trials more challenging than the evaluation of trials in other health areas.

1.3. Hypothesis, Aim, and Objectives

The basic hypothesis of this study is that there is no difference in treatment effect size estimates among oral health trials that meet certain methodological quality characteristics, such as: adequate randomization, adequate allocation concealment, baseline balance, adequate blinding of assessors and participants, similarity of cointerventions, a description of withdrawals, and adequate treatment compliance (among others). A secondary hypothesis is that oral health trials with different nonmethodological characteristics, such as nature of the intervention, type of outcome, and dental specialty, will not have different treatment effect size estimates.

The study's overall aim is to examine the empirical evidence quantifying bias associated with methodologic characteristics in oral health randomized trials. This study was conducted in four interconnected phases with the following specific objectives were to:

1) identify systematic reviews published in the domain of oral health research and evaluate them in terms of their epidemiological and descriptive characteristics;
2) describe how often and by what means the risk of bias in clinical trials is assessed in reviews of oral health interventions, and to identify factors associated with risk of bias assessments;
3) examine the current state of oral health randomized trials, and how this state has evolved over time, by evaluating reporting characteristics, methodological characteristics, and the risk of bias;
4) measure associations between methodological characteristics and treatment effect size estimates and determine the impact of specific features (such as dental specialty, type of outcome) on treatment effect size estimates.

References

[1] Yusof ZY, Han LJ, San PP, Ramli AS. Evidence-based practice among a group of Malaysian dental practitioners. J Dent Educ 2008; 72:1333-42.

[2] Iqbal A, Glenny AM. General dental practitioners' knowledge of and attitudes towards evidence based practice. Br Dent J 2002; 193:587-91.

[3] Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000; 342:1887-92.

[4] Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 2003; 7:1-76.

[5] Needleman I, Moles DR, Worthington H. Evidence-based periodontology, systematic reviews and research quality. Periodontol 2000 2005; 37:12-28.

[6] Richards D. Critically appraising systematic reviews. Evid Based Dent 2010; 11:27-9.

[7] Saltaji H, Cummings GG, Armijo-Olivo S, Major MP, Amin M, Major PW, et al. A descriptive analysis of oral health systematic reviews published 1991-2012: cross sectional study. PloS One 2013; 8:e74545.

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

[9] Shea B, Moher D, Graham I, Pham B, Tugwell P. A comparison of the quality of Cochrane reviews and systematic reviews published in paper-based journals. Eval Health Prof 2002; 25:116-29.

[10] 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; 35:244-8.

[11] Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001; 134:663-94.

[12] Higgins JP, Green S, eds. Cochrane Handbook For Systematic Reviews of Interventions version 5.0.0. Cochrane Collaboration 2008.

[13] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010; 8:336-41.

[14] Panic N, Leoncini E, de Belvis G, Ricciardi W, Boccia S. Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. PloS One 2013; 8:e83138.

[15] Needleman I, Worthington H, Moher D, Schulz K, Altman DG. Improving the completeness and transparency of reports of randomized trials in oral health: the CONSORT statement. Am J Dent 2008; 21:7-12.

[16] Pandis N, Fleming PS, Worthington H, Salanti G. The quality of the evidence according to GRADE is predominantly low or very low in oral health systematic reviews. PloS One 2015; 10:e0131644.

[17] Lempesi E, Koletsi D, Fleming PS, Pandis N. The reporting quality of randomized controlled trials in orthodontics. J Evid Based Dent Pract 2014; 14:46-52.

[18] Kiriakou J, Pandis N, Fleming PS, Madianos P, Polychronopoulou A. Reporting quality of systematic review abstracts in leading oral implantology journals. J Dent 2013; 41:1181-7.

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

[20] Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001; 323:42-6.

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

[22] Armijo-Olivo S, Fuentes J, Ospina M, Saltaji H, Hartling L. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: a descriptive analysis. BMC Med Res Methodol 2013; 13:116.

[23] Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. J Clin Epidemiol 2006; 59:1249-56.

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

[25] Nuesch E, Reichenbach S, Trelle S, Rutjes AWS, Liewald K, Sterchi R, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: A meta-epidemiologic study. Arthritis Rheum 2009; 61:1633-41.

[26] Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med 2012; 157:429-38.

[27] Armijo-Olivo S, Cummings GG, Fuentes J, Saltaji H, Ha C, Chisholm A, et al. Identifying items to assess methodological quality in physical therapy trials: a factor analysis. Phys Ther 2014; 94:1272-84.

[28] Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. J Clin Epidemiol 2006; 59(12):1249-56.

[29] Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010; 63:e1-37.

[30] Pandis N, Turpin DL. Enhancing CONSORT compliance for improved reporting of randomized controlled trials. Am J Orthod Dentofacial Orthop 2014; 145(1):1.

[31] Turpin DL. Challenges faced by journal editors. Am J Orthod Dentofacial Orthop 2015; 147:285-9.

[32] Hrobjartsson A, Boutron I, Turner L, Altman DG, Moher D. Assessing risk of bias in randomised clinical trials included in Cochrane Reviews: the why is easy, the how is a challenge. Cochrane Database Syst Rev 2013; 4:ED000058.

[33] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008; 336:924-6.

[34] Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 2001; 285:1987-91.

[35] Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996; 276:637-9.

[36] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011; 64:383-94.

[37] De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. Lancet 2004; 364:911-2.

[38] De Angelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Is this clinical trial fully registered? —A statement from the International Committee of Medical Journal Editors. N Engl J Med 2005; 352:2436-8.

[39] Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ 2010; 340:c723.

[40] Hooft L, Korevaar DA, Molenaar N, Bossuyt PM, Scholten RJ. Endorsement of ICMJE's Clinical Trial Registration Policy: a survey among journal editors. Neth J Med 2014; 72:349-55.

[41] Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Syst Rev 2012; 1:60.

[42] Montenegro R, Needleman I, Moles D, Tonetti M. Quality of RCTs in periodontology—a systematic review. J Dent Res 2002; 81:866-70.

[43] Nieri M, Clauser C, Franceschi D, Pagliaro U, Saletta D, Pini-Prato G. Randomized clinical trials in implant therapy: relationships among methodological, statistical, clinical, paratextual features and number of citations. Clin Oral Implants Res 2007; 18:419-31.

[44] Vere J, Joshi R. Quality assessment of randomised controlled trials of dental implant surgery and prosthodontics published from 2004 to 2008: a systematic review. Clin Oral Implants Res 2011; 22:1338-45.

[45] Harrison JE. Clinical trials in orthodontics II: assessment of the quality of reporting of clinical trials published in three orthodontic journals between 1989 and 1998. J Orthod 2003; 30:309-15.

[46] Pandis N, Polychronopoulou A, Eliades T. An assessment of quality characteristics of randomised control trials published in dental journals. J Dent 2010; 38:713-21.

[47] Agha R, Cooper D, Muir G. The reporting quality of randomised controlled trials in surgery: a systematic review. Int J Surg 2007; 5:413-22.

[48] Reveiz L, Chapman E, Asial S, Munoz S, Bonfill X, Alonso-Coello P. Risk of bias of randomized trials over time. J Clin Epidemiol 2015; 68:1036-45.

[49] Dechartres A, Charles P, Hopewell S, Ravaud P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. J Clin Epidemiol 2011; 64:136-44.

[50] Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. Health Technol Assess 1999; 3:i-iv, 1-98.

[51] Pandis N, Polychronopoulou A, Madianos P, Makou M, Eliades T. Reporting of research quality characteristics of studies published in 6 major clinical dental specialty journals. J Evid Based Dent Pract 2011; 11:75-83.

[52] Seehra J, Wright NS, Polychronopoulou A, Cobourne MT, Pandis N. Reporting quality of abstracts of randomized controlled trials published in dental specialty journals. J Evid Based Dent Pract 2013; 13:1-8.

[53] Pandis N, Polychronopoulou A, Eliades T. An assessment of quality characteristics of randomised control trials published in dental journals. J Dent 2010; 38 (9):713-21.

[54] Dumbrigue HB, Al-Bayat MI, Ng CC, Wakefield CW. Assessment of bias in methodology for randomized controlled trials published on implant dentistry. J Prosthodont 2006; 15:257-63.

[55] Koletsi D, Pandis N, Fleming PS. Sample size in orthodontic randomized controlled trials: are numbers justified? Eur J Orthod 2013.

[56] Polychronopoulou A, Pandis N, Eliades T. Appropriateness of reporting statistical results in orthodontics: the dominance of P values over confidence intervals. Eur J Orthod 2011; 33:22-5.

[57] Koletsi D, Pandis N, Polychronopoulou A, Eliades T. Does published orthodontic research account for clustering effects during statistical data analysis? Eur J Orthod 2012; 34:287-92.

[58] Koletsi D, Pandis N, Polychronopoulou A, Eliades T. What's in a title? An assessment of whether randomized controlled trial in a title means that it is one. Am J Orthod Dentofacial Orthop 2012; 141:679-85.

[59] Koletsi D, Pandis N, Polychronopoulou A, Eliades T. Mislabeling controlled clinical trials (CCTs) as "randomized clinical trials (RCTs)" in dental specialty journals. J Evid Based Dent Pract 2012; 12:124-30.

[60] Needleman I, Worthington H, Moher D, Schulz K, Altman DG. Improving the completeness and transparency of reports of randomized trials in oral health: The CONSORT Statement. Am J Dent 2008; 21:7-12.

[61] Fenwick J, Needleman IG, Moles DR. The effect of bias on the magnitude of clinical outcomes in periodontology: A pilot study. J Clin Periodontol 2008; 35:775-82.

[62] Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, Flores-Mir C. Methodological characteristics and treatment effect sizes in oral health randomised controlled trials: Is there a relationship? Protocol for a meta-epidemiological study. BMJ Open 2014; 4:e004527.

[63] Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. BMJ 1997; 315:617-9.

[64] 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; 336:601-5.

[65] van Tulder MW, Suttorp M, Morton S, Bouter LM, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine 2009; 34:1685-92.

[66] Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2001; 135:982-9.

[67] Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 2002; 287:2973-82.

[68] Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. Stat Med 2007; 26:2745-58.

[69] Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998; 352:609-13.

[70] Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 2007: MR000012.

[71] Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gotzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int J Epidemiol 2007; 36:847-57.

[72] Hartling L, Hamm MP, Fernandes RM, Dryden DM, Vandermeer B. Quantifying bias in randomized controlled trials in child health: a meta-epidemiological study. PloS One 2014; 9:e88008.

[73] Koletsi D, Spineli LM, Lempesi E, Pandis N. Risk of bias and magnitude of effect in orthodontic randomized controlled trials: a meta-epidemiological review. Eur J Orthod 2016; 38(3):308-12.

Chapter 2

A descriptive analysis of oral health systematic reviews published 1991-2012: cross sectional study*

Humam Saltaji, Greta G. Cummings, Susan Armijo-Olivo, Michael P. Major, Maryam Amin, Paul W. Major, Lisa Hartling, Carlos Flores-Mir

A copy of this chapter has been published in the PloS One journal; Saltaji et al. PLoS One. 2013 Sep 30;8(9):e74545. doi: 10.1371/journal.pone.0074545. eCollection 2013.

Abstract

Objectives: To identify all systematic reviews (SRs) published in the domain of oral health research and describe them in terms of their epidemiological and descriptive characteristics.

Design: Cross sectional, descriptive study.

Methods: An electronic search of seven databases was performed from inception through May 2012; bibliographies of relevant publications were also reviewed. Studies were considered for inclusion if they were oral health SRs defined as therapeutic or non-therapeutic investigations that studied a topic or an intervention related to dental, oral or craniofacial diseases/disorders. Data were extracted from all the SRs based on a number of epidemiological and descriptive characteristics. Data were analysed descriptively for all the SRs, within each of the nine dental specialities, and for Cochrane and non-Cochrane SRs separately.

Results: 1188 oral health (126 Cochrane and 1062 non-Cochrane) SRs published from 1991 through May 2012 were identified, encompassing the nine dental specialties. Over half (n=676; 56.9%) of the SRs were published in specialty oral health journals, with almost all (n=1178; 99.2%) of the SRs published in English and almost none of

the non-Cochrane SRs (n=11; 0.9%) consisting of updates of previously published SRs. 75.3% of the SRs were categorized as therapeutic, with 64.5% examining non-drug interventions, while approximately half (n=150/294; 51%) of the non-therapeutic SRs were classified as epidemiological SRs. The SRs included a median of 15 studies, with a meta-analysis conducted in 43.6%, in which a median of 9 studies/1 randomized trial were included in the largest meta-analysis conducted. Funding was received for 25.1% of the SRs, including nearly three-quarters (n=96; 76.2%) of the Cochrane SRs.

Conclusion: Epidemiological and descriptive characteristics of the 1188 oral health SRs varied across the nine dental specialties and by SR category (Cochrane vs. non-Cochrane). There is a clear need for more updates of SRs in all the dental specialties.

2.1. Introduction

A systematic review (SR) is a useful tool that serves to identify, appraise and integrate the findings of studies on a specific topic using a systematic approach [1-3]. It has become the gold standard for decision-making by clinicians and policy makers, and foundational to evidence-based practice approach [2]. Since the inception of the evidence-based practice approach in dentistry, the number of published SRs conducted in dental fields has rapidly increased [4]. One of the valuable sources for SRs is The Cochrane Collaboration, an international organization that aims to help health care professionals make well-informed decisions about treatment interventions by conducting high quality SRs. It has been acknowledged that SRs produced by this collaboration differ in their characteristics and reporting qualities from non-Cochrane SRs [5-7].

In the field of oral health, there has been no comprehensive evaluation of all the published SRs. A few evaluations [8-9] in the last decade have set out to examine

characteristics of a sample of dental SRs; however, the value of these evaluations is limited. Their limitations include: not examining all the pertinent epidemiologic and descriptive characteristics of oral health SRs; not considering the SR category (Cochrane vs. non-Cochrane) or the dental specialty in the analysis; not examining controversial areas relevant to SRs (e.g., publishing updates of SRs); nor providing a comprehensive evaluation of all the SRs published in the field of oral health research, but rather including a limited number of years in their searches (e.g., 2000 to 2005) and limiting it to the English language) [8].

Given the need for more evidence to guide informed decision-making by dental practitioners, the knowledge gained from a comprehensive description of all the oral health SRs and within each specialty would be of paramount importance. This work would help to: identify gaps where evidence is limited, as well as where more oral health SRs and further development are needed, direct future developments in the field of evidence-based dentistry, and provide information for future methodological and meta-epidemiological studies that are clearly needed to quantify the bias associated with methodologies in oral health randomized clinical trials.

The purpose of this cross-sectional descriptive study is to provide a first step in the development of a database of all SRs published in the domain of oral health research. The objectives were to: (1) identify all of the oral health SRs published from inception through May 2012; and (2) describe the oral health SRs in terms of their epidemiological and descriptive characteristics.

2.2. Materials and Methods

2.2.1. Data sources and searches

Electronic searches up to May 2nd, 2012, were conducted using the following electronic bibliographic databases:

- PubMed (1966 to May 2012, week 1)
- MEDLINE (1980 to 2012, week 18)
- EMBASE (1980 to 2012, week 18)
- ISI Web of Science (1965 to May 2, 2012)
- Evidence-Based Medicine Reviews – Cochrane Database of Systematic Reviews (1991 to second quarter of 2012)
- Health STAR (1966 to May 2012)

The key words used in the search were "systematic review," "meta-analysis," "dentistry," "tooth," "orthodontics," "oral surgery," "endodontics," "periodontics," "prosthodontics," "pedodontics," "pediatric dentistry," "dental public health," and "oral pathology." Subject subheadings and some word truncations, according to each database, were used as well to map all possible key words. The initial search strategy was designed for PubMed **(Table 2.1)** and adapted to other databases. The details of the specific search terms and combinations used in each individual database are listed in **Appendix 2.A**. The electronic searches were developed with the assistance of a librarian specializing in health science databases.

We also searched the American Dental Association (ADA)-Evidence-based Dentistry website [10] on May 18-20, 2012. In addition, we have searched the bibliographies of articles that focused on the quality of SRs in the dental fields. The searches were not limited to the English language nor restricted by other means. The references resulting from the searches were entered in EndNote X5, and duplicates were removed.

2.2.2. Study selection and data extraction

Appropriate reports to be included met the following pre-established eligibility criteria:

- Reports fit within the following definition: Oral health SR was defined as one that studied a therapeutic or non-therapeutic topic related to dental, oral or craniofacial diseases/disorders as defined by the ADA scope of practice [11]. We considered a report to be a SR if the authors set out to summarize evidence from several studies and reported explicit methods to identify and evaluate relevant studies [9, 12].
- The SR should be a full-length report.
- SRs in all languages were eligible.
- If a duplicate involving a Cochrane SR and a non-Cochrane SR generated from it was identified, only the Cochrane SR was included.

Two researchers (H.S & T.K) independently reviewed the list of titles and abstracts for inclusion. Once potentially relevant abstracts were selected, the full reports were retrieved for a final selection process. If the abstract was judged to contain insufficient information to ascertain the appropriateness of the work for inclusion, the full report was obtained and reviewed before a final decision was made. Any discrepancies in the inclusion of reports between researchers were addressed through discussion until a consensus was reached. The selected SRs were classified according to one of the following dental specialties as defined by the ADA [11]:

- Dental public health
- Endodontics
- Oral medicine and pathology
- Oral and maxillofacial radiology
- Oral and maxillofacial surgery
- Orthodontics and dentofacial orthopedics
- Pediatric dentistry
- Periodontics
- Restorative dentistry and prosthodontics

We modified the ADA classification [11] by adding oral medicine to “oral and

maxillofacial pathology", and "restorative dentistry" to "prosthodontics".

A data extraction template was designed using Microsoft Excel and pilot tested. Data were extracted on the following characteristics: [5, 8, 13] dental specialty, year of publication, country of corresponding author, continent of corresponding author, number of authors, number of schools/affiliations, career type of the primary author (e.g., academic, private practice, public health, industry), name of journal, type of journal (e.g., general dentistry, specialty dentistry, non-dental), impact factor of journal, source of funding (e.g., industry, government, foundation, academic), type and focus of review (e.g., therapeutic, non- therapeutic: diagnosis/prognosis, epidemiology, psychological/educational), nature of intervention (e.g., drug, surgical, device, dental material, psychological, educational, policy), language of review, design of included studies, number of included studies, number of included randomized controlled trials (RCTs), whether eligible studies were found, whether a meta-analysis (MA) was conducted, number of studies and RCTs contributing data to the largest MA conducted, and whether the review is an update of a previous report.

Complete data extraction was achieved by a non-blinded assessor (H.S), among which a random sample of roughly 20% (250 SRs) was performed in duplicate by two assessors (H.S & T.K/M.A) to assess accuracy. Discrepancies were resolved through discussion until a consensus was reached.

2.2.3. Data analysis

Data were analyzed descriptively as frequency, median, or interquartile range (IQR). The data were analyzed for all the SRs, within each of the dental specialities, and for Cochrane and non- Cochrane SRs separately. Data analysis was performed using the Statistical Package for Social Sciences (SPSS, Version 18.0; IBM, Armonk, NY) for Windows (Microsoft Corporation, Redmond, WA).

2.3. Results

2.3.1. Literature search

The search returned 9669 potential records for inclusion, including 2854 duplicates. The search results from different electronic databases are listed in **Appendix 2.A**. Through the process of screening, 5414 records were excluded based on title/abstract. The remaining 1401 full-text reports were retrieved for a more detailed evaluation, of which 1002 reports fulfilled the inclusion-exclusion criteria. An additional 186 reports were identified through the ADA-Evidence-based Dentistry website [10] search or reference list search, and 1188 reports were finally included. A flow diagram of the data search is given in **Figure 2.1**. The main reasons for exclusion were not being within the scope of any of the dental fields or not using explicit methods to identify relevant studies.

2.3.2. Prevalence and specialties of oral health SRs

The majority of the SRs were published either in the fields of periodontics (n = 212; 17.8%), prosthodontics and restorative dentistry (n= 198; 16.7%), or dental public health (n= 184 =15.5%). Oral health SRs published in the remaining dental specialities included: oral medicine and oral pathology (n= 162; 13.6%), oral and maxillofacial surgery (n= 59; 13.4%), orthodontics and dentofacial orthopedics (n= 138; 11.6%), endodontics (n= 54; 4.5%), pediatric dentistry (n= 50; 4.2%), and oral and maxillofacial radiology (n= 31; 2.6%). **Table 2.2** provides further details of the number of oral health SRs within each of the nine dental specialities and for Cochrane and non-Cochrane SRs separately.

2.3.3. Characteristics of oral health SRs

The 1188 SRs were published between 1991 and 2012. The median date of

publication of oral health SRs was 2008, ranging from 2006 for dental public health publications to 2009 for oral and maxillofacial radiology publications. **Figure 2.2** shows the increase of oral health SRs, with each year, from 1991 to 2011. The majority of the published SRs were non-Cochrane SRs (n= 1062; 89.4%), while Cochrane SRs contributed only 10% of the total number of SRs (n= 126; 10.6%).

The SRs were published in 194 (96 oral health & 98 non-oral health) journals. More than half of the SRs were published in specialty oral health journals (n= 676; 56.9%), while 373 SRs (31.4%), including all of the Cochrane SRs, were published in general oral health journals. Nearly one third of the non-Cochrane SRs (n =335; 32%) were published in eight (one general and seven specialty) oral health journals, namely the *Journal of Clinical Periodontology* (n= 75; 6.3%), *Clinical Oral Implants Research* (n= 59; 5.0%), the *Journal of Periodontology* (n= 40; 3.4%), the *Angle Orthodontist* (n= 35; 2.9%), the *American Journal of Orthodontics and Dentofacial Orthopedics* (n= 34; 2.9%), the *International Journal of Oral & Maxillofacial implants* (n= 34; 2.9%), t*he Journal of Oral and Maxillofacial Surgery* (n= 30; 2.5%), and the *Journal of the American Dental Association* (n= 28; 2.4%) **(Table 2.3)**. Almost half of the non-Cochrane SRs (n =489; 47%) were published in journals with a relatively high impact factor for the field of dentistry (>1.5); while 7.1% (n= 84) of the non-Cochrane SRs were published in oral health journals that did not have an impact factor **(Table 2.4)**.

The corresponding authors of the SRs were most frequently from Europe (Cochrane SRs: n= 99; 78.6% & non-Cochrane SRs: 546; 51.4%) followed by North America, with one country (UK) accounting for nearly two-thirds (n= 82; 65.1%) of the Cochrane SRs, another country (USA) accounting for nearly one-quarter (n =217; 20.4%) of the non-Cochrane SRs, and four countries (the United States, the United Kingdom, Canada, and the Netherlands) accounting for nearly half (n= 581; 48.9%) of all oral health SRs **(Table 2.4)**. Approximately half of the SRs had authors from

multiple centers (median of two affiliations for non-Cochrane SRs and three affiliations for Cochrane SRs), and included four to six authors (median of three authors for non-Cochrane SRs and five authors for Cochrane SRs) although 78 (7.3%) of the non- Cochrane SRs were single-authored **(Table 2.5 & Appendix 2.D)**. The primary authors were from an academic background in the vast majority of the oral health SRs (n= 1084; 91.2%), with a small proportion published by private practice clinicians (n =47; 4.0%), researchers from policy/public health organizations (n= 39; 3.3%), and researchers from dental companies (n =18; 1.5%).

Three-quarters (n= 894; 75.3%) of the SRs, including all the Cochrane SRs, were categorized as therapeutic; the vast majority (approximately 90%) of the SRs in the fields of prosthodontics and restorative dentistry, oral and maxillofacial surgery, and endodontics were categorized as therapeutic, and the vast majority (n= 29; 93.5%) of the SRs in the field of oral and maxillo-facial radiology were categorized as non-therapeutic. Approximately half (n= 150/294; 51%) of the non-therapeutic SRs were classified as epidemiology SRs, including the majority (n= 56/82; 68.3%) of the SRs in the field of oral medicine and oral pathology, and 38.1% (112/294) as diagnostic/prognostic SRs, including the vast majority (n= 25/29; 86.2%) of the SRs in the field of oral and maxillo-facial radiology **(Table 2.5 & Appendix 2.D)**.

The nature of intervention varied across the dental specialties, with nearly two-thirds (n= 577/894; 64.5%) of all the therapeutic SRs examining non-drug interventions, including the vast majority (approximately 90%) of the therapeutic SRs in the fields of prosthodontics and restorative dentistry, and orthodontics and dentofacial orthopedics. Nearly three-quarters (n =651/894; 72.8%) of all the therapeutic SRs examined non-surgical interventions, including almost all of the therapeutic SRs in the fields of dental public health and pediatric dentistry. Moreover, similar ratios of therapeutic SRs reported examining surgical (n =145/894; 16.2%), device (n= 163/894; 18.2%), drug (n =194/894; 21.7%), and multiple (n= 160/894;

17.9%) interventions, with a small portion (n= 31/894; 3.5%) examining psychological or educational interventions **(Table 2.5 & Appendix 2.D)**.

One-quarter (n= 298; 25.1%) of all the SRs, including nearly three-quarters (n= 96; 76.2%) of the Cochrane SRs, reported receiving at least one source of funding. Approximately one-third (n =66/184; 35.9%) of the SRs in the field of dental public health, including all (n =21/21; 100%) the Cochrane SRs, received funding, while only a small portion (n = 2/31; 6.5%) of the SRs in the field of oral and maxillo-facial radiology reported receiving funding. The most common sources of funding for non-Cochrane SRs were foundations (n= 67/202; 33.2%) followed by academic (n =41/202; 20.3%) and government (n =37/202; 18.3%) sources. For Cochrane SRs, nearly three-quarters (n= 90; 71.4%) reported receiving an external source of funding, with "foundations" as the most common (30/48; 62.5%) external source of funding **(Table 2.5 & Figure 2.3)**.

Almost all (n= 1178; 99.2%) of the SRs were published in English, and almost none of the non-Cochrane SRs (n= 11; 0.9%) were updates of previously published SRs **(Table 2.4)**. While almost all the Cochrane SRs included RCTs only (n =97/126; 93.3%), only 17.6% (n= 186/1062) of the non-Cochrane SRs included only RCTs. The research design of studies included in non- Cochrane SRs were most often non-RCTs (n= 423; 39.9%), including the majority of the SRs in the fields of oral and maxillofacial radiology (n= 25/31; 80.6%) and oral medicine and oral pathology (83/140; 59.3%), followed by RCTs and other designs (n =325; 30.7%) and RCTs only (n= 186; 17.6%).

Non-Cochrane SRs included a median of 15 studies, ranging from 12 for orthodontics and dentofacial orthopedics to 16.5 for oral medicine and oral pathology; while the median number of studies included in Cochrane SRs was five, ranging from two for oral medicine and oral pathology to twelve for dental public health **(Table 2.6**

& Appendix 2.F). The median number of RCTs included in the non-Cochrane SRs was one, ranging from zero for oral medicine and & oral pathology, pediatric dentistry and orthodontics and dentofacial orthopedics to four for dental public health, while the Cochrane SRs included a median of five RCTs, ranging from two for orthodontics and dentofacial orthopedics and oral and maxillofacial surgery to twelve for dental public health. There were no eligible studies in 22 (17.5%) of the Cochrane SRs, while only three (0.3%) of the non-Cochrane SRs included no relevant studies.

Less than half of the SRs (n= 518; 43.6%) conducted quantitative analyses (meta-analyses). A median of nine studies and a median of two RCTs were included in the largest MA conducted **(Table 2.6 & Appendix 2.G)**. This varied across dental specialties and the category of the review, with a median of 5.5 studies and 4.5 RCTs included in the largest MA conducted in the Cochrane SRs, and a median of nine studies and one RCT included in the largest MA conducted in the non-Cochrane SRs. 152 (29.4%) SRs (32 Cochrane and 120 non-Cochrane), in which a MA was conducted, included at least five RCTs. **Appendices 2.B** to **2.G** provide further details of the epidemiological and descriptive characteristics of all of the oral health SRs, within each of the dental specialities, and for Cochrane and non-Cochrane SRs separately.

2.4. Discussion

SRs are important tools for researchers, clinicians and policy makers because they serve to systematically identify and appraise the available evidence on a specific topic, and to integrate it into an evidence-based conclusion [1-3]. This study demonstrates variation in the characteristics of SRs across the nine dental specialties and according to SR category (Cochrane vs. non- Cochrane). Our findings show that the number of SRs published in the domain of oral health research and within each dental specialty has steadily increased over the last two decades, similar to the results

published in previous reports examining dental SRs [9, 12, 14] and medical SRs [5, 13, 15]. However, there was a decline observed in 2011, which was also observed in previously published reports, [12, 13] and could be attributed to the fact that oral health SRs published in late 2011 would not necessarily be indexed by May 2nd, 2012, a so called time lag. The increased volume of SRs may not necessarily reflect a steady improvement in the methodological quality of the published SRs though. Previously published reports demonstrated that oral health SRs improved as a whole over a period of five years, [8, 12] with some specialities (e.g., periodontics) performing better at meeting the methodological quality criteria [12]. In order to avoid biased results and misleading decision-making in the dental practice, it is necessary that the increase in the quantity of published dental SRs be associated with an increase in the methodological quality of these SRs. Our study did not provide detailed information on methodological quality criteria, as our overall goal was to provide the reader with a detailed descriptive analysis of all SRs published in the field of dentistry.

Dental specialities were ranked according to the proportion of the total published SRs as follows (in descending order): periodontics, prosthodontics and restorative dentistry, dental public health, oral medicine and oral pathology, oral and maxillo-facial surgery, orthodontics and dentofacial orthopedics, endodontics, pediatric dentistry, and oral and maxillofacial radiology. Despite the steady increase in the number of published oral health SRs, there have only been a few SRs published in the fields of oral and maxillofacial radiology (31 SRs), pediatric dentistry (50 SRs), and endodontics (54 SRs); therefore, more SRs are specifically needed in these fields. However, it should be noted that many pediatric-related SRs were found to be better classified in the field of dental public health (e.g., “Fluoride supplements for preventing dental caries in children” [16]); ergo it is likely that the resulting number of published pediatric dental SRs in this study are underestimated and may not be representative of reality. Additionally, given that the ADA classification [11] was utilized for categorizing the selected SRs, implantology-related SRs were not

classified in an individual field, but in one of three specialties (periodontics, oral and maxillofacial surgery, or prosthodontics). Given that the field of implantology is a relatively new and quickly growing dental field, future studies should consider it as an individual dental specialty in order not to inflate the SR count of other specialties.

Oral health SRs appear to be published more often in specialty journals. Our results showed that more than half of the SRs were published in specialty oral health journals, with almost half of the SRs published in journals with a high impact factor. Nearly half of the SRs were from four countries: the United States, the United Kingdom, Canada, and the Netherlands. This trend is similar to what was found in recently published reports, [14, 17] and could be attributed to an increased interest of the public sector and government agencies in these countries to make decisions regarding financing dental services based on the findings of the SRs [14].

The current study revealed that many characteristics of the published oral health SRs still require improvement. For example, only 11 out of the 1062 non-Cochrane SRs were updates of previously published SRs. Furthermore, none of the 11 updates identified in our research were considered “up-to-date” according to the Cochrane policy, which requires updating the SR every two years [6]. This is a disappointing fact given that “up-to-date” evidenced-based conclusions are considered essential for decision making [18]. This might be explained by the fact that updates are usually given lower priority by funding agencies and editors, who tend not to publish updates with results that are the same as previously published versions [5, 18]. Therefore, updates of SRs in the domain of oral health research are clearly needed. In light of this, examining where updates are needed and identifying specific mechanics are a priority in order to ensure that decision-making processes in the dental fields are based on the best up-to-date evidence. This finding does not apply completely to Cochrane SRs, given that authors of Cochrane SRs are supposed to update their reports every two years according to Cochrane standards, [6, 19] although a previously

published report [13] identified a considerable portion (38%) of the Cochrane child-related SRs as not up-to-date based on the Cochrane criteria.

The results showed that 78 (7.3%) of the non-Cochrane SRs were single-authored, while nearly half of the SRs involved authors from multiple locations and included four to six authors. Having at least two assessors to select relevant reports and extract data in duplicate reduces the potential selection and extraction bias and decreases the possibility of accidental exclusion of relevant reports and inaccurate extraction of relevant data, which may lead to distorted conclusions [20-22]. In addition, only one or two databases were searched by approximately half of the non-Cochrane SRs. This is problematic because failure to search multiple databases may lead to missing relevant studies, which can produce biased results and possibly mislead decision-making related to dental practice [23-26].

The results also revealed that the research design of the included studies varied across dental specialties and by type of the SR. While almost all the Cochrane SRs included RCTs only, a small proportion (17.6%) of the non-Cochrane SRs exclusively included RCTs. This may be attributed to Cochrane policy and guidance, which has historically focused on reviews of health care interventions and inclusion of only RCT. This policy explains why all the retrieved Cochrane SRs were therapeutic, while only 72% of the non-Cochrane SRs were therapeutic. Moreover, the nature of the interventions varied across the dental specialties, with nearly two-thirds of all the therapeutic SRs examining non-drug related interventions. This proportion is higher than the proportion found in previous reports examined in medical SRs, [5, 12] and possibly reflects the greater variability in oral health interventions compared to medical interventions. Interestingly, a sizable proportion of the Cochrane SRs (17.5%), including nearly a third of the SRs in the field of oral and maxillofacial surgery, found no appropriate trials to be included. This may be explained by Cochrane's selective policy of only including RCTs in study selection, considering

MAs of RCTs with low risk of bias as the highest level of evidence on the efficacy of treatment interventions [2]. This proportion is higher than the proportion of child related Cochrane SRs (9.3%) found by Bow et al [13], and possibly highlights the need for more trials to be conducted in the dental specialties, specifically related to oral and maxilla-facial surgery. Similarly, the number of included studies varied across dental specialties and by type of SR. The median number of studies included in Cochrane SRs was five, ranging from two in oral medicine and oral pathology to 12 for dental public health. This median number is less than the number found in child-related Cochrane SRs (seven studies), [13] and again reflects a clear need for more studies to be conducted in the dental specialties.

2.4.1. Strengths and limitations

This cross-sectional observational study provides a comprehensive descriptive analysis of all SRs published in the domain of oral health research from inception through May 2012. Our data searches covered six different databases in addition to the ADA Evidence-based Dentistry website [10], which contains a list of systematic/literature reviews related to oral health research. The addition of this website in our search complemented the other databases searched, making it more comprehensive. However, one of the clear limitations in our research is the data extraction method, which was performed by one assessor. This is problematic because it creates the potential for bias, even though accuracy was assessed by having a 20% random sample (250 SRs) examined in duplicate by two assessors. A further limitation is that we extracted data based on what was reported by the authors of the SRs and, thus, it is possible that some characteristics, such as the type of study included in the SRs, were inappropriately reported by the authors or altogether omitted (which occurred with the source of funding). Another potential limitation is that the implantology-related SRs were categorized in one of three specialties (periodontics, oral and maxillofacial surgery, or prosthodontics), as the ADA classification [11]

utilized in our study does not classify “implantology” as an individual specialty. Future methodological studies should consider “implantology” as an individual dental field. Additionally, we may have missed some characteristics in our data extraction such as SR registration which is not very well-known to oral health systematic reviewers. Finally, we may have included SRs in our sample that are not directly related to oral health research but are relevant to dental/oral diseases, such as “orofacial pain in patients receiving cancer therapy” [27].

2.5. Conclusions

We have identified and described a total of 1188 oral health (126 Cochrane and 1062 non-Cochrane) SRs published from 1991 through May 2012, encompassing the nine dental specialties. Epidemiological and descriptive characteristics of the oral health SRs varied across the nine dental specialties and by SR category (Cochrane vs. non-Cochrane). There is a clear need for more regular updating of SRs. This includes the examination of where updates are needed and the development of mechanisms to regularly update SRs to ensure that dental practice decision-making is based on up-to-date information. Oral health SRs require improvement with respect to having multiple assessors and searching more than one database. Finally, future methodological studies should consider ‘‘implantology’’ as an individual dental specialty.

Table 2.1. Search strategy in PubMed

#1 systematic review* OR meta-analys*
#2 dent* OR tooth OR teeth OR orthodon* OR oral surg* OR endodon* OR periodon* OR prosthodon* OR pedodon* OR pediatric* dentistry OR paediatric* dentistry OR dent* public health OR oral pathology
#3 #1 AND #2

Table 2.2. Specialties of oral health systematic reviews, N (% total)

Dental Specialty	Overall (Cochrane & Non-Cochrane SRs) N=1188	Non-Cochrane SRs (N=1062)	Cochrane SRs (N=126)
Periodontics	212 (17.8)	203 (19.1)	9 (7.1)
Prosthodontics & Restorative Dentistry	198 (16.7)	179 (16.9)	19 (15.1)
Dental Public Health	184 (15.5)	163 (15.3)	21 (16.7)
Oral Medicine & Oral Pathology	162 (13.6)	140 (13.2)	22 (17.5)
Oral and Maxillofacial Surgery	159 (13.4)	134 (12.6)	25 (19.8)
Orthodontics and Dentofacial Orthopedics	138 (11.6)	123 (11.6)	15 (11.9)
Endodontics	54 (4.5)	47 (4.4)	7 (5.6)
Pediatric Dentistry	50 (4.2)	42 (4.0)	8 (6.3)
Oral and Maxillofacial Radiology	31 (2.6)	31 (2.9)	0 (0.0)
Total	1188 (100)	1062 (100)	126 (100)

Table 2.3. Journals in which oral health systematic reviews were published

Journal Title	Classification	No. (%) of 1188 SRs (Cochrane and Non-Cochrane SRs)	Rank†	Impact Factor†
1. Journal of Clinical Periodontology	Specialty	75 (6.3)	5	2.996
2. Clinical Oral Implants Research	Specialty	59 (5.0)	13	2.514
3. Journal of Periodontology	Specialty	40 (3.4)	11	2.602
4. Angle Orthodontist	Specialty	35 (2.9)	40	1.207
5. American Journal of Orthodontics and Dentofacial Orthopedics	Specialty	34 (2.9)	35	1.381
5. The International Journal of Oral & Maxillofacial implants	Specialty	34 (2.9)	21	1.776
6. Journal of Oral and Maxillofacial Surgery	Specialty	30 (2.5)	27	1.640
7. The Journal of the American Dental Association	General	28 (2.4)	22	1.773
8. Journal of Endodontics	Specialty	26 (2.2)	7	2.880
8. Journal of Dentistry	General	26 (2.2)	6	2.947
9. Journal of Dental Research‡	General	25 (2.1)	3	3.486
9. Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontology	General	25 (2.1)	33	1.457
10. The International journal of prosthodontics	Specialty	24 (2.0)	36	1.376
11. Journal of Dental Education	Specialty	23 (1.9)	61	0.906
12. The Journal of Prosthetic Dentistry	Specialty	21 (1.8)	37	1.324
13. Journal of Oral Rehabilitation	Specialty	19 (1.6)	30	1.529
14. Community Dentistry and Oral Epidemiology	Specialty	18 (1.5)	19	1.894
15. Dentomaxillofacial Radiology	Specialty	16 (1.3)	49	1.081
16. International journal of oral and maxillofacial surgery	Specialty	14 (1.2)	32	1.506
17. International Journal of Dental Hygiene	Specialty	13 (1.1)	63	0.871
18. Dental Materials	Specialty	12 (1.0)	4	3.135
18. International Dental Journal	General	12 (1.0)	58	0.963
18. Acta odontologica scandinavica	General	12 (1.0)	50	1.066
Cochrane Database of Systematic Reviews	General	126 (10.6)	N/A	5.912
Other oral health journals (with IF)	General/ Specialty	218 (18.3)	-	-
Other oral health journals (IF is not found)	General/ Specialty	84 (7.1)	-	Not found
Non-oral health journals	Non-dental	139 (11.7)	-	-
Total number of oral health journals (1094 SRs)	96 (63 with IF & 33 without IF)			
Total number of non-oral health journals (139 SRs)	98			

† 2011 Journal Citation Reports® (Thomson Reuters, 2012).
‡SRs published in *Critical Reviews in Oral Biology & Medicine* were included in *Journal of Dental Research*. *Critical Reviews in Oral Biology & Medicine* was merged into the *Journal of Dental Research* (last issue Nov 2004).
IF, impact factor; N/A, not applicable.

Table 2.4. Characteristics of oral health systematic reviews

Characteristic	No. (%) of 1188 SRs (Cochrane and Non-Cochrane SRs)	No. (%) of 126 Cochrane SRs	No. (%) of 1062 Non-Cochrane SRs
Year of publication, median			
	2008	Protocol: 2004; Review: 2007	2008
Continent of corresponding author, n (% total)			
Europe	645 (54.3)	99 (78.6)	546 (51.4)
North America	303 (25.5)	2 (1.6)	301 (28.3)
Asia	99 (8.3)	13 (10.3)	86 (8.1)
South America	61 (5.1)	10 (7.9)	51 (4.8)
Australia	47 (4.0)	0 (0.0)	47 (4.4)
Africa	33 (2.8)	2 (1.6)	31 (2.9)
Country of corresponding author, n (% total)			
No. of countries	47	20	47
USA	218 (18.4)	1 (0.8)	217 (20.4)
UK	196 (16.5)	82 (65.1)	114 (10.7)
Canada	85 (7.2)	1 (0.8)	84 (7.9)
The Netherlands	82 (6.9)	1 (0.8)	81 (7.6)
Switzerland	67 (5.6)	0 (0.0)	67 (6.3)
Italy	65 (5.5)	4 (3.2)	61 (5.7)
Brazil	57 (4.8)	9 (7.1)	48 (4.5)
Germany	46 (3.9)	4 (3.2)	42 (4.0)
Sweden	40 (3.4)	0 (0.0)	40 (3.8)
China	40 (3.4)	5 (4.0)	35 (32.9)
Greece	28 (2.4)	0 (0.0)	28 (2.6)
Australia	28 (2.4)	0 (0.0)	28 (2.6)
Spain	25 (2.1)	0 (0.0)	25 (2.4)
South Africa	25 (2.1)	1 (0.8)	24 (2.3)
Other	186 (15.6)	18 (14.3)	168 (15.8)
Career type of the primary author, n (% total)			
Academic	1084 (91.2)	105 (83.3)	979 (92.2)
Private practice	47 (4.0)	5 (4)	42 (4)
Policy/Public health	39 (3.3)	16 (2.7)	23 (2.2)
Industry	18 (1.5)	0 (0.0)	18 (1.7)
Journal impact factor‡, n (% total)			
0.0-1.000	122 (10.3)	0 (0.0)	122 (11.5)
1.001-1.500	219 (18.4)	0 (0.0)	219 (20.6)

1.501-2.000	170 (14.3)	0 (0.0)	170 (16.0)
2.001-3.000	282 (23.7)	0 (0.0)	282 (26.5)
3.001-4.000	46 (3.9)	0 (0.0)	46 (4.3)
>4.001§	126 (10.6)	126 (100)	0 (0.0)
Not found*	84 (7.1)	0 (0.0)	84 (7.9)
N/A¶	139(11.7)	0 (0.0)	139 (13.1)
Journal type†, n (% total)			
General Dentistry	373 (31.4)	126 (100.0)	247 (23.3)
Specialty Dentistry	676 (56.9)	0 (0.0)	676 (63.7)
Non-Dental	139 (11.7)	0 (0.0)	139 (13.1)
Language, n (% total)			
English	1178 (99.2)	126 (100.0)	1052 (99.1)
Bilingual English	6 (0.5)	0 (0.0)	6 (0.6)
Other	4 (0.3)	0 (0.0)	4 (0.4)
Update of previous review‡, n (% total)			
Yes	11 (0.9)	N/A	11 (1.0)
No	1051 (88.5)	N/A	1051 (99.0)
Number of databases, n (% total)			
1-2	518 (43.6)	1 (0.8)	517 (48.7)
3-4	373 (31.4)	62 (49.2)	311 (29.3)
>4	253 (21.3)	63 (50.0)	190 (17.9)
Unclear/Not reported	44 (3.7)	0 (0.0)	44 (4.1)

‡ 2011 Journal Citation Reports® (Thomson Reuters, 2012). The highest impact factor for oral health journals is 3.961 (*Periodontology 2000*).
†Cochrane Database of Systematic Reviews (CDSR), where Cochrane SRs are published, was classified as a general journal.
§Includes Cochrane SRs only (CDSR's impact factor = 5.912).
*Includes SRs published in oral health journals without impact factor.
¶ Includes SRs published in non-oral health journals.
‡ Does not equal 100 % for overall, as Cochrane SRs were not considered in the analysis.
N/A, not applicable.

Table 2.5. Characteristics of oral health systematic reviews

Characteristic	No. (%) of 1188 SRs (Cochrane and Non-Cochrane SRs)	No. (%) of 126 Cochrane SRs	No. (%) of 1062 Non-Cochrane SRs
Number of Authors			
Number of authors, median (IQR)			
	4 (2, 5)	5 (4, 6)	3 (2, 5)
Number of authors, n (% total)			
1	78 (6.6)	0 (0.0)	78 (7.3)
2-3	505 (42.5)	26 (20.6)	479 (45.1)
4-6	520 (43.8)	81 (64.3)	439 (41.3)
≥ 7	85 (7.2)	19 (15.1)	66 (6.2)
Number of Schools or Affiliations			
Number of schools, median (IQR)			
	2 (1, 3)	3 (2, 4)	2 (1, 3)
Number of schools, n (% total)			
1	454 (38.2)	19 (15.1)	435 (41.0)
2-3	573 (48.2)	57 (45.2)	516 (48.6)
≥ 4	161 (13.6)	50 (39.7)	111 (10.5)
Type of Review, N (% Total)			
Therapeutic	894 (75.3)	126 (100)	768 (72.3)
Non-therapeutic	294 (24.7)	0 (0.0)	294 (27.7)
Focus of Non-therapeutic SRs, N (% Total)			
Total Number	N=294	N=0	N=294
Diagnosis/Prognosis	112 (38.1)	0 (0.0)	112 (38.1)
Epidemiology	150 (51)	0 (0.0)	150 (51)
Psychological/Educational/Policy/ Quality of studies	32 (10.9)	0 (0.0)	32 (10.9)
Type of Intervention in Therapeutic SRs, N (% Total)			
Classification I, N (% Total)			
Total Number	N=894	N=126	N=768
Drug	219 (24.5)	34 (27.0)	185 (24.1)
Non-drug	577 (64.5)	74 (58.7)	503 (65.5)
Both	98 (11.0)	18 (14.3)	80 (10.4)
Classification II, N (% Total)			
Total Number	N=894	N=126	N=768
Surgical	151 (16.9)	25 (19.8)	126 (16.4)
Non-surgical	651 (72.8)	96 (76.2)	555 (72.3)
Both	92 (10.3)	5 (4.0)	87 (11.3)
Classification III, N (% Total)			
Total Number	N=894	N=126	N=768

Surgical	145 (16.2)	22 (17.5)	123 (16.0)
Device	163 (18.2)	12 (9.5)	151 (19.7)
Drug	194 (21.7)	35 (27.8)	159 (20.7)
Dental Material	96 (10.7)	12 (9.5)	84 (10.9)
Psychological/Educational/Policy	31 (3.5)	7 (55.6)	24 (3.1)
Other	105 (11.7)	22 (17.5)	83 (10.8)
Multiple/Combined	160 (17.9)	16 (12.7)	144 (18.7)
Source of Funding, N (% Total)			
Classification I, N (% Total)			
Yes	298 (25.1)	96 (76.2)	202 (19.0)
No	58 (4.9)	1 (0.8)	57 (5.4)
Not reported	832 (70.0)	29 (23)	803 (75.6)
Classification II, N (% Total)			
Total Number	N/A	N=48‡	N=202
Industry	-	1 (2.1)	20 (9.9)
Government	-	7 (14.6)	37 (18.3)
Foundation	-	30 (62.5)	67 (33.2)
Academic	-	1 (2.1)	41 (20.3)
Multiple	-	9 (18.8)	33 (16.3)
Unclear	-	0 (0.0)	4 (2.0)
Classification III, N (% Total)			
Internal only	-	49 (38.9)	-
External only	-	6 (4.8)	-
Both internal and external	-	41 (32.5)	-
Not reported	-	29 (23.0)	-
No	-	1 (0.8)	-
‡ External funding only; N/A, not applicable.			

Table 2.6. Characteristics of included studies in oral health systematic reviews

Characteristic	No. (%) of 1188 SRs (Cochrane and Non-Cochrane SRs)	No. (%) of 126 Cochrane SRs	No. (%) of 1062 Non-Cochrane SRs
Study Designs of SRs with Eligible Studies, N (% Total)			
Total Number	N=1163	N=104	N=1059
RCTs only	283 (24.3)	97 (93.3)	186 (17.6)
CCTs only	10 (0.9)	1 (1.0)	9 (0.8)
RCTs and CCTs	71 (6.1)	4 (3.8)	67 (6.3)
RCTs and other designs	326 (28.0)	1 (1.0)	325 (30.7)
Non-RCTs	424 (36.5)	1 (1.0)	423 (39.9)
Unclear/Not reported	49 (4.2)	0 (0.0)	49 (4.6)
Number of Included Studies			
Number of included studies, median (IQR)			
	14 (7, 28)	5 (1, 13)	15 (8, 29)
Number of included studies, n (% total)			
0	25 (2.1)	22 (17.5)	3 (0.3)
1-5	166 (14.0)	45 (35.7)	121 (11.4)
6-15	433 (36.4)	32 (25.4)	401 (37.8)
16-30	261 (22.0)	17 (13.5)	244 (23.0)
>30	251 (21.1)	10 (7.9)	241 (22.7)
Unclear/Not reported	52 (4.4)	0 (0.0)	52 (4.9)
Number of Included RCTs			
Number of included RCTs, median (IQR)			
	1 (0, 7)	5 (1, 12)	1 (0, 6)
Number of included RCTs, n (% total)			
0	461 (38.3)	24 (19)	437 (41.1)
1-2	116 (9.8)	27 (21.4)	89 (8.4)
3-4	72 (6.1)	11 (8.7)	61 (5.7)
5-10	183 (15.4)	27 (21.4)	156 (14.7)
11-20	96 (8.1)	18 (14.3)	78 (7.3)
>20	75 (75)	19 (15.1)	56 (5.3)
Unclear/Not reported	185 (15.6)	0 (0.0)	185 (17.4)
Meta-Analysis Conducted, N (% Total)			
Yes	518 (43.6)	64 (50.8)	454 (42.7)
No	670 (56.4)	62 (49.2)	608 (57.3)
Number of Studies Contributed Data to the Largest Meta-Analysis Conducted			
Total Number	N=518	N=64	N=454
Number of studies in largest meta-analysis, median (IQR)			
	9 (5, 18)	5.5 (3, 9)	9 (6, 19)
Number of studies in largest meta-analysis, n (% total)			

Total Number	N=518	N=64	N=454
2-4	100 (19.3)	31 (48.4)	69 (15.2)
5-10	200 (38.6)	20 (31.2)	180 (39.6)
11-20	108 (20.8)	7 (10.9)	101 (22.2)
>20	104 (20.1)	6 (9.4)	98 (21.6)
Unclear/Not reported	6 (1.2)	0 (0.0)	6 (1.3)
Number of RCTs in largest meta-analysis, median (IQR)			
	2 (0, 6)	4.5 (2, 9)	1 (0, 6)
Number of RCTs in largest meta-analysis, n (% total)			
0	188 (36.3)	0 (0.0)	188 (41.4)
2-4	107 (20.7)	32 (50.0)	75 (16.5)
5-10	104 (20.1)	19 (29.7)	85 (18.7)
11-20	27 (5.2)	7 (10.9)	20 (4.4)
>20	21 (4.1)	6 (9.4)	15 (3.3)
Unclear/Not reported	71 (13.7)	0 (0.0)	71 (15.6)
RCTs, randomized controlled trials; CCTs, controlled clinical trials; N/A, not applicable			

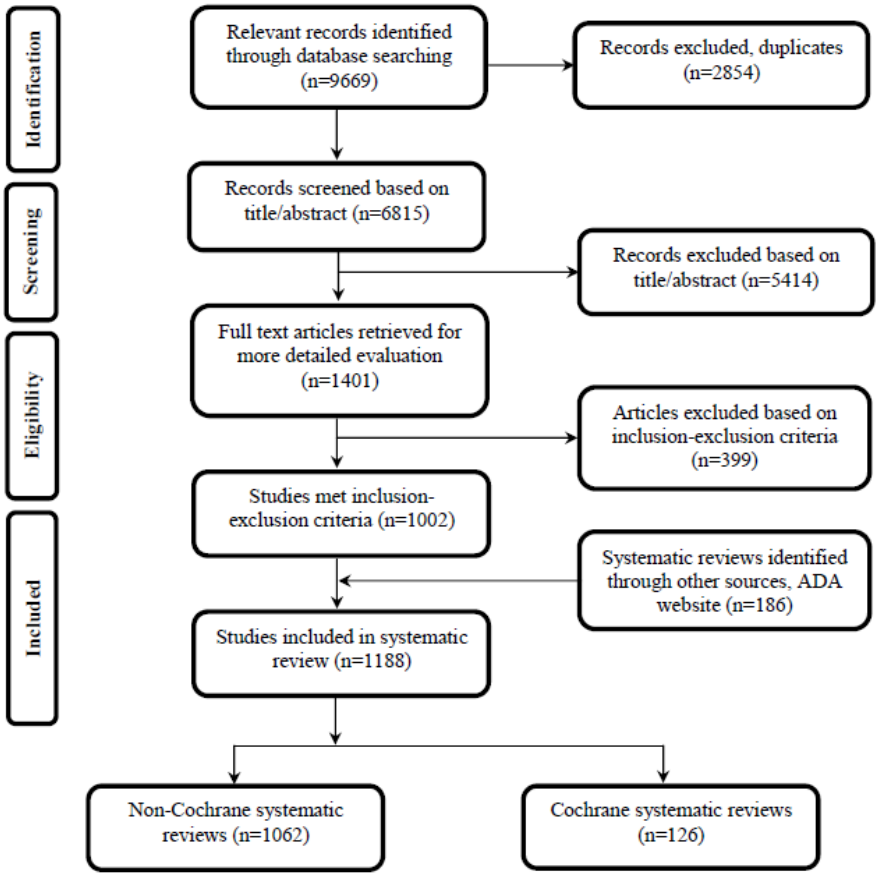

Figure 2.1. Flow diagram of the literature search according to the PRISMA [28].

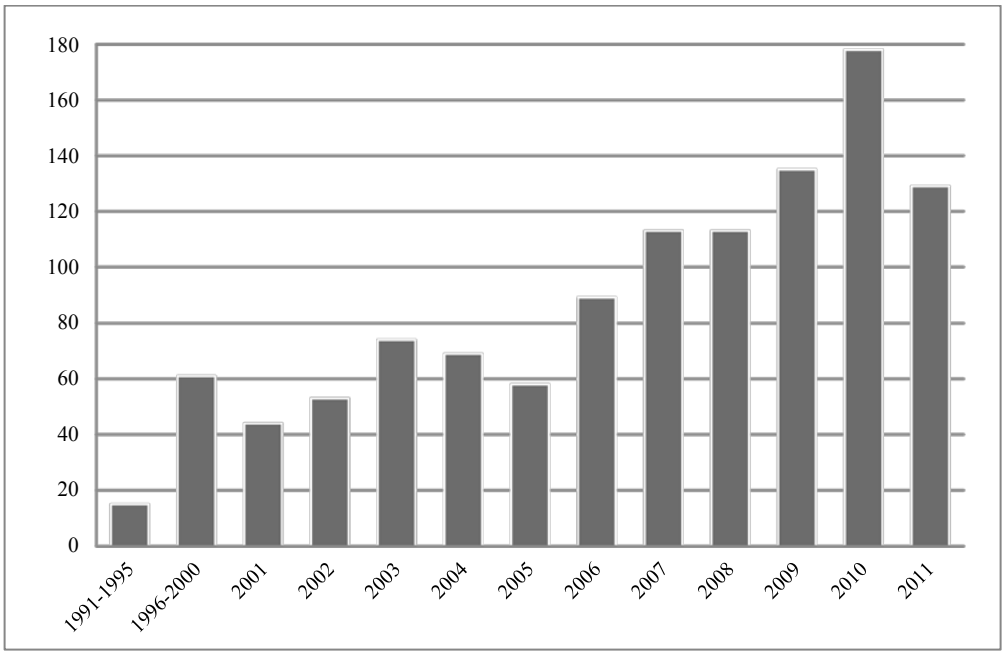

Figure 2.2. Number of systematic reviews published by year; 2012 was not included in the figure because the full year was not searched (Y axis represents numbers of reviews).

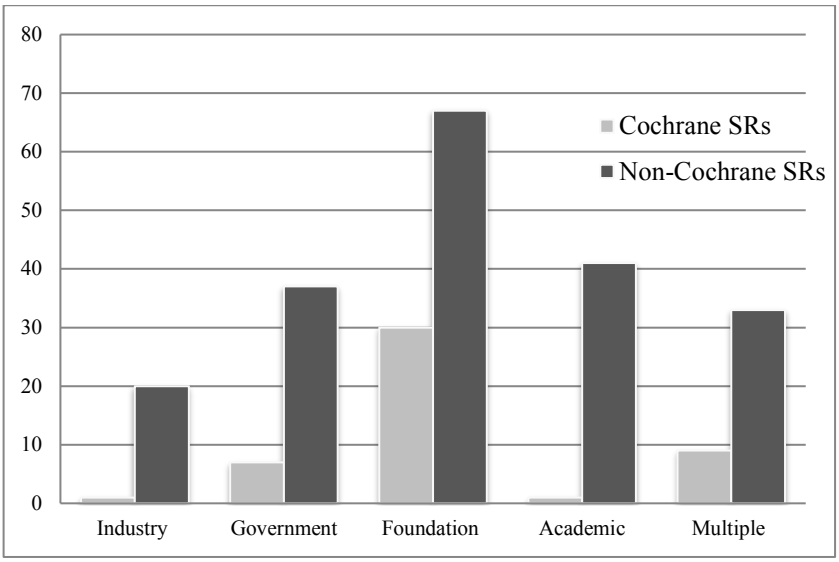

Figure 2.3. Number of oral health systematic reviews by source of funding.

References

[1] Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 2003; 7:1-76.

[2] Richards D. Critically appraising systematic reviews. Evid Based Dent 2010; 11:27-9.

[3] Needleman I, Moles DR, Worthington H. Evidence-based periodontology, systematic reviews and research quality. Periodontol 2000. 2005; 37:12-28.

[4] Bader J, Ismail A. Survey of systematic reviews in dentistry. J Am Dent Assoc. 2004; 135:464-73.

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

[6] Shea B, Moher D, Graham I, Pham B, Tugwell P. A comparison of the quality of Cochrane reviews and systematic reviews published in paper-based journals. Eval Health Prof. 2002; 25:116-29.

[7] 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. Eur J Orthod 2013; 35(2):244-8.

[8] Major MP, Major PW, Flores-Mir C. An evaluation of search and selection methods used in dental systematic reviews published in English. J Am Dent Assoc 2006; 137:1252-7.

[9] Glenny AM, Esposito M, Coulthard P, Worthington HV. The assessment of systematic reviews in dentistry. Eur J Oral Sci 2003; 111:85-92.

[10] American Dental Association - Database of Systematic Reviews. Available at: http://ebd.ada.org/SystematicReviews.aspx. Accessed May 2012.

[11] American Dental Association - Specialty Definitions. Available at: http://www.ada.org/495.aspx. Accessed April 2012.

[12] Major MP, Major PW, Flores-Mir C. Benchmarking of reported search and selection methods of systematic reviews by dental speciality. Evid Based Dent 2007; 8:66-70.

[13] Bow S, Klassen J, Chisholm A, Tjosvold L, Thomson D, Klassen TP, et al. A descriptive analysis of child-relevant systematic reviews in the Cochrane Database of Systematic Reviews. BMC Pediatrics 2010; 10:34.

[14] Papageorgiou SN, Papadopoulos MA, Athanasiou AE. Evaluation of methodology and quality characteristics of systematic reviews in orthodontics. Orthodontics & Craniofacial Research 2011; 14:116-37.

[15] Turner L, Galipeau J, Garritty C, Manheimer E, Wieland LS, Yazdi F, et al. An Evaluation of Epidemiological and Reporting Characteristics of Complementary and Alternative Medicine (CAM) Systematic Reviews (SRs). PloS One 2013; 8:e53536.

[16] 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.

[17] Aziz T, Compton S, Nassar U, Matthews D, Ansari K, Flores-Mir C. Methodological quality and descriptive characteristics of prosthodontic-related systematic reviews. J Oral Rehabil. J Oral Rehabil. 2013; 40(4):263-78.

[18] Moher D, Tsertsvadze A. Systematic reviews: when is an update an update? Lancet 2006; 367:881-3.

[19] Higgins JP, Green S, eds. Cochrane handbook for systematic reviews of interventions version 5.0.0. Cochrane Collaboration. 2008.

[20] Edwards P, Clarke M, DiGuiseppi C, Pratap S, Roberts I, Wentz R. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. Stat Med 2002; 21:1635-40.

[21] Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. BMJ 2001; 323:101-5.

[22] Millett D. Bias in systematic reviews? J Orthod 2011; 38:158-60.

[23] Suarez-Almazor ME, Belseck E, Homik J, Dorgan M, Ramos-Remus C. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. Control Clin Trials 2000; 21:476-87.

[24] Hopewell S, Clarke M, Lusher A, Lefebvre C, Westby M. A comparison of handsearching versus MEDLINE searching to identify reports of randomized controlled trials. Stat Med 2002; 21:1625-34.

[25] Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PloS One 2008;3: e3081.

[26] Koletsi D, Karagianni A, Pandis N, Makou M, Polychronopoulou A, Eliades T. Are studies reporting significant results more likely to be published? Am J Orthod Dentofacial Orthop 2009; 136:632 e1-5.

[27] Epstein JB, Hong C, Logan RM, Barasch A, Gordon SM, Oberle-Edwards L, et al. A systematic review of orofacial pain in patients receiving cancer therapy. Support Care Cancer 2010; 18:1023-31.

[28] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009; 339:b2700.

Chapter 3

Evaluation of risk of bias assessment of trials in systematic reviews of oral health interventions, 1991-2014: a methodology study*

Humam Saltaji, Maria B. Ospina, Susan Armijo-Olivo, Shruti Agarwal, Greta G. Cummings, Maryam Amin, Carlos Flores-Mir

A copy of this chapter has been published in the Journal of the American Dental Association (JADA); Saltaji et al. J Am Dent Assoc. 2016 May 4. pii: S0002-8177(16)30299-9. doi: 10.1016/j.adaj.2016.03.017.

Abstract

Background: The authors aimed to describe how often and by what means investigators assessed the risk of bias of clinical trials in systematic reviews of oral health interventions and to identify factors associated with risk of bias assessments.

Methods: The authors selected therapeutic oral health systematic reviews published from 1991 through 2014. They extracted data related to the tools used for risk of bias assessment of primary studies and data related to other review characteristics. They descriptively analyzed the data and used multivariate logistic regression.

Results: The authors identified 1,114 oral health systematic reviews (130 Cochrane reviews and 984 non-Cochrane reviews). The investigators of the primary studies assessed risk of bias in 61.4% of the reviews, and the risk of bias assessments occurred more often in Cochrane reviews than in non-Cochrane reviews (100% versus 56.3%; P < .001) and in reviews published after the dissemination of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (odds ratio [OR], 1.55; 95% confidence interval [CI], 1.17-2.06). Compared with the investigators of reviews of public oral health interventions, investigators of reviews of oral surgery were less likely to assess risk of bias (OR, 0.41; 95% CI, 0.25-0.67). Furthermore, the investigators of systematic reviews published in dental journals were less likely to

assess risk of bias of individual trials (OR, 0.28; 95% CI, 0.19-0.41) compared with the investigators of reviews published in non-dental journals.

Conclusions: The investigators of primary studies did not undertake risk of bias assessment in a considerable portion of non-Cochrane oral health systematic reviews. The investigators of reviews published in dental journals were less likely to assess risk of bias than the investigators of reviews published in non-dental journals. The results of this study provide evidence of the need for improving the conduct and reporting of oral health systematic reviews with respect to risk of bias assessment.

Practical Implications: Clinicians should determine to what extent the findings of a systematic review are valid on the basis of whether the investigators assessed and considered risk of bias during the interpretation of findings.

3.1. Introduction

Systematic reviews and meta-analyses of randomized controlled clinical trials are considered to be a criterion standard form of evidence to indicate the efficacy and effectiveness of therapeutic interventions in health sciences [1]. The authors of systematic reviews use a comprehensive search strategy to identify all potentially relevant trials, predefine eligibility criteria to minimize the impact of bias in study selection, and use reproducible methods to assess the risk of bias found in individual trials and to consider that risk when synthesizing their results [2]. As with any research design, the value of a systematic review depends on how well its authors conduct and report the results. The endorsement by journal editors, reviewers, and authors of reporting guidelines such as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [3] has resulted in increases in both the reporting and the methodological quality of published reviews [4].

In the area of oral health, approximately 50 dentistry-related trials are published per month, and this number increases every year [5]. Similarly, the number

of systematic reviews published in oral health and within dental specialties has grown steadily over the last 2 decades [6]. Evidence from the results of methodological studies has shown that the reporting of methodological aspects of systematic reviews in oral health was below an acceptable level and that an area of particular weakness was the risk of bias assessment for the primary studies [7-9].

The extent to which clinicians can interpret and use findings from a systematic review relies heavily on the scope and internal validity of the included studies; the latter is determined largely by the extent to which the investigators who designed, conducted, and analyzed the included trials followed the highest possible standards to minimize multiple biases and thus ensured that the findings could be attributed to the intervention [2, 10]. For this reason, it is essential to critically appraise the risk of bias—a critical component to overall methodological quality—of trials included in systematic reviews that focus on therapeutic interventions [11]. Numerous tools exist to assess the risk of bias of randomized clinical trials; however, few investigators have conducted extensive testing of these tools to determine their reliability or validity [12, 13]. Because investigators have not assessed the measurement properties of these instruments, it is unknown whether or to what extent the instruments can tap the construct of risk of bias in ways that can discriminate between trials that have biased and unbiased results.

The investigators of a 2014 report [14] examined the risk of bias approaches used in periodontal systematic reviews that included meta-analysis (n=159) and found that risk of bias assessments varied greatly among the reviews. Because the investigators performed that study in 1 dental specialty only and restricted it to reviews with meta-analysis, clinicians cannot generalize the findings to reviews in other dental specialties. Consequently, whether the authors of systematic reviews of therapeutic oral health interventions more frequently assess the risk of bias of trials and which factors of systematic reviews are associated with risk of bias are still largely unknown. Thus, our objectives for this study were to describe the approaches

used by systematic reviewers of oral health interventions for assessing the risk of bias of trials and to identify potential factors associated with performing risk of bias assessment as they relate to dental specialty and publication source.

3.2. Methods

We conducted comprehensive searches of the literature in 6 electronic databases (PubMed, MEDLINE, Embase, Web of Science, Cochrane Database of Systematic Reviews [Evidence-Based Medicine Reviews], and Ovid HealthSTAR) from databases' inceptions to May 2014. We planned the search strategy with the assistance of a health sciences librarian and included a combination of index terms and key words relating to systematic reviews and oral health. The search strategy for MEDLINE can be found in the **Appendix 3.A**; we adapted the search using controlled vocabulary for each database. In addition, we searched the American Dental Association (ADA) Evidence-based Dentistry database (http://ebd.ada.org/en/evidence/systematic-reviews/) and hand searched the reference lists of potentially relevant studies that focused on the quality of systematic reviews in oral health that we identified in the main search. We did not limit the searches to articles written in the English language nor did we restrict the search with other limitations.

We included systematic reviews that examined a therapeutic intervention related to dental, oral, or craniofacial diseases as defined by the ADA scope of practice [15]. We considered a report to be a systematic review if the authors summarized the evidence from individual studies and reported methods to search, identify, and evaluate the evidence [16].

Two reviewers (either H.S. and T.K. or H.S. and S.A.) independently screened the titles and abstracts retrieved from the search strategy. We retrieved for full screening the full text of relevant systematic reviews and articles with insufficient information in the abstract. Two independent reviewers (either H.S. and T.K. or H.S.

and S.A.) determined the final eligibility of full texts, with disagreements resolved through consensus. We created a flow diagram of study selection according to the PRISMA statement (**Figure 3.1**) [3].

Two reviewers (H.S., S.A.) classified relevant systematic reviews into the following dental specialties by adapting the ADA definitions [15]: dental public health, oral and maxillofacial radiology, endodontics, oral medicine and pathology, orthodontics and dentofacial orthopedics, oral and maxillofacial surgery, periodontics, pediatric dentistry, restorative dentistry, and prosthodontics. We adapted the ADA definition [15] by adding oral medicine topics to "oral and maxillofacial pathology" and restorative dentistry topics to "prosthodontics".

We extracted the following data elements from the systematic reviews: publication year, type of review (Cochrane versus non-Cochrane), journal of publication (dental versus non-dental; we classified Cochrane reviews as being in dental publications), journal impact factor, and which methodological quality tool, risk of bias assessment tool, or both, were used. We tested double data extraction on a random sample of 20% of the reviews to assess the completeness and accuracy of the data extraction; we resolved any discrepancies by consensus.

To describe the pool of systematic reviews included, we conducted descriptive analyses (that is, proportions and percentages for categorical data such as risk of bias assessment, and mean and standard deviations [SD] or median and interquartile range [IQR] for continuous data such as year of publication, as appropriate). Then we grouped systematic reviews according to whether the investigators of the primary studies had assessed risk of bias. We implemented multivariate logistic regression to explore the associations between risk of bias assessment and the characteristics of the review publication: journal impact factor, journal of publication (dental versus non-dental), time of publication (that is, before or after the publication of the PRISMA statement [3]), and dental specialty. We reported odds ratios (OR) with 95%

confidence intervals (CI), and we set the statistical significance at $P < .05$. We performed statistical analyses using Stata Version 14.0 (StataCorp).

3.3. Results

The search strategy identified 8,076 titles and abstracts, of which we judged 1,878 articles to be potentially relevant; of these, we determined that 1,114 articles satisfied the eligibility criteria (**Figure 3.1**). The complete list of excluded articles is available on request. **Figure 3.2** illustrates the number of systematic reviews of oral heath interventions published from 1991 through 2014.

Of the 1,114 systematic reviews whose authors assessed therapeutic interventions in oral health and were published from 1991 through 2014 (median year of publication, 2009; IQR, 2006-2012), 78.8% were published in dental journals and 21.2% were published in other health science journals. The median impact factor of the journals of publication was 1.99 (IQR, 1.27-3.12). Overall, 88.3% were not Cochrane reviews, and 11.7% were Cochrane reviews.

Most systematic reviews were published in the fields of prosthodontics and restorative dentistry (20.4%), periodontics (18.8%), and oral and maxillofacial surgery (15.8%). Other dental specialties represented included dental public health (12.4%), orthodontics and dentofacial orthopedics (12.0%), oral medicine and pathology (11.6%), endodontics (4.8%), pediatric dentistry (4.0%), and oral and maxillofacial radiology (0.2%). **Table 3.1** provides details of the numbers of Cochrane and non-Cochrane systematic reviews in oral health therapeutics by specialty.

The proportion of oral health systematic reviews published before (54.6%) and after (45.4%) the publication of the PRISMA statement [3] in 2009 was relatively similar. We found that investigators had conducted risk of bias assessments of individual trials in 61.4% (n = 684) of reviews, and that this type of assessment had

occurred more often in Cochrane reviews than in non-Cochrane reviews (100% versus 56.3%; P < .001). **Table 3.2** provides details of the numbers of reviews with completed risk of bias assessments by year of publication. Overall, the frequency of risk of bias assessment ranged from 76.1% among systematic reviews in dental public health to 39.2% among systematic reviews in prosthodontics and restorative dentistry. **Figure 3.3** illustrates the distribution of risk of bias assessment by dental specialty.

Among the systematic reviews whose investigators had assessed the risk of bias of the individual studies (n = 684), the investigators of 43.2% of these systematic reviews had used tools for which the measurement properties had been assessed formally in the scientific literature [7,17]. Of these, most (47.1%) had used the Cochrane Risk of Bias tool, [10] individual items recommended in the Cochrane Handbook for Systematic Reviews of Interventions [2, 18] (26.5%), or the Jadad scale [19] (15.3%). The investigators of 38% of reviews had used unvalidated trial risk of bias assessment consisting of items extracted from a variety of tools, whereas the investigators of 18.8% of reviews had assessed trials by using risk of bias instruments that were not designed explicitly for trial evaluation. **Table 3.3** [1, 2, 10, 18-22] provides details of the risk of bias approaches used in the therapeutic oral health reviews.

Results of the logistic regression analyses showed that the investigators of systematic reviews that were published after the dissemination of the PRISMA statement [3] in 2009 were more likely to have assessed the risk of bias of individual trials (OR, 1.55; 95% CI, 1.17-2.06).

Compared with the investigators of systematic reviews of public oral health interventions, the investigators of systematic reviews that examined interventions in the following dental specialties were less likely to have conducted risk of bias assessment of individual trials: oral and maxillofacial surgery and radiology (OR, 0.41; 95% CI, 0.25-0.67), orthodontics and dentofacial orthopedics (OR, 0.62; 95%

CI, 0.37-1.04), periodontics (OR, 0.53; 95% CI, 0.33-0.86), and prosthodontics and restorative dentistry (OR, 0.24; 95% CI, 0.16-0.38).

The investigators of systematic reviews that were published in dental journals were less likely to have assessed the risk of bias of individual trials (OR, 0.28; 95% CI, 0.19-0.41) compared with the investigators of reviews published in other health science journals. Finally, the investigators of systematic reviews that were published in journals with impact factors above the median impact factor of publication of this sample of reviews (1.9 for dental journals and 5.9 for non-dental journals) were less likely to have assessed the quality of trials (OR, 0.53; 95% CI, 0.40-0.70). **Table 3.4** [3] provides details of the regression models for assessing the risk of bias of primary studies.

3.4. Discussion

Assessing the risk of bias of primary trials is an essential step when synthesizing and interpreting evidence in systematic reviews of therapeutic interventions. To the best of our knowledge, this is the first methodological study in the domain of oral health research to evaluate the extent of risk of bias assessments of trials in therapeutic systematic reviews and of the factors associated with the risk of bias assessment.

When examining risk of bias approaches used in 159 periodontal reviews, Faggion and colleagues [14] reported that 28% of the reviews' investigators had used domain-based tools and 26% of the reviews' investigators had used more than 1 tool, whereas only 39% of the reviews' investigators had used validated tools. However, the authors of this study [14] restricted the study selection to reviews that included meta-analysis and that had been conducted in the field of periodontology. Our study results showed that investigators had assessed risk of bias in primary studies in almost two-thirds of the reviews and that assessments had occurred more often in Cochrane reviews than in non-Cochrane reviews. The fact that the investigators of only two-

thirds of the oral heath reviews had assessed the risk of bias is concerning for clinicians who interpret the results because of the limited ability to assess how the estimates of effect may have been biased owing to study conduct; therefore, clinical decision making may not be made on the basis of valid findings provided by the best evidence reviewed. The use of risk of bias assessments calculated in this study for oral heath interventions (61.40% for all of the reviews, and 66.67% for reviews published in 2014) is less than that reported in a 2016 study (71.8%; n = 222) [23] in which investigators examined a sample of 309 therapeutic and nontherapeutic systematic reviews published in the domain of medical research in a 3-month period during 2014. This suggests that research in the dental fields is relatively falling behind medical research standards.

Our results revealed that the Cochrane risk of bias tool [10] is the most commonly used tool for assessing the risk of bias of studies included in oral health reviews. The proportion of systematic reviews whose investigators used this tool (47.1%) is higher than that reported by Seehra and colleagues [23] (26.1%; n = 58) in their study of medical research systematic reviews. In 2008, the Cochrane Collaboration proposed the risk of bias tool, which included 6 methodological domains [10]: sequence generation, allocation concealment, masking, incomplete outcome data, selective outcome reporting, and other sources of bias. Since its inception, the Cochrane risk of bias tool has evolved continually. Although the Cochrane tool has played an important role in leading the way in improving risk of bias assessment in health research, the tool requires further development and improvement [24]. The developers of the Cochrane risk of bias tool highlighted that the tool's domains need to be expanded on the basis of its use in different health areas [25]. They [25] and others [26-28] called for more meta-epidemiologic studies in a wider range of health disciplines to support existing domains, add new domains, or both. The tool's criterion validity has been tested using the Jadad scale, [19] Schulz allocation concealment, [29,30] and the Effective Public Health Practice Project

Quality Assessment Tool [28]. However, investigators have reported the inter-rater reliability of this tool as poor [27,30]. Some methodological reports have suggested the need for additional evaluations of the Cochrane risk of bias tool [10] in a wide range of health research areas with a need for additional testing related to its reliability and validity [28,29]. Other investigators have recommended developing more consistent and clearer guidelines for scoring for the Cochrane risk of bias tool [10,31].

The Jadad scale [19] was the second most commonly used tool; it was used in 15.3% of the reviews. This psychometrically evaluated tool includes 3 items: randomization, double-masking, and a description of dropouts and withdrawals. However, double-masking accounts for 40% of the total score when using this tool, which was developed to assess pharmacologic trials; the high value placed on double-masking may not be as useful for oral health trials involving surgical or device interventions, such as orthodontic trials, in which masking patients is not feasible. In addition, allocation concealment is not included in this tool despite its well-known relationship with the internal validity of randomized clinical trials [32-34]. Furthermore, investigators have reported that the use of the Jadad scale [19] has been associated with inadequate responsiveness to differentiate between different quality grades [35] Thus, future investigators should re-evaluate the wide use and the validity of the Jadad scale [19] in oral health systematic reviews.

The results revealed that investigators used unvalidated risk of bias assessment tools (consisting of items extracted from a variety of tools) in 38% of reviews. Modification of risk of bias tools is likely to affect the validity and applicability of the results. Without having validation of the newly developed tool or group of quality items, clinicians could question the interpretation of the review findings. This is especially critical when using an overall quality and risk of bias score, which may differ conceptually among tools (for example, placing more or less weight on masking and leaving out concealment). The individual weighting of scored items should undergo a validation process.

Furthermore, by examining our study results, we determined that although investigators have used a number of risk of bias tools, there is no tool that is specifically designed for assessing the quality of oral health trials. There is also a lack of empirical evidence demonstrating how the results of any assessments, particularly with respect to distinct constructs in the tools, relate to the over- or underestimation of estimates of treatment effects within oral health trials [36]. Moreover, using different tools to assess the oral health trial's quality could lead to different results and interpretations, which ultimately could affect recommendations for oral health care.

The investigators of reviews published in dental journals were less likely than the investigators of reviews published in non-dental journals to have assessed the risk of bias of individual trials. This raises concerns regarding the quality of evidence stemming from systematic reviews published in dental journals compared with journals published in other medical fields. These findings call for dental journal editors to pay close attention to the assessment of risk of bias of primary trials in systematic reviews, given that 1 of the main aims of systematic reviews is to evaluate the quality of the evidence and inform clinical practice guidelines to provide accurate recommendations for clinical practice [16].

The risk of bias assessment of individual studies in a systematic review is a key component of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) [37] approach to evaluate the strength of a body of evidence. Importantly, research into the synthesis of dental knowledge would be enhanced by establishing a clear definition of terms as they pertain to the evaluation of risk of bias (that is, determining the extent to which study results are close to the truth), methodological quality (that is, conducting research by using the highest possible standards), and quality of the evidence (that is, having the confidence that an estimate of the effect is near the true value for an outcome across studies) [2]. Adopting the GRADE approach in the synthesis of evidence in dentistry would be a major step

toward informing clinical dental practice on the basis of the analysis of the quality of the evidence and the strength of recommendations from dental systematic reviews.

The results of our study demonstrated variations across the 9 dental specialties; clearly, improvement in the conduct and reporting of systematic reviews in specific dental specialties (such as prosthodontics and restorative dentistry and oral and maxillofacial surgery) is needed. Furthermore, reviews published in journals with higher impact factors were less likely to assess the quality of trials. These findings, which may seem surprising at first, could be attributed to the fact that the impact factor has a limited utility and is not always a valid measure of research quality [38]. Moreover, this unexpected association between absence of risk assessment and higher impact factor could be explained by noting the different impact factors of journals in different dental specialties. For example, the 2014 impact factor (according to Journal Citation Reports) of the European Journal of Orthodontics (the highest-ranking journal in the specialty of orthodontics) is 1.483; this rating is much lower than the impact factor of the highest-ranking journal (4.01) in the specialty of periodontics (Journal of Clinical Periodontology) [39].

Clinicians can draw several implications from our study. From a methodological perspective, authors of systematic reviews need to use tested or validated items and assessment tools when assessing risk of bias of individual studies, and they should explicitly report the results for each quality item or risk of bias domain. They should state what domain they are considering to be the most important for their assessment of quality and interpretation of results and how they are condensing individual items into a final score. The use of items from different risk of bias assessment tools may be more acceptable than using an overall score for some oral health trials as long as the items are linked to important potential biases. Some study investigators [40, 41] have criticized the common use of summary quality scores, in which trials receive points for criteria met. The impact of quality criteria (such as randomization) could be weakened by the summary quality score, specifically

if the summary score includes criteria not related to potential bias [42]. Also, systematic review authors should consider conducting sensitivity analysis in a meta-analysis on the basis of the risk of bias of the included studies. Peer reviewers and editors of dental journals should require authors of systematic reviews to adhere to the PRISMA guidelines [3] and insist that authors show adequate conduct and reporting of risk of bias assessment.

Although we determined that no tool was identified as being specifically designed for assessing the risk of bias of oral health trials, we did note that using the Cochrane risk of bias tool, [10] which was developed based on empirical evidence of associations between methodological characteristics and treatment effect size, has potential value for oral health systematic reviews. Thus, while acknowledging that the Cochrane risk of bias tool [10] is not without its problems, the tool could be considered the best available approach to assessing risk of bias in oral health systematic reviews. The Cochrane risk of bias tool [10] should continue to be improved to facilitate assessing the risk of bias. Furthermore, systematic review authors should take into consideration the potential design characteristics of oral health trials. Compared with randomized clinical trials in other fields, oral health randomized clinical trials tend to have some unique design characteristics, such as the use of a broad range of concomitant interventions (surgical, nonsurgical, drug, and nondrug), difficulty in applying masking, and a common use of split-mouth, crossover, or cluster designs. These features can add complexity with respect to reporting and applying strategies that reduce the increased potential for biases threatening a study's internal validity, and, hence, also a study's external validity.

From a clinical perspective, clinicians should have the knowledge to correctly identify the type of primary studies, adequately appraise the quality of included studies, and effectively take into consideration the risk of bias assessment when reading systematic review reports and interpreting their findings. They should determine to what extent the findings of a systematic review are valid on the basis of

whether the investigators had assessed risk of bias and how the investigators considered that bias during interpretation of findings. Relying on reviews of data from high-quality studies may help ensure that clinicians will achieve the best possible results for their patients and practices.

3.5. Conclusions

Our study results identified that the investigators of a considerable portion of non-Cochrane oral health systematic reviews published from 1991 through 2014 did not assess the risk of bias of primary studies. Investigators of reviews published in dental journals were less likely to have assessed the risk of bias of individual trials than the investigators of reviews published in non-dental journals. The Cochrane risk of bias tool was used most commonly; however, we did not identify the use of a tool that was designed specifically for assessing the methodological quality of oral health trials. The results of our methodological study provide evidence for the need for oral health systematic review authors to improve the conduct and reporting of risk of bias assessment, and for dental journal reviewers and editors to insist on adequacy in these areas.

Table 3.1. The number of Cochrane and non-Cochrane oral health systematic reviews by specialty

Dental Specialty	Non-Cochrane Reviews, % (n)	Cochrane Reviews, % (n)
Dental public health	11.8 (116)	16.9 (22)
Endodontics	4.9 (48)	4.6 (6)
Oral and maxillofacial radiology	0.2 (2)	0.0 (0)
Oral and maxillofacial surgery	15.4 (152)	18.5 (24)
Oral medicine and pathology	10.8 (106)	17.7 (23)
Orthodontics and dentofacial orthopedics	12.1 (119)	11.5 (15)
Pediatric dentistry	3.7 (36)	6.9 (9)
Periodontics	19.9 (196)	10.0 (13)
Prosthodontics and restorative dentistry	21.2 (209)	13.8 (18)
Total	100 (984)	100 (130)

Table 3.2. The number of non-Cochrane reviews with completed risk of bias assessments by year of publication

Year of Publication	Non-Cochrane* Reviews, % (n) (n = 984)
1991-1995	18.18 (2)
1996-2000	42.86 (18)
2001	36.00 (9)
2002	58.62 (17)
2003	59.52 (25)
2004	43.59 (17)
2005	43.59 (17)
2006	63.16 (36)
2007	46.38 (32)
2008	46.25 (37)
2009	46.74 (43)
2010	61.74 (71)
2011	70.59 (60)
2012	64.29 (72)
2013	66.67 (68)
2014	66.67 (30)
Total	56.3 (554)

* Cochrane reviews (n = 130) were not considered in the analysis because risk of bias was assessed in all of the Cochrane reviews.

Table 3.3. Frequency of risk of bias assessment tools and approaches in oral health systematic reviews

Risk of Bias Assessment Tool or Approach	Systematic Reviews, No. (%) (n = 1,114)
No Risk of Bias Assessment	430 (38.6)
Chalmers Tool*	6 (0.5)
Cochrane Risk of Bias Tool†	139 (12.5)
Consolidated Statement for Reporting Trials Statement‡	15 (1.3)
Delphi List§	3 (0.3)
Hadorn Criteria¶	5 (0.4)
Items from the Cochrane Handbook for Systematic Reviews #	79 (7.1)
Items or Checklist Adapted from More Than 1 Tool	68 (6.1)
Items or Checklist Based on Authors' Knowledge (Without Reporting a Reference)	149 (13.4)
Items or Checklist Used in a Previous Review	43 (3.9)
Jadad Scale**	45 (4.0)
More Than 1 Tool Based on Study Design of Selected Studies	17 (1.5)
Other (Formal Randomized Controlled Trial Tool)	3 (0.3)
Other (Nonrandomized Controlled Trial or Diverse Study Design Tool)	112 (10.1)

* Source: Chalmers and colleagues [20].
† Sources: Higgins and colleagues [10] and Higgins and Altman [18].
‡ Source: Altman and colleagues [1].
§ Source: Verhagen and colleagues [21].
¶ Source: Hadorn and colleagues [22].
# Sources: Higgins and Green [2] and Higgins and Altman [18].
** Source: Jadad and colleague [19].

Table 3.4. Multivariate logistic regression model (factors associated with assessing risk of bias in oral health systematic reviews)

Variable	Assessing Risk of Bias	
	Odds Ratio (95% Confidence Interval)	*P* Value
Type of Journal		
Non-dental (Medical)	1.00	
Dental	0.28 (0.19-0.41)	< 0.001
Journal Impact Factor		
Below the median impact factor	1.00	
Above the median impact factor	0.53 (0.40-0.70)	< 0.001
Time of Publication		
Before PRISMA* statement publication	1.00	
After PRISMA statement publication	1.55 (1.17-2.06)	0.002
Dental Specialty		
Dental public health and pediatric dentistry	1.00	
Periodontics	0.53 (0.33-0.86)	0.010
Orthodontics and dentofacial orthopedics	0.62 (0.37-1.04)	0.077
Endodontics, prosthodontics and restorative dentistry	0.24 (0.16-0.38)	< 0.001
Oral and maxillofacial surgery	0.41 (0.25-0.67)	< 0.001
Oral medicine and pathology	0.63 (0.37-1.07)	0.090

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

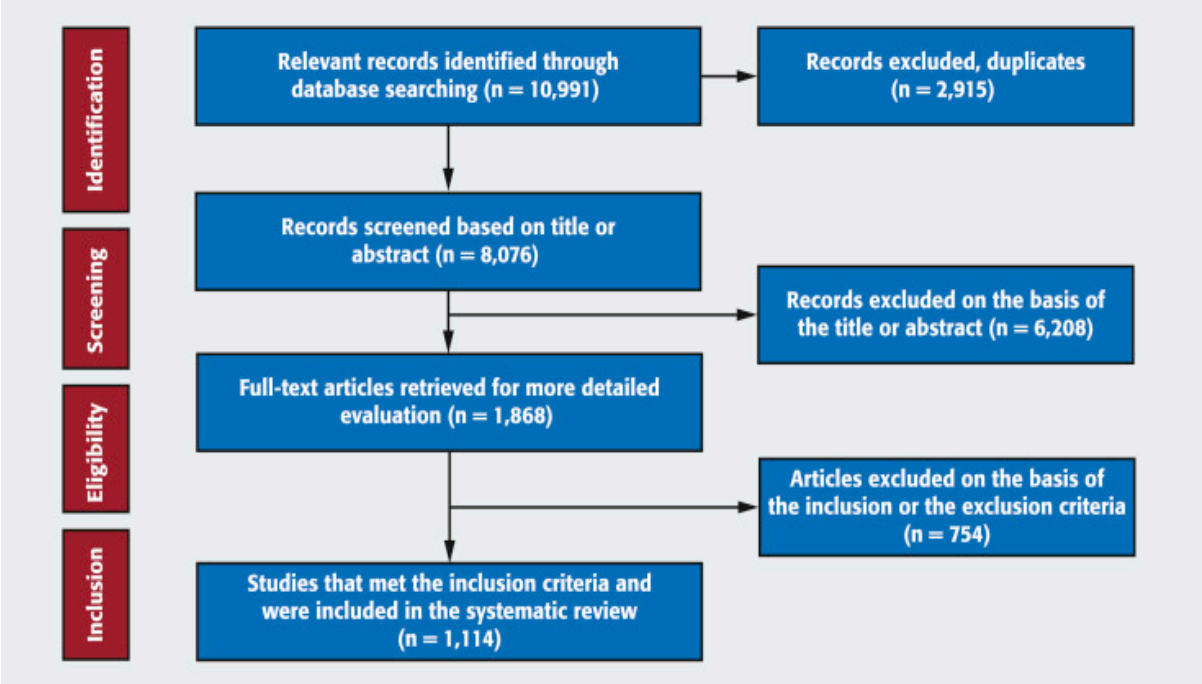

Figure 3.1. Flow diagram of the literature search according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement. Source: Moher and colleagues [3].

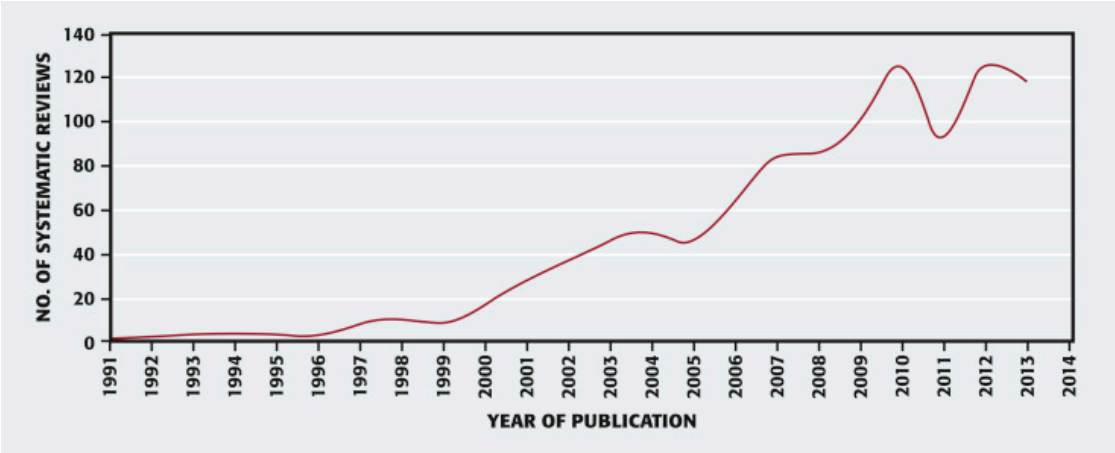

Figure 3.2. Number of systematic reviews of oral heath interventions published from 1991 through 2014.

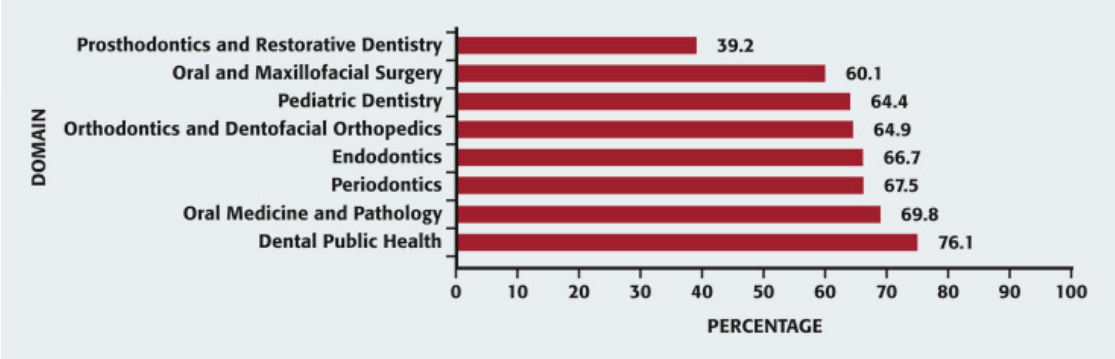

Figure 3.3. Distribution of completed risk of bias assessments by dental specialty. Data for oral and maxillofacial radiology are not shown as no therapeutic reviews were found for this subject.

References

[1] Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001;134(8):663-694.

[2] Higgins JP, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.0. Cochrane Collaboration; 2008.

[3] 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(10):1006-1012.

[4] Panic N, Leoncini E, de Belvis G, Ricciardi W, Boccia S. Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. PloS One 2013;8(12):e83138.

[5] Needleman I, Worthington H, Moher D, Schulz K, Altman DG. Improving the completeness and transparency of reports of randomized trials in oral health: the CONSORT statement. Am J Dent 2008;21(1):7-12.

[6] Saltaji H, Cummings GG, Armijo-Olivo S, et al. A descriptive analysis of oral health systematic reviews published 1991-2012: cross sectional study. PloS One 2013;8(9):e74545.

[7] Pandis N, Fleming PS, Worthington H, Salanti G. The quality of the evidence according to GRADE is predominantly low or very low in oral health systematic reviews. PLoS One 2015;10(7):e0131644.

[8] Lempesi E, Koletsi D, Fleming PS, Pandis N. The reporting quality of randomized controlled trials in orthodontics. J Evid Based Dent Pract 2014; 14(2):46-52.

[9] 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(12):1181-1187.

[10] Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343: d5928.

[11] Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. BMJ 2001;323(7303):42-46.

[12] Armijo-Olivo S, Fuentes CJ, Ospina M, Saltaji H, Hartling L. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: a descriptive analysis. BMC Med Res Methodol 2013;13:116.

[13] Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials 1995;16(1):62-73.

[14] Faggion CM, Huda F, Wasiak J. Use of methodological tools for assessing the quality of studies in periodontology and implant dentistry: a systematic review. J Clin Periodontol 2014;41(6):625-631.

[15] American Dental Association. Specialty definitions: definitions of recognized dental specialties. Available at: http://www.ada.org/495.aspx. Accessed April 1, 2016.

[16] Cook DA, West CP. Conducting systematic reviews in medical education: a stepwise approach. Med Educ 2012;46(10):943-952.

[17] Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. Phys Ther 2008;88(2):156-175.

[18] Higgins J, Altman DG. Assessing risk of bias in included studies (Chapter 8). In: Higgins J, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 (updated March 2011). Available at: http://handbook.cochrane.org/. Accessed April 1, 2016.

[19] Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996;17(1):1-12.

[20] Chalmers TC, Smith H Jr, Blackburn B, et al. A method for assessing the quality of a randomized control trial. Control Clin Trials 1981;2(1):31-49.

[21] Verhagen AP, de Vet HC, de Bie RA, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 1998; 51(12):1235-1241.

[22] Hadorn DC, Baker D, Hodges JS, Hicks N. Rating the quality of evidence for clinical practice guidelines. J Clin Epidemiol 1996;49(7):749-754.

[23] Seehra J, Pandis N, Koletsi D, Fleming PS. Use of quality assessment tools in systematic reviews was varied and inconsistent. J Clin Epidemiol 2016;69:179-184.e5.

[24] Hrobjartsson A, Boutron I, Turner L, Altman DG, Moher D. Assessing risk of bias in randomised clinical trials included in Cochrane Reviews: the why is easy, the how is a challenge. Cochrane Database Syst Rev 2013;4:ED000058.

[25] Savovic J, Weeks L, Sterne JA, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. Syst Rev 2014;3:37.

[26] Hartling L, Hamm MP, Milne A, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. J Clin Epidemiol 2012;66(9):973-981.

[27] Armijo-Olivo S, Ospina M, da Costa BR, et al. Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. PloS One 2014;9(5): e96920.

[28] Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool—methodological research. J Eval Clin Pract 2012;18(1):12-18.

[29] Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. BMJ 2009;339:b4102.

[30] Hartling L, Bond K, Vandermeer B, et al. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. PloS One 2011;6(2):e17242.

[31] Boutron I, Ravaud P. Classification systems to improve assessment of risk of bias. J Clin Epidemiol 2012;65(3):236-238.

[32] Nuesch E, Reichenbach S, Trelle S, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. Arthritis Rheum 2009;61(12):1633-1641.

[33] Savovic J, Jones HE, Altman DG, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med 2012;157(6):429-438.

[34] Armijo-Olivo S, Cummings GG, Fuentes J, et al. Identifying items to assess methodological quality in physical therapy trials: a factor analysis. Phys Ther 2014;94(9):1272-1284.

[35] Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. J Clin Epidemiol 2006;59(12):1249-1256.

[36] Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, Flores-Mir C. Methodological characteristics and treatment effect sizes in oral health randomised controlled trials: is there a relationship? Protocol for a meta-epidemiological study. BMJ Open 2014;4(2):e004527.

[37] Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336(7650):924-926.

[38] Moustafa K. The disaster of the impact factor. Sci Eng Ethics 2015;21(1):139-142.

[39] 2014 Journal Citation Reports Science Edition. Thomson Reuters, 2015.

[40] Greenland S. Quality scores are useless and potentially misleading: reply to “re: a critical look at some popular analytic methods.” Am J Epidemiol 1994;140(3):300-301.

[41] Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282(11): 1054-1060.

[42] Armijo-Olivo S, da Costa BR, Cummings GG, et al. PEDro or Cochrane to assess the quality of clinical trials? A meta-epidemiological study. PloS One 2015;10(7):e0132634.

Chapter 4

Methodological characteristics and treatment effect sizes in oral health randomised controlled trials: Is there a relationship? Protocol for a meta-epidemiological study*

Humam Saltaji, Susan Armijo-Olivo, Greta G. Cummings, Maryam Amin, Paul W. Major, Carlos Flores-Mir

A copy of this chapter has been published in the BMJ Open journal; Saltaji et al. BMJ Open 2014;4:e004527. doi:10.1136/bmjopen-2013- 004527

Abstract

Introduction: It is fundamental that randomized controlled trials (RCTs) are properly conducted in order to reach well-supported conclusions. However, there is emerging evidence that RCTs are subject to biases which can overestimate or underestimate the true treatment effect, due to flaws in the study design characteristics of such trials. The extent to which this holds true in oral health RCTs, which have some unique design characteristics compared to RCTs in other health fields, is unclear. As such, we aim to examine the empirical evidence quantifying the extent of bias associated with methodological and non-methodological characteristics in oral health RCTs.

Methods and analysis: We plan to perform a meta-epidemiological study, where a sample size of 60 meta-analyses (MAs) including approximately 600 RCTs will be selected. The MAs will be randomly obtained from the Oral Health Database of Systematic Reviews using a random number table; and will be considered for inclusion if they include a minimum of five RCTs, and examine a therapeutic intervention related to one of the recognised dental specialties. RCTs identified in selected MAs will be subsequently included if their study design includes a comparison between an intervention group and a placebo group or another intervention group. Data will be extracted from selected trials included in MAs based

on a number of methodological and non-methodological characteristics. Moreover, the risk of bias will be assessed using the Cochrane Risk of Bias tool. Effect size estimates and measures of variability for the main outcome will be extracted from each RCT included in selected MAs, and a two-level analysis will be conducted using a meta-meta-analytic approach with a random effects model to allow for intra-MA and inter-MA heterogeneity.

Ethics and dissemination: The intended audiences of the findings will include dental clinicians, oral health researchers, policymakers and graduate students. The aforementioned will be introduced to the findings through workshops, seminars, round table discussions and targeted individual meetings. Other opportunities for knowledge transfer will be pursued such as key dental conferences. Finally, the results will be published as a scientific report in a dental peer-reviewed journal.

4.1. Introduction

Current scientific knowledge for clinical research is based on randomised control trials (RCTs) that have been synthesised in systematic reviews (SRs) and meta-analyses (MAs), which together comprise the 'gold standard' of scientific evidence [1, 2]. The abundance of RCTs is continually increasing with about 50 new clinical trials published every month in the field of dentistry [3]. Since these sources are considered the highest level of evidence for the efficacy of treatment interventions, the information gathered from them is used to guide clinical practice and policy decisions [4].

In the field of oral health research, several investigations have assessed the methodological/reporting quality of oral health RCTs [5-8], and examined important aspects related to the conduct and reporting of these trials such as: clustering effects [9], reporting statistical findings [10], sample size justification [11] and randomisation process [12, 13]. Nonetheless, the quality of these forms of evidence has not yet been

fully scrutinised to identify design flaws and their impact on treatment effect estimates in the field of oral health research [3, 14].

Ideally, RCTs should be properly conducted and accurately reported in order to reach well-supported conclusions for decision–making that are both valid and generalisable to patients who will receive the interventions in clinical practice [4, 15]. However, evidence is emerging that some RCTs are biased and overestimate the magnitude of the effect size due to flaws in their design and/or reported study characteristics [15, 16]. These studies will likely skew the overall conclusions of MAs once pooled, possibly leading to faulty treatment decisions [2, 17]. Generally, the poor methodological quality of these RCTs has resulted in the tendency to exaggerate or overestimate the true treatment effect (effect size) [18]. Among the flawed characteristics observed in previously published reports and found to have an impact on the true treatment effect were: the lack of randomisation and concealment [19, 20], inadequate blinding [16, 20, 21], and industrial funding [22], although not all the studies confirmed these associations [23, 24]. This could lead clinicians to the implementation of treatment choices, which might be inappropriate or ineffective and have negative effects on treatment outcomes in dental practices [3].

The assumption of the association between trial quality and true effect size is derived from published 'meta-epidemiological' studies, which are investigations that quantify the extent of bias in the effect sizes related to trial quality in a group of meta-analyses [25]. There are a few meta-epidemiological investigations that have been conducted in the field of medicine [16, 18, 23–26]; however, the value of the conclusions of some of these investigations to other healthcare fields is limited by numerous factors including: the examination of quality items that were related to reporting quality and not methodological quality or bias [23, 25, 26], the failure to examine continuous outcomes (which occurred in the majority of studies) [16, 23, 25] based on a preference for evaluating dichotomous outcomes which could limit applying their conclusions to RCTs with continuous outcomes, and the presentation of

inconsistent findings regarding the methodological criteria that are associated with effect size [16, 23].

More importantly, meta-epidemiological studies reported that bias in effect size associated with methodological characteristics may vary between medical fields or different areas of health research [27] and differ based on the type of intervention [20, 25]. To our current knowledge, no meta-epidemiological study has been conducted in any of the nine specialised dental fields that examined the extent of bias related to the quality of oral health RCTs. It is unclear to what extent this holds true in oral health RCTs, which have some unique design characteristics compared with RCTs in other health fields, such as: difficulty in applying blinding, use of a broad range of different interventions (surgical, nonsurgical, drug and non-drug) [28], use of multiple outcomes, common use of split-mouth and crossover designs and clustering effects [9, 29], which make the evaluation of these trials more challenging compared to trials in other health areas.

As such, the purpose of the proposed study is to provide a first step in the development of a research framework for appraising, reporting and conducting RCTs in oral health research. The objectives are to: (1) examine the empirical evidence for associations between methodological trial characteristics (eg, adequacy of randomisation, adequacy of allocation concealment, baseline comparability, blinding of assessors and participants, similarity of co-interventions, adequacy of compliance to the treatment, among others) and treatment effect estimates (effect sizes) in oral health RCTs and (2) determine if other non-methodological study characteristics (eg, the nature of intervention, specialty, type of outcomes, number of centres, type of funding, among others) are associated with effect sizes in oral health RCTs.

The hypothesis of the proposed study is that there is no difference between treatment effect estimates (effect sizes) for oral health trials meeting certain methodological quality characteristics versus trials not meeting those quality

characteristics, such as: adequacy of randomisation, adequacy of allocation concealment, baseline comparability, blinding of assessors and participants, similarity of co-interventions, similarity of outcome assessment, description of withdrawals and adequacy of compliance to the treatment, among others; and that trials with different non-methodological characteristics such as nature of intervention, type of outcomes, study design, number of centres, type of funding, sample size and speciality, among others, will not have different treatment effect estimates (effect sizes).

4.2. Methods and Analysis

4.2.1. Design

A meta-epidemiological study.

4.2.2. Study selection

Selection of MAs

MAs/SRs will be included if they fulfil the following inclusion criteria:

1. Reports should be therapeutic oral health MAs defined as reviews that examined therapeutic interventions related to dental/oral diseases as defined by the American Dental Association (ADA) scope of practice [28, 30]. Reports will be considered as MAs if they explicitly identified and summarized evidence from several published reports through quantitative analyses [31, 32].
2. MAs should include a minimum of five RCTs and provide quantitative data of treatment estimates.
3. The MA should examine at least one continuous outcome.
4. MAs should be full-length reports.

Selection of RCTs

All RCTs included in selected MAs will be eligible if they meet the following inclusion–exclusion criteria:

1. The study design is reported to be an RCT [33];
2. Comparison is between an intervention and a placebo or another intervention;
3. RCTs evaluated a therapeutic intervention related to one of the dental specialties defined by the ADA [30];

RCTs will be excluded if the results are reported in a way that does not allow for effect sizes to be calculated.

4.2.3. Sample size calculation

Previously published meta-epidemiological studies have been reported to be underpowered because of their small sample sizes and their highly heterogeneous samples [27]. Accordingly, it has been suggested compiling a set of RCTs that are specific to clinical fields to decrease heterogeneity and improve power of meta-epidemiological reports. Our study will focus on oral health RCTs published in the recognised nine dental specialties. Furthermore, a minimum of 60 MAs containing approximately 600 RCTs will be assessed for this meta-epidemiological study. Given the previous reports [34, 35], it could be anticipated that a difference in effect sizes of at least 0.15 will be obtained between trials with and without selected quality domains. This magnitude of difference has been argued to correspond to one-quarter to one-half of the typical treatment effect found for interventions in areas similar to dentistry. Thus, this difference should also be relevant to the field of dentistry.

As such, we planned a sample size of 60 MAs expecting that a sample of 600 RCTs would come from these MAs. To the best of our knowledge, this number of RCTs will represent the largest sample size examined in a meta-epidemiological study aimed at examining bias related to trial quality using continuous outcomes. The sample of MAs will be selected from the Oral Health Database of Systematic Reviews

[28], developed by the authors and include all of the oral health SRs published between 1999 and 2012, encompassing the nine dental specialties defined by the ADA [30]. This database contains 153 MAs (39 Cochrane and 114 non-Cochrane), out of the 1188 SRs included in the database, which potentially meet the inclusion–exclusion criteria for this study. **Figure 4.1** provides further details of the SRs identified in the database and the number of MAs potentially meeting the eligibility criteria for this study, within each of the nine dental specialties and for Cochrane and non-Cochrane SRs separately.

4.2.4. Data extraction

A data extraction template will be designed in Microsoft ACCESS and pilot tested. With regard to assessors, a panel of assessors from varied health research backgrounds will perform data extraction. One of the team members will perform the training for all assessors and will make sure that all of them have a clear understanding of the data extraction process. Training of these assessors will be carried out through 10–15 separate articles, not included in the set of articles to be reviewed. Each of the 10–15 training articles will be independently reviewed by all the members of the review panel and then discussed by the panel. In order to ensure good agreement between the assessors, the training exercise will be repeated to address any issues identified in the first exercise. For actual data extraction, two assessors will independently complete data extraction with a consensus meeting utilised to resolve any disagreement between the assessors. If a consensus could not be achieved, then the two assessors will consult with a third assessor (HS or SA-O) to achieve full consensus, and only consensus answers will be used for all analyses.

Data will be extracted on the following items:

<u>Non-methodological characteristics</u>

Dental speciality (eg, dental public health, endodontics, oral medicine and oral pathology, oral and maxillofacial radiology, oral and maxillofacial surgery, orthodontics and dentofacial orthopaedics, pediatric dentistry, periodontics, prosthodontics and restorative dentistry), year of publication, source of funding (eg, industry, government, foundation and academic), type of intervention (eg, drug, surgical, device, dental material, psychological, educational and policy), number of randomized groups, number of centres (eg, multicentre and single centre), study design (eg, parallel, crossover and factorial) and type of outcome (eg, subjective and objective).

Methodological characteristics

Methodological characteristics will be based on preliminary work performed by our research team and will be extracted from commonly used tools to evaluate the methodological quality of RCT in health research [36], such as: adequacy of randomisation, adequacy of allocation concealment, baseline comparability, blinding of assessors and participants, similarity of co-interventions, similarity of outcome assessment, description of withdrawals and adequacy of compliance to the treatment. Guidelines for decision-making will be formed based on the previous work of our team, in order to increase consistency [36, 37].

Risk of bias

Risk of bias will be assessed using the Cochrane risk of bias tool as ‘high’, ‘low’ or ‘unclear’ and will follow the guidelines of the Cochrane Collaboration for scoring clinical trials [33].

Treatment effect estimates

Treatment effect estimates, measures of variability (SDs and 95% CIs) and respective sample sizes will be extracted for the main outcome.

4.2.5. Data analysis

Descriptive statistics (means and SDs) will be extracted for one continuous outcome per RCT. RCTs that are included in more than one MA will be eligible for inclusion in the study only once, and will be extracted from the MA with the fewer number of RCTs. STATA statistical software V.12 will be employed to perform the planned statistical analyses.

We propose to use a two-level meta-meta-analytic approach with the use of a random effects model. This analysis will permit the evaluation of the heterogeneity intra-MAs and inter-MAs [38]. The first step will consist in obtaining the standardised effect size estimates for the primary outcome of each trial using the guidelines established by Cohen [39]. The second step will involve pooling the results of the previous analysis, using a combined difference approach, to demonstrate the different components of MAs across all MAs. Moreover, the data acquired will be used to evaluate certain components of the methodological assessment—such as allocation concealment, randomisation, blinding, etc—and will subsequently be used to divide the data set into two groups: one group having adequately addressed the said component, and the other not addressing it. Thus, for each MA, we will conduct meta-regression techniques to derive the difference between pooled estimates from trials with (eg, allocation concealment) and without the characteristic of interest. Formal tests of interaction will be performed separately for each MA based on Z scores for estimated differences in effect sizes between trials with and without the characteristic of interest and the corresponding SEs. Therefore, two pooled effect sizes will be calculated for each MA. The effect sizes at this stage will be combined using the DerSimonian and Laird random effect models to allow for appropriate inter-MA heterogeneity assessment [40]. P values will be two sided. Analysis will be performed by a statistician specialised in the meta-epidemiological approach.

4.3. Discussion

The findings of the proposed research could most likely have important implications for oral health research, dental practice decision-making and oral health policy. The proposed research will be the first meta-epidemiological study that provides empirical evidence regarding biases related to the quality of RCTs in the field of dentistry. This work, in combination with some of the current knowledge the oral health community already has, should have an important impact on the quality of future oral health RCTs, SRs and MAs by providing new and important insights about potential biases that exist in RCTs as well as factors associated with bias in oral health RCTs. Additionally, it will provide an improved framework when conducting, appraising and reporting oral health RCTs.

More importantly, this additional information will update dental professionals about proper, evidence-based decision-making when treating patients and could assist guideline developers and policymakers in making informed decisions about the implementation of dental interventions. Finally, the outcomes generated from this work should most likely be of value for developing and disseminating future research framework for the conduct and reporting of oral health RCTs.

Dissemination of the developed framework will be achieved through an array of means to maximise exposure. The intended audiences will include dental clinicians, oral health researchers, policymakers and graduate students. The aforementioned persons will be introduced to this research framework through workshops, seminars, round table discussions and targeted individual meetings. Moreover, key organisations will be used to strengthen the dissemination strategy, such as the International Association for Dental Research (IADR), the American Dental Association (ADA), the Canadian Dental Association (CDA) and the Cochrane Bias Methods Group. Other opportunities for knowledge transfer will be pursued such as key conferences (eg, the annual meeting of the International Association for Dental Research). Finally, the results of this study will be published as a scientific report in a dental peer-reviewed journal.

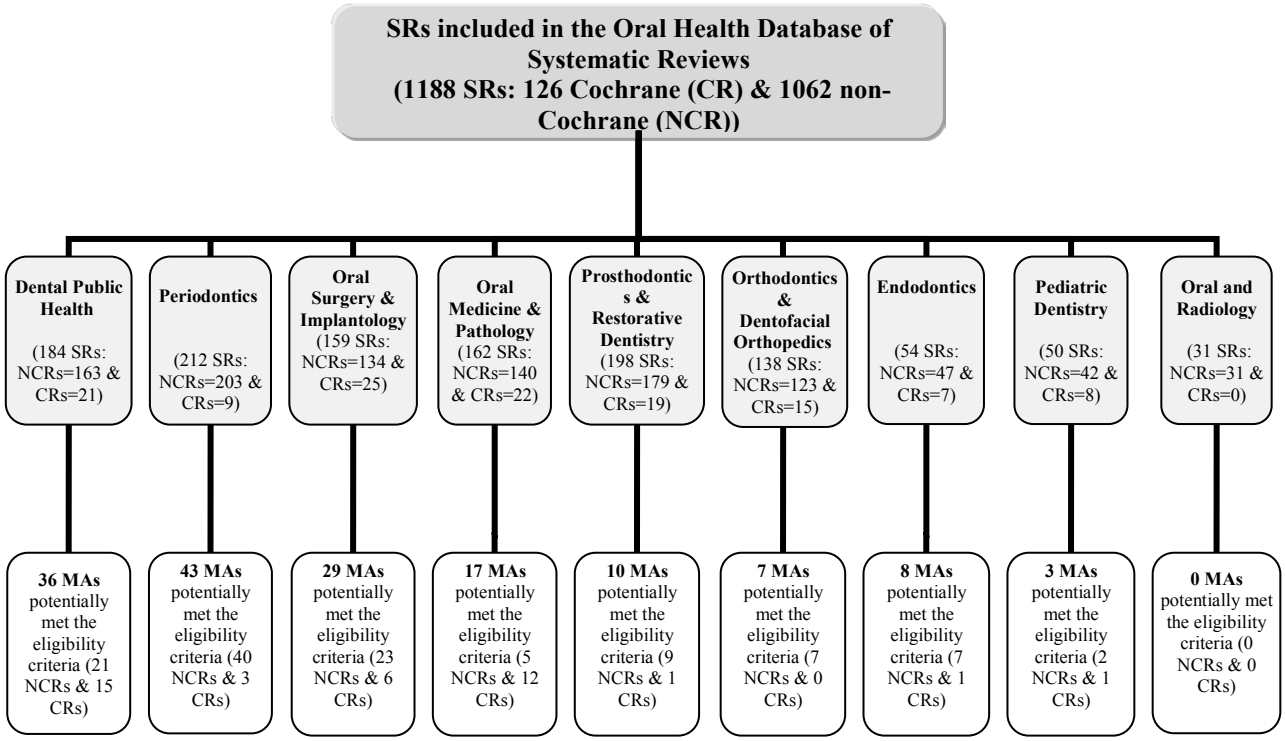

Figure 4.1. Systematic reviews identified in the database and the number of MAs potentially meeting the eligibility criteria for this study.

References

[1] Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001;134:663–94.

[2] Gluud LL. Bias in clinical intervention research. Am J Epidemiol 2006;163:493–501.

[3] Needleman I, Worthington H, Moher D, et al. Improving the completeness and transparency of reports of randomized trials in oral health: the CONSORT Statement. Am J Dentist 2008;21:7–12.

[4] Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10:28–55.

[5] Pandis N, Polychronopoulou A, Madianos P, et al. Reporting of research quality characteristics of studies published in 6 major clinical dental specialty journals. J Evid Based Dent Pract 2011;11:75–83.

[6] Seehra J, Wright NS, Polychronopoulou A, et al. Reporting quality of abstracts of randomized controlled trials published in dental specialty journals. J Evid Based Dent Pract 2013;13:1–8.

[7] Pandis N, Polychronopoulou A, Eliades T. An assessment of quality characteristics of randomised control trials published in dental journals J Dent. 2010;38:713–21.

[8] Dumbrigue HB, Al-Bayat MI, Ng CC, et al. Assessment of bias in methodology for randomized controlled trials published on implant dentistry J Prosthodont. 2006;15:257–63.

[9] Koletsi D, Pandis N, Polychronopoulou A, et al. Does published orthodontic research account for clustering effects during statistical data analysis? Eur J Orthod 2012;34:287–92.

[10] Polychronopoulou A, Pandis N, Eliades T. Appropriateness of reporting statistical results in orthodontics: the dominance of P values over confidence intervals. Eur J Orthod 2011;33:22–5.

[11] Koletsi D, Pandis N, Fleming PS. Sample size in orthodontic randomized controlled trials: are numbers justified? Eur J Orthod 2013;1:1.

[12] Koletsi D, Pandis N, Polychronopoulou A, et al. What's in a title? An assessment of whether randomized controlled trial in a title means that it is one. Am J Orthod Dentofacial Orthop 2012;141:679–85.

[13] Koletsi D, Pandis N, Polychronopoulou A, et al. Mislabeling controlled clinical trials (CCTs) as 'randomized clinical trials (RCTs)' in dental specialty journals. J Evid Based Dent Pract 2012;12:124–30.

[14] Fenwick J, Needleman IG, Moles DR. The effect of bias on the magnitude of clinical outcomes in periodontology: a pilot study. J Clin Periodontol 2008;35:775–82.

[15] Moher D, Cook DJ, Jadad AR, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. Health Technol Assess 1999;3:i–iv, 1–98.

[16] Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609–13.

[17] Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 2002;287:2973–82.

[18] van Tulder MW, Suttorp M, Morton S, et al. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine 2009;34:1685–92.

[19] Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2001;135:982–9.

[20] Egger M, Juni P, Bartlett C, et al. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 2003;7:1–76.

[21] Savovic J, Jones HE, Altman DG, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med 2012;157:429–38.

[22] Lexchin J, Bero LA, Djulbegovic B, et al. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 2003;326:1167–70.

[23] Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 2002;287:2973–82.

[24] Siersma V, Als-Nielsen B, Chen W, et al. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. Stat Med 2007;26:2745–58.

[25] Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008;336:601–5.

[26] Pildal J, Hrobjartsson A, Jorgensen KJ, et al. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int J Epidemiol 2007;36:847–57.

[27] Hempel S, Miles J, Suttorp MJ, et al. Detection of associations between trial quality and effect sizes. Rockville, MD: Agency for Healthcare Research and Quality (US), 2012.

[28] Saltaji H, Cummings GG, Armijo-Olivo S, et al. A Descriptive Analysis of Oral Health Systematic Reviews Published 1991–2012: cross Sectional Study. PLoS One 2013;8:e74545.

[29] Fleming PS, Koletsi D, Polychronopoulou A, et al. Are clustering effects accounted for in statistical analysis in leading dental specialty journals? J Dent 2013;41:265–70.

[30] American Dental Association. Specialty Definitions. http://www.ada. org/495.aspx (accessed April 2012).

[31] Major MP, Major PW, Flores-Mir C. Benchmarking of reported search and selection methods of systematic reviews by dental speciality. Evid Based Dent 2007;8:66–70.

[32] Glenny AM, Esposito M, Coulthard P, et al. The assessment of systematic reviews in dentistry. Eur J Oral Sci 2003;111: 85–92.

[33] Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0. The Cochrane Collaboration, 2008. http://www.cochrane-handbook.org

[34] Nuesch E, Trelle S, Reichenbach S, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. BMJ 2009;339:b3244.

[35] Nuesch E, Reichenbach S, Trelle S, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. Arthritis Rheum 2009;61:1633–41.

[36] Armijo-Olivo S, Fuentes J, Ospina M, et al. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: a descriptive analysis. BMC Med Res Methodol 2013;13:116.

[37] Armijo-Olivo S, Fuentes J, Rogers T, et al. How should we evaluate the risk of bias of physical therapy trials? a psychometric and meta-epidemiological approach towards developing guidelines for the design, conduct, and reporting of RCTs in Physical Therapy (PT) area: a study protocol. Syst Rev 2013;2:88.

[38] Sterne JA, Juni P, Schulz KF, et al. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med 2002;21:1513–24.

[39] Cohen J. Chapter 1: the concepts of power analysis. In: Cohen J, ed. Statistical power analysis for the behavioral sciences. 2nd edn. Hillsdale, NJ: Academic Press, INC, 1988:1–17.

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

Chapter 5

Randomized clinical trials in dentistry: Risk of bias, reporting quality, and methodologic quality over time, 1955–2013

5.1. Background

Randomized controlled trials (RCTs) are often referred to as the ideal type of clinical research to examine the effectiveness of treatment interventions in health sciences [1]. The value and significance of a randomized trial depend on the control of potential biases, how rigorously the trial was conducted, and how thoroughly the results were reported. The Consolidated Standards of Reporting Trials (CONSORT) Statement [2, 3], and recent initiatives such as the International Committee of Medical Journal Editors (ICMJE) statement on clinical trial registration [4, 5], have led to improvements in the "quality of evidence," including both the methodological and reporting quality of medical RCTs [6-8]. Adhering to these initiatives is critical to oral health research and practice, as high quality RCTs contribute largely to the body of evidence measured in systematic reviews and meta-analyses, especially when assesing therapeutic interventions.

Currently, it is estimated that nearly 50 clinical trials of oral health interventions are published every month, and it is expected that this number will increase over time [9]. There is emerging evidence from methodological reports published in the various fields of oral health over the last decade (periodontics [10], prosthodontics [11], implantology [12], orthodontics [13], and dentistry [14]) that the reported methodological quality of oral heath randomized trials is below an acceptable level to adequately lead clinical decision making. Moreover, there is evidence that some trials are biased and, due to weaknesses in their methodological characteristics, they tend to exaggerate the magnitude of the treatment effect [15]. This emerging evidence raises questions about the validity of the results of RCTs of oral health

interventions, which dental practitioners use when making day-to-day clinical decisions in dental practice, and which policy makers use more generally when developing clinical practice guidelines.

In the context of medical research methodology, "methodological quality" concerns the external and internal validity of a trial (the latter is determined by the extent to which the conduct and design of a trial are precisely and rigorously performed to generally acceptable standards so that biases are minimized) and "reporting quality" refers to the reporting of the conduct and design of a trial [16, 17]. While it is generally difficult to differentiate between reporting quality and methodological quality, "risk of bias" (which concerns the internal validity of a trial) [18] is often distinguishable from methodological quality. Interestingly, a trial may have a considerable risk of bias (e.g., due to the impossibility of applying blinding) yet still be conducted with and attain the highest acceptable principles [18].

In the field of oral health, to our knowledge, no study has assessed changes over time in the quality of reporting, in methodological characteristics, and in the risk of bias in RCTs of oral health interventions. A recent report by Reveiz et al. [19] described the results of an examination of different risk of bias domains in a sample of medical RCTs (identified from a cohort of Cochrane reviews). Reveiz's report stated that the rate of trials found with a low risk of bias consistently increased with time. However, since the RCTs in Reveiz's sample were performed in the field of medicine and were dependent on the risk of bias assessment performed by investigators presenting published reviews (rather than by conducting standardized data extraction from each trial), the findings cannot be compared to findings from trials in the field of dentistry which tend to have different design characteristics, such as difficulty in applying blinding and common use of the split-mouth design. Furthermore, a recent overview by Dechartres et al. [20], which evaluated a cohort of methodological reports assessing the quality of trials, indicated that these reports inadequately described the quality of the items used and the studies employed different approaches in evaluating

methodologic quality and risk of bias.

Consequently, it is unclear if the increase in number of published trials of oral health interventions over time is associated with changes in the conduct and reporting of the trials. To improve the conduct and reporting of RCTs, we set out to assess whether the reporting quality, the methodological quality, and the risk of bias in RCTs of oral health interventions, have improved over time. Our objectives were (1) to examine the reporting quality, the methodological characteristics, the risk of bias, and the general trial characteristics of RCTs of oral health interventions; and (2) to determine whether (and to what extent) the methodological quality, reporting quality, and risk of bias have improved over time.

5.2. Methods

5.2.1. Study sample

We used the Oral Health Database of Systematic Reviews [21] which includes all meta-analyses published in the field of oral health research between 1991 and 2014. From this database, we selected a sample of meta-analyses and their associated RCTs that met the following criteria: the meta-analysis was (1) published in any language and (2) conducted in a dental field that examined an intervention concerning craniofacial, oral, or dental diseases (as defined by the American Dental Association [ADA] scope of practice) [22]. An RCT was defined as "an experiment in which two or more interventions (possibly including a control intervention or no intervention) are compared, by being randomly allocated to participants" [18]. Further details regarding the study selection included in the final database of systematic reviews have been published [21]. Briefly, two reviewers (dentists with oral health research backgrounds) initially selected relevant reports and independently determined the final eligibility of the full texts (any disagreements were resolved through consensus). Ultimately, 540 RCTs that met the predefined eligibility criteria were selected and utilized in this study.

5.2.2. Data extraction

A panel of five reviewers from diverse health research areas (dentistry, pediatrics, and physical therapy) extracted the data. To ensure consistency during data extraction, two team members (H.S., S.A.) facilitated a reviewer training process, similar to the process followed in our team's previous investigations [23, 24]. In this process, the review panel evaluated and discussed 10 RCTs not included in the final set of trials.

Once agreement on data extraction protocols and interpretation was achieved, the review panel performed data extraction in duplicate. Two assessors independently carried out data extraction for each included RCT (consensus meetings were employed to resolve any disagreements). One assessor (H.S.) who has a background in oral health research performed a complete data extraction (n = 540, 100%), while other members of the review panel (C.H., J.S., J.F.) who have medical (nonoral health) research backgrounds, acted as secondary assessors. If two assessors could not reach an agreement, then a third assessor (S.A.) assisted with consensus. Only data that received consensus were used for data analyses. We used a structured and pilot-tested data extraction template, designed using Microsoft Office Access, for data extraction. We extracted details from each of the selected RCTs with respect to publication and trial characteristics, methodological quality, reporting quality, and risk of bias, as described below.

Publication and trial characteristics

Data elements related to publication and trial characteristics included the following information: publication year, dental speciality as classified by the American Dental Association (ADA) (e.g., periodontics, dental public health, prosthodontics and restorative dentistry, oral medicine and oral pathology, implantology, oral and maxillofacial surgery, orthodontics and dentofacial orthopedics, pediatric dentistry, endodontics [22]), country and continent of first

author, number of authors, funding source (e.g., foundation, government, industry, academic), type of journal (e.g., specialty oral health, general oral health, nonoral health), type of intervention (e.g., surgical, drug, dental material, device, psychological, educational, policy), age of participants, number of centres (e.g., multicentres, single centre), design (e.g., parallel, cross-over, split-mouth, cluster), type of outcome (e.g., subjective, objective), and sample size (see **Appendix 5.E**).

Methodological quality and reporting quality

Reporting quality and methodological quality are difficult to distinguish and often overlap to some extent. Methodological quality is defined as "the confidence that the trial design, conduct, and analysis has minimized or avoided biases in its treatment comparisons" [16, 17] (e.g., the sequence generation was appropriate). Reporting quality involves the provision of "information about the design, conduct, and analysis of the trial" [16, 17] (e.g., this was a randomized study). Accordingly, based on preliminary work performed by the research team [25, 26], we obtained 40 quality assessment criteria and their classifications ("reporting" vs. "conduct") from the most commonly used quality assessment tools in health care research [27-34]. Of the 40 quality criteria selected, 15 criteria assess "reporting" quality, 21 criteria assess "methodological" quality, and four quality criteria assess both reporting quality and methodological quality. We classified the items that evaluated methodological quality according to type of bias as follows [25, 26] (see **Appendix 5.A**): selection bias (6 criteria), performance and detection bias (4 criteria), performance bias (9 criteria), performance and compliance bias (2 criteria), information bias (3 criteria), reporting bias (3 criteria), attrition bias (5 criteria), detection bias (2 criteria), statistical bias (1 criterion), threats to precision (3 criteria), and multiple biases (2 criteria). We also grouped the selected quality criteria according to the following categories [18]: patient selection (inclusion and exclusion criteria, description of subjects); assignment, randomization, and allocation concealment; blinding; interventions; attrition, follow up and protocol deviation; outcomes; statistical analysis; and funding. Using original

tools as guidelines, the definitions and methods were derived for each criterion, using a three-part answering scheme (yes, no, unclear) for each item. We established decision rules and guidelines to ensure consistency (see **Appendices 5.A** and **5.B** for further details on the quality criteria used).

Risk of bias

We employed the Cochrane risk of bias tool [1] (introduced by the Cochrane Collaboration in 2008), which contains six domains and seven items, namely, "sequence generation," "allocation concealment," "blinding of outcome assessors," "blinding of participants," "incomplete outcome data," "selective outcome reporting," and "other sources of bias." We used the Cochrane Collaboration guidelines to score domains (e.g., high, low, unclear). However, we developed specific rules to make final decisions (see **Appendices 5.C** and **5.D)**. For "other sources of bias," we examined baseline comparability, control for cointerventions, and whether treatment compliance was acceptable [35]. For the overall assessment of risk of bias, if *one* domain was assessed as having a high risk, the overall risk of bias assessment was labelled "high risk." A randomized trial was considered to be at low risk of bias if it was assessed as "low risk" in *all* individual domains. If the assessment was "unclear" in *at least* one domain (and other domains were unclear or low) the overall risk of bias assessment was designated "unclear" [36, 37].

5.2.3. Data analysis

We conducted descriptive analyses for each trial characteristic, quality assessment item, and risk of bias domain (using means and standard deviations [SD] or median and interquartile range [IQR] for continuous outcomes, and proportions and percentages for categorical outcomes, where appropriate). To evaluate whether the quality of RCTs has improved over years, trials were grouped according to four time periods of publication year: before 1990, 1990–1999, 2000–2006, 2007–2013. We employed this classification after descriptively analysing the "publication year" of the

chosen trials (median: 2000; IQR: 1990, 2007). We implemented Chi-square statistics and two-tailed Fisher exact tests to examine the difference in proportion with respect to time periods for all quality assessment items and risk of bias domains. Furthermore, we used a logistic regression to explore the relationship between each criterion and time; we entered time into the logistic regression model as a continuous variable (publication year) and a categorical variable (time period of publication year; < 1990 was used as a reference category). The outcome of each analysis was each methodological criterion dichotomized in low risk vs. others (unclear, high risk of bias), or yes vs. others (no, unclear, not-reported).

We reported odds ratios (OR) with 95% confidence intervals (CIs) and we set the statistical significance at $P < .05$. We performed statistical analyses using Stata Version 14.0 (StataCorp) and the Statistical Package for Social Sciences for Windows (SPSS) Version 18.0 (IBM, Armonk, NY).

5.3. Results

5.3.1. RCT publication and trial characteristics

From the selected 540 trials published between 1955 and 2013 (median year of publication: 2000; IQR: 1990, 2007) (see **Figure 5.1**), the majority of trials were published either in periodontics (n = 233; 43.1%), dental public health (n = 124; 23.0%), or prosthodontic and restorative dentistry (n = 54; 10.0%). More than half of the trials were published in journals that specialized in oral health (n = 304; 56.3%).

The trials' first authors were most frequently from Europe (n = 239; 44.3%) followed by North America (n = 202; 37.4%). Three countries (the United Kingdom, Italy, and the United States) accounted for nearly half of all trials (n = 280; 51.9%). Approximately one fifth of the trials were multicenter trials, nearly half of the trials involved four to six authors (n = 249; 46.1%), and one third included two to three authors (n = 169; 31.3%). In approximately half of the trials, the authors did not

declare whether they received a source of funding (n = 256; 47.4%), while nearly one third of trials received funding from industry (n = 171; 31.7%).

Approximately one third of the trials were placebo-controlled (n = 204; 37.8%) and two thirds examined nondrug (n = 359; 66.5%) or nonsurgical (n = 370; 68.5%) interventions. One quarter of the trials examined pediatric patients (n = 136; 25.2%), while the majority examined adults (n = 398; 73.7%). The majority of trials used parallel design (n = 372; 68.9%), while almost one quarter used split-mouth design (n = 126; 23.3%). **Table 5.1** provides further details of publications and trial characteristics of the 540 trials.

5.3.2. Changes in risk of bias, reporting quality, and methodological characteristics over time

Sequence generation was judged to be adequate (at low risk of bias) in 32% (n = 173) of the trials, while the sequence generation in 67.6% (n = 365) of the trials was assessed as unclear. Allocation concealment was unclear in the majority of trials (n = 458; 84.8%). Blinding of participants was judged to be adequate (at low risk of bias) in 71.5% (n = 386) of the trials, and blinding of the outcome assessment was judged to be adequate (at low risk of bias) in 59.4% of the trials. Other sources of bias—baseline comparability, similarity of cointerventions, and compliance to treatment—were judged to have a low risk of bias in 77.8%, 40.2%, and 53.5% of the trials, respectively. The overall risk of bias was unclear in 73.7% (n = 398) of the trials, and a high risk of bias was assessed in 20.9% (n = 113) of the trials. Therefore, only 5.4% (n = 29) of the trials were judged to have a low risk of bias. **Table 5.2** provides further details of the proportions and percentages of trials judged to have low, high, or unclear risks of bias, for each domain of the Cochrane risk of bias tool.

We identified a significant increase ($P > 0.001$) in the proportion of trials judged as having a low risk of bias over time in five domains of the Cochrane risk of bias tool: sequence generation, allocation concealment, incomplete outcome data,

other sources of bias (including: baseline comparability, similarity of cointerventions, and compliance to treatment), and overall risk of bias. The proportion of trials assessed as having a low risk of bias, with respect to patient blinding, increased significantly ($P > 0.031$) in the sample, while change in blinding of outcome assessment and selective outcome reporting were not statistically significant. **Table 5.3** provides further details of the risk of bias assessment in different time periods.

Inclusion and exclusion criteria were clearly defined in the majority of trials, while baseline comparability was adequate in 72.8% of the trials. The method of randomization was assessed as unclear/not reported with respect to appropriateness and concealment in 65% and 86.9% of the trials, respectively. Two thirds of the trials were not described as double-blinded (n = 358; 66.3%), while the method of blinding was appropriate in 53% (n = 286) of the trials. Blinding of the assessor was reported in 59.4% (n = 321) of the trials, while blinding of subjects was unclear/not reported in slightly over half (n = 279; 51.7%) of the trials. Blinding of the principal investigator and statistician was unclear/not reported in 92% and 97.8% of the trials, respectively, that is, in the vast majority of the trials.

The treatment protocol was adequately described for treatment and control groups in the vast majority of trials, with 73.1% (n = 395) having a designated control group, and 38.1% (n = 206) using a placebo group. Whether cointerventions were avoided/comparable was assessed as unclear/not reported in 60.7% (n = 328) of the trials, while 84.1% (n = 424) of the trials did not report cointerventions for each group. Subject compliance to treatment protocol was tested in 61.1% (n = 330) of the trials, with compliance being acceptable in 50.9% (n = 275) of the trials. Withdrawals/dropouts were reported in the vast majority (89.4%, n = 483) of trials, with withdrawal/dropout rates being acceptable in 73.1% (n = 395) of the trials, and reasons for withdrawals/dropouts reported in 71.1% (n = 384) of the trials. Adverse effects were not described in nearly half (n = 259; 48%) of the trials.

Outcome measures were described in the majority of trials, while psychometric properties of main outcome measures—validity, reliability, and responsiveness—were not reported in 96.7%, 93.1%, and 96.7% of the trials, repectively. The statistical analysis was appropriate in 85.7% (n = 463) of the trials, with descriptive measures being reported in the majority of the trials. A sample size calculation before the initiation of the study was not performed in 77.8% (n = 420) of the trials, while the sample size was assessed as adequate in 17.6% (n = 95) of the trials. The clinical significance was not reported in 70.7% (n = 381) of the trials, while the intention to treat analysis was not used in 48.9% (n = 264) of the trials. The influence of the trial sponsor was assessed as being unclear in 72.8% (n = 393) of the trials, while it was assessed as appropriate in 16.7% (n = 90) of the trials. **Table 5.2** provides the proportions and percentages of trials assessed as yes, no, or unclear with respect to the methodologic quality assessment items.

The proportion of trials assessed as having adequately addressed methodological quality items increased significantly over time in 30 out of the 40 quality criteria (23 quality criteria at $P < 0.001$, seven quality criteria at $P < 0.05$). This was not statistically significant in the following items: study described as randomized, method of blinding appropriate, blinding of principle investigator, blinding of assessor, treatment protocol adequately described for the treatment group and for the comparison group, report of withdrawals/dropouts, outcome measures described, validity and responsiveness for main outcome measures reported, descriptive measures reported, and early cessation of a trial. **Table 5.4** provides further details of the quality assessment by domain over time.

The results of the logistic regression analyses showed that a significant change over time was evident in 29 out of the 36 quality criteria (that is, 10 risk of bias domains and 26 quality items) of which 26 quality criteria improved over time, while 3 criteria (study described as double blind, blinding of care-provider, and presence of placebo group) worsened over time. Conversely, 8 quality criteria (selective outcome

reporting, report of withdraws and dropouts, and 6 blinding-based criteria) did not show a significant change over time. **Table 5.5** provides further details of the results of the logistic regression analyses of the influence of time on each quality criterion.

5.4. Discussion

Bias is a threat to the quality of controlled clinical trials [38, 39]. The degree of bias in randomized trials of oral health interventions has decreased over time according to our study. We used the Cochrane Collaboration's risk of bias tool, in addition to a comprehensive set of reporting and methodological characteristics (selected from seven quality assessment tools reported to be valid), to assess the methodological quality of RCTs. Thus, this study provides an in-depth analysis of the methodological characteritics and risk of bias present in dental literature from 1955–2013.

In the majority of the quality items and risk of bias domains, our study showed that the proportion of trials having adequate quality or having a low risk of bias increased significantly over time. This encouraging trend is similar to what was identified in a recently published report by Reveiz et al. [19]. However, rather than conducting standardized data extraction from each trial, Reveiz used a risk of bias assessment reported by the investigators of reviews; this might be problematic, especially given the documented low reliability of the Cochrane Collaboration's risk of bias tool [36, 40]. The trend in our study is comparable to that found in a cohort of child-related trials [41] and medical RCTs [6]. A similar trend was identified when the methodological quality of trials of physical therapy interventions was assessed, where an improvement of nearly 0.6 points each decade was found in the total Physiotherapy Evidence Database (PEDro) score [42].

Although an improvement over time was identified in our study with respect to trials of oral health interventions, results of risk of bias and reported methodological quality assessments were unpropitious, indicating substandard quality and a high

potential for bias. We believe that sizable improvements in the conduct and reporting of oral health RCTs is possible. The fact that the proportion of trials having a low risk of bias did not exceed 60% in the majority of risk of bias domains is a concern. Because inadequate design and unrigorous conduct of a trial can bias the estimation of the treatment effect size, decisions made in dental practice might not be based on valid findings. For example, allocation concealment and sequence generation were unclear in 84.8% and 66.7% of the trials, respectively, although these factors improved significantly over time. It should be noted that an "unclear" risk of bias result in a trial may not mirror the actual design and conduct of the trial. Because journals have a word limit that may restrict authors in reporting detailed methods, all of the methodological characteristics used might not be reported [43], thus restricting the accuracy of quality assessment tools. The field of dentistry also lacks evidence to establish how a trial's design characteristics can affect overestimation or underestimation of the impact of a treatment within an oral health trial [15].

Our study revealed that more than half of the trials were published in specialist journals, and that nearly half of the trials were from the United Kingdom and the United States. These trends are similar for medical trials [43]. Possbly the interest of government and public sectors in the aforementioned countries is responsible for facilitating the financial support for such randomized trials [44].

The improvement observed in risk of bias and reported methodological quality of RCTs over time, could be attributed to efforts made by editors and reviewers of dental journals, through endorsement of the CONSORT Statement [45, 46], and by the mandatory implementation of trial registration, as recommended by the ICMJE [7, 47]. The CONSORT Statement is an accepted and widely used approach in medical and dental research to assess the reporting quality of RCTs. This approach covers the fundamental aspects of a trial's reporting quality; the CONSORT Statement aims to advance the transparency and quality of medical and dental trial reporting through the creation of reporting criteria [4, 46]. It has been endorsed during the last 10 years by

several medical journals worldwide, including the majority of high impact oral health journals [45, 46, 48]. Although the CONSORT Statement applies only to reporting quality, it is used commonly and erroneously by many dentistry researchers as a methodological quality assessment tool.

In the dental literature, the concepts "reporting quality" and "methodological quality" are often used interchangeably, contributing to conceptual ambiguity. Methodological quality depends mainly on the degree to which the design, conduct, and analysis of a trial follows the highest possible standards (to reduce multiple potential biases) and, hence, suggests that the findings can be based on the implemented intervention [1, 16, 18]. While the internal validity of a trial (which is closely connected to the risk of bias [18] and the methodological quality) should be the core of quality assessment, "reporting quality" is mistakenly used by researchers as an alternative for methodological quality; this has induced a conceptual ambiguity in the definition of trial "quality" [25]. In the context of medical research, a risk of bias assessment will benefit from an explicit and unambiguous definition of "methodological quality."

Although endorsement of the CONSORT Statement by dental journal editors and reviewers results in improvement in the reporting quality of trials, it does not guarantee compliance by trialists [46]. Furthermore, reliance on the CONSORT Statement only, may give reviewers, authors, and readers a false sense of security. Transparent reporting is desirable, but it does not necessarily raise methodological quality or lower the risk of bias [39]. For example, good reporting fails to prevent publication bias (i.e., trials of methods that have beneficial and large effects are published rapidly in journals with high impact), and selective outcome reporting (i.e., beneficial findings get publishing priority) [18]. These reported biases have exaggerated the magnitude of treatment effects in clinical trials, and can distort findings in meta-analysis [49, 50]. Implementation of the mandatory trial registration policy [4] could contribute to the improvement of trial quality and lower the risk of

bias identified in this study. Implementation of the mandatory trial registration policy started over 10 years ago by 11 leading medical journals, and is currently applied by over 300 medical journals [7], including many leading dental journals [51, 52]. Although recently only 23% of dental RCTs, published in 15 high impact dental journals, were registered [47, 53].

The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) initiative has become widely used to rate the strength of a body of evidence [54, 55]. Several agencies and societies have endorsed the use of the GRADE system, including the Agency for Healthcare Research and Quality (AHRQ), the World Health Organization (WHO), and the Cochrane Collaboration [56]. One of the key components of the GRADE initiative is classifying the "quality of evidence" in a systematic review. While several medical fields [57-60] have implemented the GRADE initiative, the adoption and use of the GRADE initiative by oral health institutions and journals (in the synthesis of evidence in dentistry) is lagging behind its use in other medical fields. The use of this approach will potentially improve the quality of evidence in oral health research [55]. Research into the synthesis of oral health evidence would benefit from a clear explanation of the following terms: methodological quality (i.e., conducting research at the highest possible standards), assessment of risk of bias (i.e., degree to which study findings are close to the truth), and methodologic quality of the evidence (i.e., confidence that the actual treatment effect size is close to the value estimated in the report) [18, 61]. Based on the findings in this study, endorsement of the GRADE initiative in the domain of oral health research would be an important step in informing clinical dental practice.

The results of this study have several implications. Dental trialists need to explicitly report their trials' results and adhere to published guidelines. Dental journal editors and reviewers should continue to be committed to international initiatives and statements developed to ensure adequate and appropriate conduct and reporting of randomized trials. Adherence to the above guidelines can reduce the risk that

inaccurate conclusions will be drawn from the research and, accordingly, will reduce inappropriate recommendations regarding treatment interventions in dental practice. Our findings call for oral health policy makers, methodologists, clinicians, and researchers to develop initiatives for improving clinical trials, which would spread such actions within the dental community. The formation of a global oral health initiative that aims to improve the conduct and reporting of oral health trials, and that prioritizes methodological criteria in oral health research, would be an example of a potentially needed measure to raise RCT standards.

5.4.1. Strengths and limitations

This cross-sectional study provided a comprehensive assessment of oral health RCTs with respect to trial characteristics, reporting quality, methodological characteristics, and risk of bias, and attempts to identify the variation of these factors over time. The range and size of our sample provided a comprehensive evaluation of oral health trials over the 58 year-period of 1955–2013. One of the strengths of our research was the data extraction method, which was performed in duplicate by two assessors to ensure high accuracy and avoid potential biases during the data extraction process. We performed a standardized data extraction rather than relying on the risk of bias assessment reported in systematic reviews, which was the case in a recent report by Reveiz et al. [19] where the risk of bias in medical RCTs was assessed.

A potential limitation of our research is that the choice of sample trials might not have been strictly random. Our sample consisted of 64 dental, oral, and craniofacial meta-analyses and was designed to cover the overall spectrum of clinical oral health research during 1955–2013, therefore, we submit that it represents a realistic cohort for that period.

Another potential limitation is that we did not contact the authors of the studied trials for missing data. A large proportion of the trials were published before the year 2000 when an author's correspondence information was sometimes not current and not

always provided in the publication. Moreover, because we extracted data based on the data reported in the published trials, the actual risk of bias potential was not visible in the majority of risk of bias domains studied due to the poor quality of the reporting identified in the studied trials. As our study did not look at factors that contributed to methodological quality improvement over time, these factors must be left to future research.

We applied an educated judgement to assign each RCT to a primary dental specialty (e.g., dental public health), although the RCT could be classified under more than one specialty (e.g., both pediatric dentistry and dental public health).

5.5. Conclusions

Our study showed a significant increase over time (1955–2013) in the proportion of trials judged to be adequate in reporting quality, methodological quality, and risk of bias. However, the proportion of trials judged as having a low risk of bias did not exceed 60% in the majority of the risk of bias domains. We found the risk of bias and the quality assessment in the studied trials to be unfavorable in general. That is, in the trials of oral health interventions the methodology and the reporting quality were substandard, resulting in a high potential for bias. We believe that a commitment to international initiatives by researchers, journal editors, and manuscript reviewers can contribute to the development of more stringent methodology and more detailed reporting of randomised trials of oral health interventions.

Table 5.1. Publication and trial characteristics of trials (N = 540)

Trial Characteristic	No. (%)
Primary dental specialty	
Periodontics	233 (43.1)
Dental public health	124 (23.0)
Prosthodontics and restorative dentistry	54 (10.0)
Oral medicine and oral pathology	42 (7.8)
Implantology	33 (6.1)
Oral and maxillofacial surgery	31 (5.7)
Orthodontics and dentofacial orthopedics	13 (2.4)
Pediatric dentistry	6 (1.1)
Endodontics	4 (0.7)
Date of publication	
Before 1990	127 (23.5)
1990-1999	135 (25.0)
2000-2006	138 (25.6)
2007-2013	140 (25.9)
Continent of first author	
Europe	239 (44.3)
North America	202 (37.4)
Asia	55 (10.2)
South America	28 (5.2)
Africa	7 (1.3)
Australia	9 (1.7)
Country of first author *(No. of countries = 45)*	
USA	187 (34.6)
UK	53 (9.8)
Italy	40 (7.4)
Sweden	27 (5.0)
Turkey	26 (4.8)
Brazil	25 (4.6)
Germany	20 (3.7)
Canada	16 (3.0)
France	13 (2.4)
China	12 (2.2)
Other	121 (22.4)
Number of authors	
1	25 (4.6)
3-2	169 (31.3)
6-4	249 (46.1)
7 ≤	97 (18.0)
Source of funding	
Industry	156 (28.9)
Government	43 (8.0)
Academics	19 (3.5)

Foundation	14 (2.6)
Government & Foundation/Academics	17 (3.1)
Industry & Government/Academics	15 (2.8)
Other combination	13 (2.4)
No funding	7 (1.3)
Funding not declared	256 (47.4)
Type of journal	
Specialty oral-health	304 (56.3)
General oral-health	171 (31.7)
Non-oral-health (medical)	65 (12.0)
Study design	
Parallel	372 (68.9)
Split-Mouth	126 (23.3)
Crossover	28 (5.2)
Cluster	10 (1.9)
Factorial	4 (0.7)
Placebo-controlled	
Yes	204 (37.8)
No	336 (62.2)
Number of centers	
Multicenter	103 (19.1)
2-5 center	51 (9.4)
6-10 center	28 (5.2)
>10 center	24 (4.4)
Single center	393 (72.8)
Unclear	44 (8.1)
Nature of intervention, classification I	
Drug	143 (26.5)
Non-drug	359 (66.5)
Both (drug and non-drug)	38 (7.0)
Nature of intervention, classification II	
Surgical	158 (29.3)
Non-surgical	370 (68.5)
Both (surgical and non-surgical)	12 (2.2)
Nature of intervention, classification III	
Drug	170 (31.5)
Surgical	163 (30.2)
Dental Material	83 (15.4)
Device	35 (6.5)
Psychological, Educational, Policy	16 (3.0)
Other	73 (13.5)
Mean age of participants	
Pediatric	136 (25.2)
Adult	398 (73.7)
Geriatric	6 (1.1)

Table 5.2. Risk of bias and quality assessments by criterion (N = 540)

Criterion	Risk of Bias Assessment, N (%)		
	Low Risk	High Risk	Unclear Risk
Sequence generation	173 (32)	2 (0.4)	365 (67.6)
Allocation concealment	76 (14.1)	6 (1.1)	458 (84.8)
Blinding of participants	386 (71.5)	7 (1.3)	147 (27.2)
Blinding of outcome assessment	321 (59.4)	16 (3.0)	203 (37.6)
Incomplete outcome data	295 (54.6)	93 (17.2)	152 (28.1)
Selective outcome reporting	519 (96.1)	5 (0.9)	16 (3.0)
Other sources of bias	286 (53.0)	1 (0.2)	253 (46.9)
Baseline comparability	420 (77.8)	0 (0.0)	120 (22.2)
Similarity of co-interventions	217 (40.2)	0 (0.0)	323 (59.8)
Compliance to the treatment	289 (53.5)	1 (0.2)	250 (46.3)
Overall risk of bias	29 (5.4)	113 (20.9)	398 (73.7)
	Quality Assessment, N (%)		
	Yes	No	Unclear/NR
Patient Selection (Inclusion and Exclusion and Description of Subjects)			
Inclusion criteria clearly defined	497 (92)	7 (1.3)	36 (6.7)
Exclusion criteria clearly defined	486 (90)	24 (4.4)	30 (5.6)
Baseline comparability (group equivalence)	393 (72.8)	3 (0.6)	144 (26.7)
Assignment, Randomization, and Allocation Concealment			
Study described as randomized	517 (95.7)	16 (3.0)	7 (1.3)
Method of randomization appropriate	181 (33.5)	8 (1.5)	351 (65)
Method of randomization concealed	64 (11.9)	7 (1.3)	469 (86.9)
Blinding			
Study described as double-blind	181 (33.5)	358 (66.3)	1 (0.2)
Method of blinding appropriate	286 (53)	17 (3.1)	237 (43.9)
Blinding of principal investigator	33 (6.1)	10 (1.9)	497 (92.0)
Blinding of assessor	321 (59.4)	16 (3.0)	203 (37.6)
Blinding of patients	192 (35.6)	69 (12.8)	279 (51.7)
Blinding of therapists/care-providers	134 (24.8)	356 (65.9)	50 (9.3)
Blinding of data analyst	9 (1.7)	3 (0.6)	528 (97.8)
Interventions			
Treatment protocol adequately described for treatment group	532 (98.5)	2 (0.4)	6 (1.1)
Treatment protocol adequately described for control group	528 (97.8)	4 (0.7)	8 (1.5)
Treatment protocol adequately described for comparison group †	227 (98.7)	1 (0.4)	2 (0.9)

Presence of a control group	395 (73.1)	143 (26.5)	2 (0.4)
Presence of a placebo group	206 (38.1)	334 (61.9)	0 (0.0)
Co-interventions avoided/comparable	208 (38.5)	4 (0.7)	328 (60.7)
Co-interventions reported for each group §	77 (15.3)	424 (84.1)	3 (0.6)
Testing of subject compliance to treatment protocol	330 (61.1)	11 (2)	199 (36.9)
Compliance acceptable (80% of treatment received)	275 (50.9)	5 (0.9)	260 (48.1)
Attrition, Follow-up and Protocol Deviation			
Report of withdraws and dropouts	483 (89.4)	20 (3.7)	37 (6.9)
Withdrawal/dropouts rate acceptable (< than 20%)	395 (73.1)	93 (17.2)	52 (9.6)
Reasons for withdraws/dropouts reported	384 (71.1)	109 (20.2)	47 (8.7)
Adverse effects described	276 (51.1)	259 (48.0)	5 (0.9)
Short follow-up measurement performed	509 (94.3)	31 (5.7)	0 (0.0)
Long term follow-up measurement performed	307 (68.7)	140 (31.3)	0 (0.0)
Outcomes			
Outcome measures described	528 (97.8)	6 (1.1)	6 (1.1)
Validity for main outcome measures reported	18 (3.3)	522 (96.7)	0 (0.0)
Reliability for main outcome measures reported	37 (6.9)	503 (93.1)	0 (0.0)
Responsiveness for main outcome measures reported	17 (3.1)	522 (96.7)	1 (0.2)
Data Analysis			
Descriptive measures identified and reported	534 (98.9)	4 (0.7)	2 (0.4)
Appropriate statistical analysis used	463 (85.7)	4 (0.7)	73 (13.5)
Sample size calculation done prior to study initiation	113 (20.9)	420 (77.8)	7 (1.3)
Adequate sample size	95 (17.6)	18 (3.3)	427 (79.1)
Intention to treat analysis used	218 (40.4)	264 (48.9)	58 (10.7)
Clinical significance reported	157 (29.1)	381 (70.7)	1 (0.2)
Funding			
Appropriate influence of trial sponsor	90 (16.7)	57 (10.6)	393 (72.8)
Early stopping of trial	5 (0.9)	535 (99.1)	0 (0.0)

† Does not equal 100 % for overall, as the item was not applicable in 310 trials.
§Does not equal 100 % for overall, as the item was not applicable in 36 trials.
‡ Does not equal 100 % for overall, as the item was not applicable in 93 trials.

Table 5.3. Risk of bias assessment by domain over time (N = 540), N (%)

Domain	Judgment	<1990	1990-1999	2000-2006	2007-2013	*P* value
Sequence generation	Low risk	15 (11.8)	27 (20)	59 (42.7)	72 (51.4)	<0.001
	High risk	1 (0.79)	0 (0.0)	0 (0.0)	1 (0.71)	-
	Unclear risk	111 (87.4)	108 (80)	79 (57.3)	67 (47.9)	<0.001
Allocation concealment	Low risk	6 (4.7)	16 (11.9)	26 (18.8)	39 (27.86)	<0.001
	High risk	1 (0.8)	1 (0.7)	1 (0.7)	3 (2.1)	-
	Unclear risk	120 (94.5)	118 (87.4)	111 (80.4)	98 (70)	<0.001
Blinding of participants	Low risk	83 (65.35	96 (71.1)	97 (70.3)	110 (78.6)	0.156
	High risk	0 (0.0)	0 (0.0)	4 (2.9)	3 (2.1)	0.072
	Unclear risk	44 (34.65)	39 (28.9)	37 (26.8)	27 (19.3)	0.057
Blinding of outcome assessment	Low risk	78 (61.4)	70 (51.9)	86 (62.3)	87 (62.1)	0.201
	High risk	0 (0.0)	3 (2.2)	6 (4.3)	7 (5)	0.069
	Unclear risk	49 (38.6)	62 (45.9)	46 (33.3)	46 (32.9)	0.109
Incomplete outcome data	Low risk	18 (14.2)	82 (60.7)	98 (71.0)	97 (69.3)	<0.001
	High risk	61 (48.0)	10 (7.4)	14 (10.1)	8 (5.7)	<0.001
	Unclear risk	48 (37.8)	43 (31.9)	26 (18.8)	35 (25)	0.077
Selective outcome reporting	Low risk	118 (92.9)	130 (96.3)	136 (98.6)	135 (96.4)	0.126
	High risk	3 (2.4)	1 (0.7)	0 (0.0)	1 (0.7)	0.234
	Unclear risk	6 (4.7)	4 (3)	2 (1.4)	4 (2.9)	0.480
Other sources of bias	Low risk	20 (15.7)	77 (57)	93 (67.4)	96 (68.6)	<0.001
	High risk	1 (0.8)	0 (0.0)	0 (0.0)	0 (0.0)	0.354
	Unclear risk	106 (83.5)	58 (43)	45 (32.6)	44 (31.4)	<0.001
Baseline comparability	Low risk	68 (53.5)	114 (84.4)	121 (87.7)	117 (83.6)	<0.001
	High risk	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	-
	Unclear risk	59 (46.5)	21 (15.6)	17 (12.3)	23 (16.4)	<0.001
Similarity of co-interventions	Low risk	17 (13.4)	60 (44.4)	70 (50.7)	70 (50)	<0.001
	High risk	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	-
	Unclear risk	110 (86.6)	75 (55.6)	68 (49.3)	70 (50)	<0.001
Compliance to the treatment	Low risk	23 (18.1)	75 (55.6)	91 (65.9)	100 (71.4)	<0.001
	High risk	1 (0.8)	0 (0.0)	0 (0.0)	0 (0.0)	0.354
	Unclear risk	103 (81.1)	60 (44.4)	47 (34.1)	40 (28.6)	<0.001
Overall risk of bias	Low risk	0 (0.0)	6 (4.4)	9 (6.5)	14 (10)	0.003
	High risk	69 (54.3)	14 (10.4)	20 (14.5)	10 (7.1)	<0.001
	Unclear risk	58 (45.7)	115 (85.2)	109 (79)	116 (82.9)	<0.001

Table 5.4. Quality assessment by item over time (N = 540), N (%)

Criterion	Judgment	<1990	1990-1999	2000-2006	2007-2013	*P*-value
Patient Selection (Inclusion and Exclusion and Description of Subjects)						
Inclusion criteria clearly defined	Yes	102 (80.3)	122 (90.4)	135 (97.8)	138 (98.6)	<0.001
	No	2 (1.6)	4 (3)	1 (0.7)	0 (0.0)	0.158
	Unclear/NR	23 (18.1)	9 (6.7)	2 (1.4)	2 (1.4)	<0.001
Exclusion criteria clearly defined	Yes	100 (78.7)	117 (86.7)	133 (96.4)	136 (97.1)	<0.001
	No	13 (10.2)	10 (7.4)	1 (0.7)	0 (0.0)	<0.001
	Unclear/NR	14 (11)	8 (5.9)	4 (2.9)	4 (2.9)	0.011
Baseline comparability (group equivalence)	Yes	70 (55.1)	104 (77)	112 (81.2)	107 (76.4)	<0.001
	No	3 (2.4)	0 (0.0)	0 (0.0)	0 (0.0)	0.020
	Unclear/NR	54 (42.5)	31 (23)	26 (18.8)	33 (23.6)	<0.001
Assignment, Randomization, and Allocation Concealment						
Study described as randomized	Yes	119 (93.7)	125 (92.6)	133 (96.4)	140 (100)	0.012
	No	6 (4.7)	7 (5.2)	3 (2.2)	0 (0.0)	0.041
	Unclear/NR	2 (1.6)	3 (2.2)	2 (1.4)	0 (0.0)	0.416
Method of randomization appropriate	Yes	18 (14.2)	32 (23.7)	58 (42)	73 (52.1)	<0.001
	No	4 (3.1)	2 (1.5)	2 (1.4)	0 (0.0)	0.210
	Unclear/NR	105 (82.7)	101 (74.8)	78 (56.5)	67 (47.9)	<0.001
Method of randomization concealed	Yes	3 (2.4)	12 (8.9)	19 (13.8)	30 (21.4)	<0.001
	No	1 (0.8)	1 (0.7)	3 (2.2)	2 (1.4)	0.698
	Unclear/NR	123 (96.9)	122 (90.4)	116 (84.1)	108 (77.1)	<0.001
Blinding						
Study described as double blind	Yes	61 (48)	48 (35.6)	30 (21.7)	42 (30.0)	<0.001
	No	66 (52)	87 (64.4)	107 (77.5)	98 (70.0)	<0.001
	Unclear/NR	0 (0.0)	0 (0.0)	1 (0.7)	0 (0.0)	0.318
Method of blinding appropriate	Yes	74 (58.3)	65 (48.1)	68 (49.3)	79 (56.4)	0.249
	No	0 (0.0)	3 (2.2)	8 (5.8)	6 (4.3)	0.041
	Unclear/NR	53 (41.7)	67 (49.6)	62 (44.9)	55 (39.3)	0.346
Blinding of principal investigator	Yes	9 (7.1)	3 (2.2)	8 (5.8)	13 (9.3)	0.100
	No	1 (0.8)	3 (2.2)	1 (0.7)	5 (3.6)	0.247
	Unclear/NR	117 (92.1)	129 (95.6)	129 (93.5)	122 (87.1)	0.064
Blinding of assessor	Yes	78 (61.4)	69 (51.1)	86 (62.3)	87 (62.1)	0.174
	No	0 (0.0)	3 (2.2)	6 (4.3)	7 (5)	0.069
	Unclear/NR	49 (38.6)	63 (46.7)	46 (33.3)	46 (32.9)	0.067
Blinding of subjects /patients	Yes	62 (48.8)	43 (31.9)	36 (26.1)	51 (36.4)	0.001
	No	10 (7.9)	20 (14.8)	24 (17.4)	15 (10.7)	0.093
	Unclear/NR	55 (43.3)	72 (53.3)	78 (56.5)	74 (52.9)	0.166
Blinding of therapists /care- providers	Yes	56 (44.1)	35 (25.9)	22 (15.9)	21 (15)	<0.001
	No	42 (33.1)	93 (68.9)	106 (76.8)	115 (82.1)	<0.001
	Unclear/NR	29 (22.8)	7 (5.2)	10 (7.2)	4 (2.9)	<0.001
Blinding of data analyst	Yes	0 (0.0)	0 (0.0)	2 (1.4)	7 (5)	0.003
	No	1 (0.8)	1 (0.7)	0 (0.0)	1 (0.7)	0.791
	Unclear/NR	126 (99.2)	134 (99.3)	136 (98.6)	132 (94.3)	0.013
Interventions						
Treatment protocol adequately described for treatment group	Yes	123 (96.9)	132 (97.8)	137 (99.3)	140 (100)	0.134
	No	1 (0.8)	0 (0.0)	1 (0.7)	0 (0.0)	0.554
	Unclear/NR	3 (2.4)	3 (2.2)	0 (0.0)	0 (0.0)	0.092

Treatment protocol adequately described for control group	Yes	120 (94.5)	132 (97.8)	136 (98.6)	140 (100)	0.020
	No	2 (1.6)	0 (0.0)	2 (1.4)	0 (0.0)	0.241
	Unclear/NR	5 (3.9)	3 (2.2)	0 (0.0)	0 (0.0)	0.019
Treatment protocol adequately described for comparison group	Yes	78 (97.5)	58 (98.3)	45 (100)	46 (100)	0.540
	No	0 (0.0)	1 (1.7)	0 (0.0)	0 (0.0)	0.406
	Unclear/NR	2 (2.5)	0 (0.0)	0 (0.0)	0 (0.0)	0.286
Presence of a control group	Yes	65 (51.2)	98 (72.6)	116 (84.1)	116 (82.9)	<0.001
	No	62 (48.8)	36 (26.7)	22 (15.9)	23 (16.4)	<0.001
	Unclear/NR	0 (0.0)	1 (0.7)	0 (0.0)	1 (0.7)	0.586
Presence of a placebo group	Yes	88 (69.3)	51 (37.8)	33 (23.9)	34 (24.3)	<0.001
	No	39 (30.7)	84 (62.2)	105 (76.1)	106 (75.7)	<0.001
	Unclear/NR	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	-
Co-interventions avoided/comparable	Yes	16 (12.6)	56 (41.5)	71 (51.4)	65 (46.4)	<0.001
	No	0 (0.0)	1 (0.7)	1 (0.7)	2 (1.4)	0.604
	Unclear/NR	111 (87.4)	78 (57.8)	66 (47.8)	73 (52.1)	<0.001
Co-interventions reported for each group separately	Yes	5 (4)	22 (18.2)	24 (18.6)	26 (20.2)	0.001
	No	120 (96)	99 (81.8)	102 (79.1)	103 (79.8)	<0.001
	Unclear/NR	0 (0.0)	0 (0.0)	3 (2.3)	0 (0.0)	0.032
Testing of subject compliance to treatment protocol	Yes	43 (33.9)	90 (66.7)	94 (68.1)	103 (73.6)	<0.001
	No	0 (0.0)	2 (1.5)	3 (2.2)	6 (4.3)	0.093
	Unclear/NR	84 (66.1)	43 (31.9)	41 (29.7)	31 (22.1)	<0.001
Compliance acceptable (80% of treatment received)	Yes	24 (18.9)	74 (54.8)	83 (60.1)	94 (67.1)	<0.001
	No	2 (1.6)	0 (0.0)	2 (1.4)	1 (0.7)	0.508
	Unclear/NR	101 (79.5)	61 (45.2)	53 (38.4)	45 (32.1)	<0.001
Attrition, Follow-up and Protocol Deviation						
Report of withdraws and dropouts	Yes	114 (89.8)	122 (90.4)	127 (92)	120 (85.7)	0.365
	No	6 (4.7)	6 (4.4)	2 (1.4)	6 (4.3)	0.444
	Unclear/NR	7 (5.5)	7 (5.2)	9 (6.5)	14 (10)	0.370
Withdrawal/dropouts rate acceptable (less than 20%)	Yes	53 (41.7)	114 (84.4)	116 (84.1)	112 (80)	<0.001
	No	61 (48)	9 (6.7)	13 (9.4)	10 (7.1)	<0.001
	Unclear/NR	13 (10.2)	12 (8.9)	9 (6.5)	18 (12.9)	0.341
Reasons for withdraws/dropouts reported	Yes	53 (41.7)	102 (75.6)	115 (83.3)	114 (81.4)	<0.001
	No	65 (51.2)	24 (17.8)	11 (8)	9 (6.4)	<0.001
	Unclear/NR	9 (7.1)	9 (6.7)	12 (8.7)	17 (12.1)	0.361
Adverse effects described	Yes	21 (16.5)	79 (58.5)	85 (61.6)	91 (65)	<0.001
	No	106 (83.5)	56 (41.5)	49 (35.5)	48 (34.3)	<0.001
	Unclear/NR	0 (0.0)	0 (0.0)	4 (2.9)	1 (0.7)	0.039
Short follow-up measurement performed	Yes	126 (99.2)	130 (96.3)	123 (89.1)	130 (92.9)	0.003
	No	1 (0.8)	5 (3.7)	15 (10.9)	10 (7.1)	0.003
	Unclear/NR	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	-
Long term follow-up measurement performed	Yes	110 (93.2)	62 (54.9)	63 (58.3)	72 (66.7)	<0.001
	No	8 (6.8)	51 (45.1)	45 (41.7)	36 (33.3)	<0.001
	Unclear/NR	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	-
Outcomes						
Outcome measures described	Yes	121 (95.3)	132 (97.8)	138 (100)	137 (97.9)	0.079
	No	2 (1.6)	2 (1.5)	0 (0.0)	2 (1.4)	0.553
	Unclear/NR	4 (3.1)	1 (0.7)	0 (0.0)	1 (0.7)	0.081
Validity for main outcome measures reported	Yes	0 (0.0)	6 (4.4)	8 (5.8)	4 (2.9)	0.055
	No	127 (100)	129 (95.6)	130 (94.2)	136 (97.1)	0.055

	Unclear/NR	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	-
Reliability for main outcome measures reported	Yes	3 (2.4)	13 (9.6)	15 (10.9)	6 (4.3)	0.014
	No	124 (97.6)	122 (90.4)	123 (89.1)	134 (95.7)	0.014
	Unclear/NR	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	-
Responsiveness for main outcome measures reported	Yes	0 (0.0)	8 (5.9)	4 (2.9)	5 (3.6)	0.054
	No	127 (100)	127 (94.1)	133 (96.4)	135 (96.4)	0.064
	Unclear/NR	0 (0.0)	0 (0.0)	1 (0.7)	0 (0.0)	0.404
Data Analysis						
Descriptive measures identified and reported	Yes	125 (98.4)	132 (97.8)	137 (99.3)	140 (100)	0.317
	No	1 (0.8)	2 (1.5)	1 (0.7)	0 (0.0)	0.561
	Unclear/NR	1 (0.8)	1 (0.7)	0 (0.0)	0 (0.0)	0.545
Appropriate statistical analysis used	Yes	72 (56.7)	122 (90.4)	134 (97.1)	135 (96.4)	<0.001
	No	3 (2.4)	1 (0.7)	0 (0.0)	0 (0.0)	0.085
	Unclear/NR	52 (40.9)	12 (8.9)	4 (2.9)	5 (3.6)	<0.001
Sample size calculation performed prior to initiation of the study	Yes	7 (5.5)	13 (9.6)	31 (22.5)	62 (44.3)	<0.001
	No	119 (93.7)	119 (88.1)	105 (76.1)	77 (55)	<0.001
	Unclear/NR	1 (0.8)	3 (2.2)	2 (1.4)	1 (0.7)	0.669
Adequate sample size	Yes	3 (2.4)	10 (7.4)	28 (20.3)	54 (38.6)	<0.001
	No	3 (2.4)	4 (3)	4 (2.9)	7 (5)	0.633
	Unclear/NR	121 (95.3)	121 (89.6)	106 (76.8)	79 (56.4)	<0.001
Intention to treat analysis used	Yes	12 (9.4)	48 (35.6)	76 (55.1)	82 (58.6)	<0.001
	No	104 (81.9)	69 (51.1)	50 (36.2)	41 (29.3)	<0.001
	Unclear/NR	11 (8.7)	18 (13.3)	12 (8.7)	17 (12.1)	0.492
Clinical significance reported	Yes	21 (16.5)	46 (34.1)	56 (40.6)	34 (24.5)	<0.001
	No	106 (83.5)	89 (65.9)	82 (59.4)	104 (74.8)	<0.001
	Unclear/NR	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.7)	0.410
Funding						
Appropriate influence of trial sponsor	Yes	14 (11)	17 (12.6)	20 (14.5)	39 (27.9)	0.001
	No	4 (3.1)	21 (15.6)	20 (14.5)	12 (8.6)	0.003
	Unclear/NR	109 (85.8)	97 (71.9)	98 (71)	89 (63.6)	0.001
Early stopping of trial	Yes	2 (1.6)	0 (0.0)	2 (1.4)	1 (0.7)	0.508
	No	125 (98.4)	135 (100)	136 (98.6)	139 (99.3)	0.508
	Unclear/NR	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	-

Table 5.5. Results from the logistic regression analysis for low risk of bias or adequate quality criteria[†]

Criterion[§#]	Publication year[*]		Time-periods of publication year[¶]					
			<1990	1990-1999		2000-2006		2007-2013	
	OR[‡] (95% CI)	*P*-value		OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
Risk of Bias Assessment									
Sequence generation	1.080 (1.057-1.103)	<0.001	1.00	1.866 (0.941-3.700)	0.074	5.413 (2.865-10.226)	<0.001	7.905 (4.199-14.883)	<0.001
Allocation concealment	1.080 (1.049-1.111)	<0.001	1.00	2.711 (1.026-7.165)	0.044	4.681 (1.857-11.796)	0.001	7.787 (3.168-19.137)	<0.001
Blinding of participants	1.015 (1.001-1.030)	0.038	1.00	1.304 (0.774-2.198)	0.317	1.254 (0 .748- 2.102)	0.390	1.943 (1.127-3.350)	0.017
Blinding of outcome assessment	0.997 (0.984-1.011)	0.768	1.00	0.676 (0.413-1.106)	0.119	1.038 (0.632-1.706)	0.880	1.031 (0.629-1.690)	0.903
Incomplete outcome data	1.080 (1.062-1.099)	<0.001	1.00	9.368 (5.107-17.18)	<0.001	14.836 (7.984-27.567)	<0.001	13.660 (7.389-25.253)	<0.001
Selective outcome reporting	1.026 (0.995-1.058)	0.100	1.00	1.983 (0.646-6.085)	0.231	5.186 (1.098-24.481)	0.038	2.059 (0.671-6.316)	0.206
Other sources of bias	1.091 (1.071-1.111)	<0.001	1.00	7.102 (3.95-12.769)	<0.001	11.056 (6.095-20.056)	<0.001	11.672 (6.431-21.185)	<0.001
Baseline comparability	1.054 (1.037-1.071)	<0.001	1.00	4.815 (2.647-8.759)	<0.001	6.467 (3.409-12.269)	<0.001	4.012 (2.273-7.080)	<0.001
Similarity of co-interventions	1.065 (1.047-1.084)	<0.001	1.00	5.966 (3.267-10.89)	<0.001	7.208 (3.953-13.144)	<0.001	6.411 (3.523-11.668)	<0.001
Compliance to treatment	1.093 (1.073-1.114)	<0.001	1.00	6.054 (3.455-10.608)	<0.001	9.481 (5.351-16.798)	<0.001	12.398 (6.914-22.231)	<0.001
Patient Selection (Inclusion and Exclusion and Description of Subjects)									
Inclusion criteria clearly defined	1.066 (1.042-1.091)	<0.001	1.00	2.300 (1.119-4.725)	0.023	11.029 (3.240-37.540)	<0.001	16.911 (3.916-73.028)	<0.001
Exclusion criteria clearly defined	1.057 (1.036-1.079)	<0.001	1.00	1.755 (0.913-3.373)	0.092	7.182 (2.671-19.306)	<0.001	9.179 (3.113-27.068)	<0.001

Assignment, Randomization, and Allocation Concealment									
Method of randomization appropriate	1.073 (1.052-1.094)	<0.001	1.00	1.881 (0.994-3.557)	0.052	4.390 (2.403-8.018)	<0.001	6.597844 (3.625-12.008)	<0.001
Method of randomization concealed	1.105 (1.063-1.149)	<0.001	1.00	4.032 (1.11-14.642)	0.034	6.599 (1.903-22.881)	0.003	11.272 (3.347-37.964)	<0.001
Blinding									
Study described as double-blind	0.972 (0.958-0.986)	<0.001	1.00	0.559 (0.339-0.920)	0.022	0.300 (0.176-0.512)	<0.001	0.448 (0.270-0.741)	0.002
Method of blinding appropriate	0.991 (0.978-0.004)	0.210	1.00	0.665 (0.408-1.083)	0.102	0.695 (0.428-1.130)	0.143	0.927 (0.570-1.507)	0.762
Blinding of principal investigator	1.004 (0.975-1.033)	0.774	1.00	0.297 (0.078-1.126)	0.074	0.806 (0.301-2.159)	0.669	1.342 (0.553-3.255)	0.515
Blinding of assessor	0.997 (0.984-1.011)	0.768	1.00	0.676 (0.413-1.106)	0.119	1.038 (0.632-1.706)	0.880	1.031 (0.629-1.690)	0.903
Blinding of patients	0.976 (0.962-0.989)	0.001	1.00	0.473 (0.286-0.783)	0.004	0.329 (0.195-0.556)	<0.001	0.564 (0.345-0.922)	0.023
Blinding of care-providers	0.952 (0.938-0.967)	<0.001	1.00	0.443 (0.263-0.746)	0.002	0.240 (0.135-0.427)	<0.001	0.223 (0.125-0.400)	<0.001
Interventions									
Presence of a control group	1.055 (1.039-1.071)	<0.001	1.00	2.526 (1.511-4.223)	<0.001	5.029 (2.834-8.923)	<0.001	4.610 (2.631-8.075)	<0.001
Presence of a placebo group	0.936 (0.922-0.951)	<0.001	1.00	0.271 (0.162-0.452)	<0.001	0.162 (0.094-0.276)	<0.001	0.152 (0.089-0.261)	<0.001
Co-interventions avoided /comparable	1.065 (1.047-1.084)	<0.001	1.00	5.966 (3.267-10.89)	<0.001	7.208 (3.953-13.144)	<0.001	6.411 (3.523-11.668)	<0.001
Co-interventions reported for each group	1.054 (1.027-1.082)	<0.001	1.00	5.333 (1.948-14.59)	0.001	5.485 (2.021-14.888)	0.001	6.058 (2.245-16.347)	<0.001
Testing of subject compliance to treatment protocol	1.061 (1.045-1.078)	<0.001	1.00	3.906 (2.339-6.525)	<0.001	4.173 (2.498-6.971)	<0.001	5.438 (3.215-9.197)	<0.001
Compliance acceptable (80% of treatment received)	1.079 (1.060-1.098)	<0.001	1.00	5.206 (2.977-9.103)	<0.001	6.476 (3.699-11.337)	<0.001	8.769 (4.973-15.464)	<0.001
Attrition, Follow-up and Protocol Deviation									
Report of withdraws and dropouts	0.990 (0.968-1.013)	0.412	1.00	1.070 (0.476-2.405)	0.870	1.316 (0.5673-3.055)	0.522	.684 (0.325-1.439)	0.317

Withdrawal/dropouts rate acceptable (< than 20%)	1.063 (1.046-1.080)	<0.001	1.00	7.579 (4.226-13.59)	<0.001	7.361 (4.137-13.100)	<0.001	5.584 (3.241-9.621)	<0.001
Reasons for withdraws/ dropouts reported	1.065 (1.048-1.082)	<0.001	1.00	4.315 (2.546-7.315)	<0.001	6.981 (3.948-12.343)	<0.001	6.121 (3.521-10.642)	<0.001
Adverse effects described	1.068 (1.050-1.086)	<0.001	1.00	7.120 (3.986-12.71)	<0.001	8.095 (4.530-14.463)	<0.001	9.374 (5.233-16.791)	<0.001
Data Analysis									
Appropriate statistical analysis used	1.122 (1.097-1.149)	<0.001	1.00	7.168 (3.664-14.022)	<0.001	25.590 (8.913-73.466)	<0.001	20.625 (7.903-53.819)	<0.001
Sample size calculation done prior to study initiation	1.110 (1.077-1.145)	<0.001	1.00	1.826 (0.704-4.736)	0.215	4.966 (2.100-11.743)	<0.001	13.626 (5.930-31.307)	<0.001
Adequate sample size	1.135 (1.093-1.178)	<0.001	1.00	3.306 (0.888-12.303)	0.074	10.521 (3.112-35.566)	<0.001	25.953 (7.858-85.711)	<0.001
Intention to treat analysis used	1.088 (1.067-1.111)	<0.001	1.00	5.287 (2.648-10.55)	<0.001	11.747 (5.935-23.249)	<0.001	13.548 (6.842-26.826)	<0.001
Clinical significance reported	1.022 (1.006-1.038)	0.007	1.00	2.608 (1.448-4.697)	0.001	3.447 (1.933-6.147)	<0.001	1.634 (0.890-2.999)	0.113
Funding									
Appropriate influence of trial sponsor	1.043 (1.019-1.067)	<0.001	1.00	1.162 (0.547-2.468)	0.695	1.368 (0.659- 2.839)	0.400	3.116 (1.599-6.072)	0.001

† **Low** risk vs. others (both **Unclear** and **High** risk of bias); or **Yes** vs. others (both **No** and **Unclear/Not-reported)**.

* Time was entered in the logistic regression model as a continuous variable.

¶ Time was entered in the logistic regression model as a categorical variable.

§ The following criteria were not considered in the analysis because of either having a small number of trials judged as being adequate: ***overall risk of bias; blinding of data analyst; validity, reliability, and responsiveness for main outcome measures reported; study described as randomized; and early stopping of trial.***

The following criteria were not considered in the analysis because of having a small number of trials judged as being inadequate/unclear: **treatment protocol adequately described for treatment, control, or comparison group; short or long follow-up measurement performed; outcome measures described; descriptive measures identified and reported.**

‡ The factor which the odds of the quality criteria, being adequate, increased by every year.

Abbreviations: CI, confidence interval; OR, odds ratio.

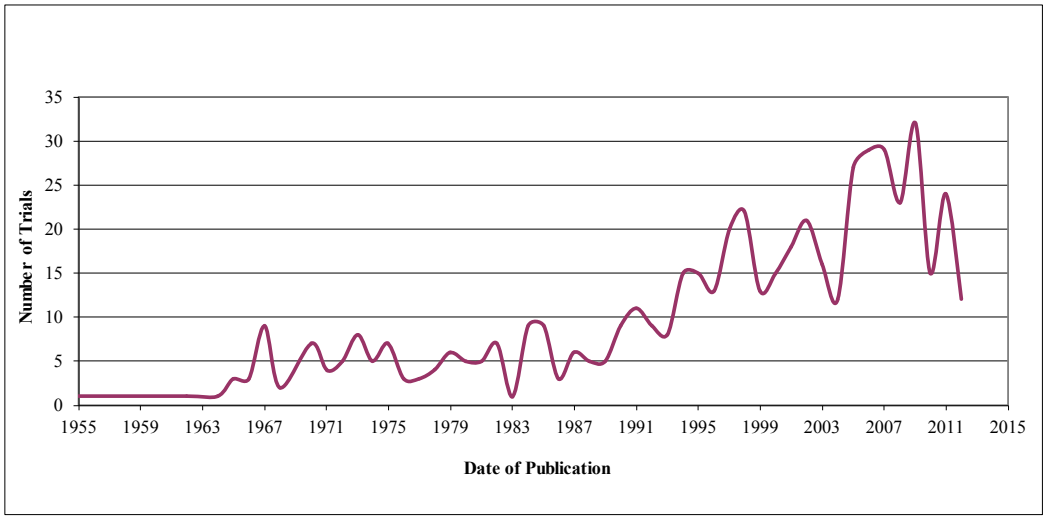

Figure 5.1. Number of oral heath trials according to year of publication.

References

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

[2] Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 2001; 285:1987-91.

[3] Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996; 276:637-9.

[4] De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. Lancet 2004; 364:911-2.

[5] De Angelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Is this clinical trial fully registered? —A statement from the International Committee of Medical Journal Editors. N Engl J Med 2005; 352:2436-8.

[6] Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ 2010; 340:c723.

[7] Hooft L, Korevaar DA, Molenaar N, Bossuyt PM, Scholten RJ. Endorsement of ICMJE's Clinical Trial Registration Policy: a survey among journal editors. Neth J Med 2014; 72:349-55.

[8] Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Syst Rev 2012; 1:60.

[9] Needleman I, Worthington H, Moher D, Schulz K, Altman DG. Improving the completeness and transparency of reports of randomized trials in oral health: the CONSORT statement. Am J Dent 2008; 21:7-12.

[10] Montenegro R, Needleman I, Moles D, Tonetti M. Quality of RCTs in periodontology—a systematic review. J Dent Res 2002; 81:866-70.

[11] Nieri M, Clauser C, Franceschi D, Pagliaro U, Saletta D, Pini-Prato G. Randomized clinical trials in implant therapy: relationships among methodological, statistical, clinical, paratextual features and number of citations. Clin Oral Implants Res 2007; 18:419-31.

[12] Vere J, Joshi R. Quality assessment of randomised controlled trials of dental implant surgery and prosthodontics published from 2004 to 2008: a systematic review. Clin Oral Implants Res 2011; 22:1338-45.

[13] Harrison JE. Clinical trials in orthodontics II: assessment of the quality of reporting of clinical trials published in three orthodontic journals between 1989 and 1998. J Orthod 2003; 30:309-15.

[14] Pandis N, Polychronopoulou A, Eliades T. An assessment of quality characteristics of randomised control trials published in dental journals. J Dent 2010; 38:713-21.

[15] Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, Flores-Mir C. Methodological characteristics and treatment effect sizes in oral health randomised controlled trials: Is there a relationship? Protocol for a meta-epidemiological study. BMJ Open 2014; 4:e004527.

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

[17] Verhagen AP, de Vet HC, de Bie RA, Boers M, van den Brandt PA. The art of quality assessment of RCTs included in systematic reviews. J Clin Epidemiol 2001; 54:651-4.

[18] Higgins JP, Green S, eds. Cochrane Handbook For Systematic Reviews of Interventions version 5.0.0. Cochrane Collaboration 2008.

[19] Reveiz L, Chapman E, Asial S, Munoz S, Bonfill X, Alonso-Coello P. Risk of bias of randomized trials over time. J Clin Epidemiol 2015; 68:1036-45.

[20] Dechartres A, Charles P, Hopewell S, Ravaud P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. J Clin Epidemiol 2011; 64:136-44.

[21] Saltaji H, Cummings GG, Armijo-Olivo S, Major MP, Amin M, Major PW, et al. A descriptive analysis of oral health systematic reviews published 1991-2012: cross sectional study. PloS One 2013; 8:e74545.

[22] American Dental Association (ADA). Specialty Definitions 2015. Available at: http://www.ada.org/495.aspx. Accessed July 2015.

[23] Armijo-Olivo S, Ospina M, da Costa BR, Egger M, Saltaji H, Fuentes CJ, et al. Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. PloS One 2014; 9(5): e96920.

[24] Armijo-Olivo S, Saltaji H, da Costa BR, Fuentes J, Ha C, Cummings GG. What is the influence of randomisation sequence generation and allocation concealment on

treatment effects of physical therapy trials? A meta-epidemiological study. BMJ Open 2015; 5:e008562.

[25] Armijo-Olivo S, Fuentes CJ, Ospina M, Saltaji H, Hartling L. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: A descriptive analysis. BMC Med Res Methodol 2013; 13:1-19.

[26] Armijo-Olivo S, Fuentes CJ, Rogers T, Hartling L, Saltaji H, Cummings G. How should we evaluate the risk of bias of physical therapy trials? A psychometric and meta-epidemiological approach towards developing guidelines for the design, conduct, and reporting of RCTs in physical therapy (PT) area: a study protocol. Syst Rev 2013; 2:88.

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

[28] Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 1998; 51:1235-41.

[29] van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. Spine 2003; 28:1290-9.

[30] van Tulder MW, Assendelft WJ, Koes BW, Bouter LM. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for Spinal Disorders. Spine 1997; 22:2323-30.

[31] Moseley AM, Herbert RD, Sherrington C, Maher CG. Evidence for physiotherapy practice: a survey of the Physiotherapy Evidence Database (PEDro). Aust J Physiother 2002; 48:43-9.

[32] Sherrington C, Herbert RD, Maher CG, Moseley AM. PEDro. A database of randomized trials and systematic reviews in physiotherapy. Man Ther 2000; 5:223-6.

[33] Bizzini M, Childs JD, Piva SR, Delitto A. Systematic review of the quality of randomized controlled trials for patellofemoral pain syndrome. J Orthop Sports Phys Ther 2003; 33:4-20.

[34] Armijo-Olivo S, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. Phys Ther 2008; 88:156-75.

[35] Furlan AD, Pennick V, Bombardier C, Van Tulder M. 2009 Updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine 2009; 34:1929-41.

[36] Hartling L, Hamm MP, Milne A, Vandermeer B, Santaguida PL, Ansari M, et al. Testing the Risk of Bias Tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. J Clin Epidemiol 2012; 66:973-81.

[37] Hartling L, Bond K, Vandermeer B, Seida J, Dryden D, Rowe B. Applying the Risk of Bias Tool in a systematic review of combination longacting betaagonists and inhaled corticosteroids for persistent asthma. PLoS Med 2011; 6.

[38] Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001; 323:42-6.

[39] Saltaji H. Bias is the key challenge in orthodontic research. Am J Orthod Dentofacial Orthop 2015; 148:8.

[40] Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials? Evaluation of risk of bias assessments in systematic reviews. BMJ 2013; 346:f1798.

[41] Thomson D, Hartling L, Cohen E, Vandermeer B, Tjosvold L, Klassen TP. Controlled trials in children: quantity, methodological quality and descriptive characteristics of pediatric controlled trials published 1948–2006. PloS One 2010; 5(9).

[42] Moseley AM, Herbert RD, Maher CG, Sherrington C, Elkins MR. Reported quality of randomized controlled trials of physiotherapy interventions has improved over time. J Clin Epidemiol 2011; 64:594-601.

[43] Hamm MP, Hartling L, Milne A, Tjosvold L, Vandermeer B, Thomson D, et al. A descriptive analysis of a representative sample of pediatric randomized controlled trials published in 2007. BMC Pediatr 2010; 10:96.

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

[45] Pandis N, Turpin DL. Enhancing CONSORT compliance for improved reporting of randomized controlled trials. Am J Orthod Dentofacial Orthop 2014; 145:1.

[46] Turpin DL. Challenges faced by journal editors. Am J Orthod Dentofacial Orthop 2015; 147:285-9.

[47] Smail-Faugeron V, Fron-Chabouis H, Durieux P. Clinical trial registration in oral health journals. J Dent Res 2015; 94:8S-13S.

[48] Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010; 63:e1-37.

[49] Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias – an updated review. PloS One 2013; 8:e66844.

[50] Bialy L, Vandermeer B, Lacaze-Masmonteil T, Dryden DM, Hartling L. A meta-epidemiological study to examine the association between bias and treatment effects in neonatal trials. Evid Based Child Health 2014; 9:1052-9.

[51] Giannobile WV. Improving clinical trials in dentistry. J Dent Res 2015; 94:6S-7S.

[52] Turpin DL. The need to register clinical trials in orthodontics. Am J Orthod Dentofacial Orthop 2006; 130:429-30.

[53] Saltaji H, Flores-Mir C, Major PW. Letter to the editor, "Clinical trial registration in oral health journals." J Dent Res 2015; 94:103S.

[54] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008; 336:924-6.

[55] Pandis N, Fleming PS, Worthington H, Salanti G. The quality of the evidence according to GRADE is predominantly low or very low in oral health systematic reviews. PloS One 2015; 10:e0131644.

[56] Organizations that have endorsed or that are using GRADE. http://www.gradeworkinggroup.org/society/index.htm. Accessed Feb 23, 2015.

[57] 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.

[58] Brito JP, Tsapas A, Griebeler ML, Wang Z, Prutsky GJ, Domecq JP, et al. Systematic reviews supporting practice guideline recommendations lack protection against bias. J Clin Epidemiol 2013; 66:633-8.

[59] Swiglo BA, Murad MH, Schunemann HJ, Kunz R, Vigersky RA, Guyatt GH, et al. 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; 93:666-73.

[60] Guyatt G, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B, et al. Grading strength of recommendations and quality of evidence in clinical guidelines: report from the American College of Chest Physicians task force. Chest 2006; 129:174-81.

[61] Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011; 64:395-400.

Chapter 6

The impact of selection bias on treatment effect size estimates in randomized trials of oral health interventions

6.1. Background

Randomized controlled trials (RCTs) are a key component in the knowledge base that clinicians consistently rely on for everyday treatment-based decisions. These trials are often grouped together to form systematic reviews and meta-analyses, which comprise the "gold standard" of scientific evidence for therapeutic interventions [1, 2]. Fortunately for the scientific community, the number of reported clinical trials is continually increasing, with nearly 50 new trials published per month in the field of dentistry alone [3, 4]. Given the importance of these trials as potential building blocks for policies and clinical decisions, it is imperative that the quality of trials is diligently monitored [5]. Unfortunately, due to shortcomings in their conduct, RCTs are potentially exposed to under- or over-estimation of the magnitude of treatment effects [6].

Selection bias takes place when individuals responsible for recruiting participants discriminate in enrolling participants into the trial, according to a likely forthcoming treatment assignment [7, 8]. An example of this type of bias occurs when trial recruiters expect the next assignment to be the intervention, or have access to the assignment list; in such instances they may try to assign severely affected cases to this treatment group [7]. This is a particular concern when participants are recruited consecutively, rather than being enrolled in the trial at the same time [9, 10]. For this reason, adequate allocation concealment (a process of concealing information about which patients are to be assigned to a new treatment versus those to be given a conventional therapy) and randomization (allocation is carried out using a chance mechanism so that neither the participant nor the investigator will know in advance

which will be assigned to an intervention) have been recognized as crucial for preventing selection bias in RCTs [11-13]. While appropriate randomization is implemented in a trial to minimize potential bias in the assignment of interventions, and to comparably allocate participants' differences to the treatment groups [1, 14, 15], allocation concealment is broadly used to ensure that individuals (participants and principal investigators) involved in a randomized trial remain uninformed of the specific nature of forthcoming assignments [16-18]. In the same way, baseline (information gathered at the beginning of a study from which variations found in the study are measured) comparability is believed to guarantee that the method of randomization is effective in ensuring that differences in baseline characteristics are not "real" [19, 20]. In fact, there is a debate whether testing for baseline comparability in RCTs is needed, given that the randomization process should account for any baseline differences between a trial's groups [20].

Evidence from methodological research published in the last decade, which quantifies bias in treatment effect size estimates in a series of meta-analyses, has found associations between inadequate randomization or inadequate allocation concealment and inflated treatment effect size estimates of 11% [2] to 51% [21] and 10% [22] to 52% [21], respectively. These findings varied based on the medical area examined and the statistical modeling used. Nevertheless, these associations were not found in some other reports [23, 24]. Similarly, associations between baseline comparability and treatment effect size estimates were not confirmed in two reports [24, 25]. Therefore, an accurate assessment of the risk of selection bias present in RCTs is critical during the synthesis and examination of research findings to allow conclusions to be relevant and important and to decipher potential recommendations, which have the potential to guide clinical decision making and future research in the field [26, 27].

To date, methodological factors, including randomization and allocation

concealment which are used to assess risk of bias, are derived only from meta-epidemiological studies of RCTs in medicine. Notably, a core set of items geared to assess the risk of selection bias has been identified by many meta-epidemiological studies in medical subspecialty fields such as cardiovascular disease, pediatrics, and surgery [28-33]. However, these meta-epidemiological studies are based mostly on dichotomous outcomes such as all-cause mortality and presence of events; only a few meta-epidemiological studies have focused on continuous outcomes [34, 35]. Therefore, generalizable evidence from meta-epidemiological studies is limited because current studies include only a few medical specialties and the studies generally employ small meta-analytic sample sizes [36]. This observation was supported by Berkman et al. [36] who recently reviewed published meta-epidemiological studies and concluded that the majority of these investigations were underpowered, leading to nonsignificant findings that did not accurately reflect associations between selection bias and estimates of the size of treatment effects. For example, the only pilot study conducted [37] at the level of meta-analyses and in the domain of oral health research quantified bias associated with periodontal trials in only three meta-analyses. That study concluded that there were no differences in the magnitude of treatment effect estimates between trials based on risk of bias related to sequence generation and allocation concealment. The current study addresses the aforementioned shortcomings.

The overall aim of this methodology study is to examine the impact of selection bias on the magnitude of treatment effect size estimates in oral health randomized clinical trials. The specific objectives were to: (1) examine associations between treatment effect size estimates and the adequacy of sequence generation, the adequacy of allocation concealment, and the baseline comparability; and (2) determine the impact of potential additional factors such as dental specialty, type of treatment, type of outcome (objective vs. subjective), and the heterogeneity of meta-analyses on treatment effect size estimates. The study results have the potential to strengthen how

oral health RCTs are designed, conducted, and reported.

6.2. Methods and Analysis

6.2.1. Protocol and registration

The protocol of this meta-epidemiological study was registered on PROSPERO (CRD42014014070); the manuscript of this protocol was peer-reviewed and published *a priori* [38].

6.2.2. Selection of meta-analyses and randomized trials

Selection of meta-analyses

Inclusion criteria for the available meta-analyses were:

1. The meta-analysis was in the field of dental, oral, or craniofacial research, and evaluated a therapeutic intervention related to oral health specialties as defined by the American Dental Association (ADA) scope of dental practice [39].
2. The meta-analysis examined a minimum of one continuous outcome, and included at least five randomized clinical trials with quantitative data of treatment effect size estimates.

Selection of trials

The randomized trials selected were included in the eligible meta-analyses and met the following inclusion-exclusion criteria:

1. The study was reported as a randomized clinical trial [8] and findings were reported in a way that allowed the calculation of treatment effect size.

2. The effect of a placebo, no-treatment, or a standard care control was compared with the effect of the intervention under investigation; however, randomized clinical trials comparing two active interventions (not including the standard care intervention) were excluded.
3. The studies examined therapeutic interventions included in at least one of the nine dental specialties recognized by the ADA [39].

6.2.3. Information sources and literature search

Six electronic databases were used to perform the searches from database inception to May 2014 (PubMed, MEDLINE, EMBASE, ISI Web of Science, Evidence-Based Medicine Reviews–Cochrane Database of Systematic Reviews, and Health STAR). Keywords used in the searches were "systematic review," "meta-analysis," "oral surgery," "endodontics," "pediatric dentistry," "dental public health," "periodontics," "prosthodontics," "pedodontics," "dentistry," "tooth," "orthodontics," and "oral pathology." A health information specialist helped to refine the search strategy. More details on specific search terms and specific combinations used in each individual database, are presented in **Appendix 3-A**. In addition, we searched the American Dental Association-Evidence-based Dentistry website [39], and reference lists of potentially relevant articles that met the inclusion criteria were hand-searched. The searches were not limited to the English language, nor were they restricted by other means.

The titles and abstracts retrieved from implementing the search strategy were screened by two independent assessors with dental research and clinical backgrounds. Citations of systematic reviews deemed potentially relevant were selected, and the full-text articles were retrieved for complete screening. The final eligibility of full texts was determined by the same two reviewers, with disagreements resolved through consensus.

6.2.4. Data extraction

Five reviewers from diverse health research areas (described below) formed the review panel that performed data extraction. Two team members (H.S.; S.A.) conducted reviewer training to ensure consistency during the data extraction phase. Training consisted of the review panel reviewing 10 RCTs not included in the final set of trials, followed by assessment and feedback. Our research team has used a similar reviewer training process in other investigations [40, 41]. Data extraction was performed only when there was agreement on extraction protocols and interpretations among review panel members.

Data were extracted in duplicate by two independent assessors, with disagreements resolved through a consensus meeting. The first assessor (H.S.), a clinician with an oral health research background, performed a complete data extraction (n = 540, 100%), while the second assessor from the review panel had a medical research background. In cases of disagreement, where consensus could not be reached, a third assessor (S.A.) was solicited to aid in achieving complete consensus (one reviewer had PhD degree in rehabilitation medicine, one had a master's degree in public health, and one had a bachelor's degree in health sciences). Only data that attained complete consensus were used for data analyses. The structured data extraction template, designed in Microsoft Office Access, was pilot-tested and was used to record data extraction. The following sections describe details of the data extracted from the selected RCTs.

<u>Nonmethodological characteristics</u>

At the meta-analysis level, the following data elements were extracted: publication year, dental specialty (e.g., dental public health, endodontics, oral medicine and oral pathology, oral and maxillofacial radiology, oral and maxillofacial surgery, orthodontics and dentofacial orthopedics, pediatric dentistry, periodontics,

prosthodontics, and restorative dentistry), primary outcome assessed, type of comparison (e.g., active intervention vs. placebo, standard care, or no-treatment intervention), and number of included trials.

Additionally, at the randomized trial level, the following elements were extracted from the RCTs: publication year, study design (e.g., parallel, split-mouth, crossover, or factorial), type of outcome (e.g., subjective vs. objective [42]), and number of centers (e.g., multicenter vs. single center).

Risk of selection bias

To assess the potential for selection bias in the selected RCTs, three methodological domains of the Cochrane Collaboration's risk of bias tool relevant to selection bias [27] were examined: random sequence generation, allocation concealment, and baseline comparability. Using the Cochrane tool as a guideline, the definition and method of each criterion used were derived and a three-part answering scheme (i.e., high, low, or unclear) was employed (see **Appendices 5.A, 5.B, 5.C,** and **5.D** for further details on the definition of items used). In addition, the methods of random sequence generation and allocation concealment implemented in included RCTs were classified [41]. For random sequence generation, three classes were set: Class I – trials with randomization (*Division 1* involved an adequate method of randomization, such as the use of computer software, minimization, and a random number table; *Division 2* involved trials with satisfactory, but less adequate than Division 1, methods of randomization, such as drawing lots, shuffling cards, and envelopes); Class II – trials with inadequate methods of sequence generation (e.g., date of birth, hospital record number, and day of admission); and Class III – trials with unclear (or unreported) random sequence generation.

Similarly, the allocation concealment method was grouped into three classes: Class I – trials with allocation concealment (*Division 1* involved trials using central

randomization. *Division 2* involved trials using sequentially numbered, sealed, and opaque envelopes); Class II – trials where allocation was not concealed, such as using a open list or a participant's date of birth; and Class III – trials with unclear (or unreported) allocation concealment. For the meta-epidemiological analysis, the three classes (in both sequence generation and allocation concealment) were further grouped into two broad categories: an "adequate" category (Class I) and an "inadequate or unclear" category (Classes II and III).

Furthermore, we categorized "baseline comparability" regarding important prognostic indicators (whether the groups were similar at the start of the trial) into three categories: (1) low risk of bias, where the trial's authors performed a comparison between groups through a statistical test or adjusted any differences statistically; (2) high risk of bias, where the trial's authors reported that groups were not equal at the baseline, and they did not adjust for any difference; and (3) unclear risk of bias, where the trial's authors did not report sufficient information to permit a "Yes" or "No" judgment. For the meta-epidemiological analysis, we classified the trial as having comparable groups (low risk of bias) versus both not having comparable groups (high risk of bias) and not reporting baseline comparability (unclear risk of bias).

Treatment effect size (ES) estimate

Data on means, measures of variability (SDs and 95% CIs), and sample sizes were extracted for the primary outcome of the review. If the review's primary outcome was not continuous, or its meta-analysis did not include at least five trials, data were extracted on the continuous outcome of the meta-analysis with the largest number of included trials.

6.2.5. Sample size calculation

Because the sample size calculation approach for meta-epidemiological

investigations is not well-established in the literature, our study sample size was estimated based on recommendations in studies by Berkman et al. [36], Hempel et al. [43], and Fenwick et al. [44]. The majority of published meta-epidemiological reports have been reported to be highly heterogeneous and underpowered, mainly because of their small sample sizes [36]. From previous meta-epidemiological investigations [34], we anticipate obtaining a difference in treatment effect size estimate of at least 0.15 (SE = 0.087) between trials with and without methodological limitations [34, 41]. This magnitude of difference in treatment effect size estimate has been claimed to resemble nearly 1/4 to 1/2 of a typical treatment effect size estimate for interventions in fields similar to the field of dentistry [34, 44]. Accordingly, we employed a sample size of more than 500 RCTs that were the subjects of more than 60 systematic reviews with meta-analyses to demonstrate a meaningful difference. This is two to three times the number of trials included in previously published meta-epidemiological investigations [34, 35, 45].

6.2.6. Data analysis

To illustrate the methodological characteristics of the RCTs included in this study, descriptive analyses were conducted (e.g., proportions and percentages for categorical data, such as the proportion of trials) according to allocation concealment methods).

To examine associations between the adequacy of randomization, the allocation concealment, the baseline comparability and treatment effect estimates, a two-level analysis was conducted, using a meta-meta-analytic approach with a random-effects model, based on recommendations in Sterne et al. [46]. This type of statistical analysis was appropriate because the methodology used for our meta-epidemiological analysis accounts for the heterogeneity between RCTs, within meta-analyses in the first step, and among meta-analyses in the second step [33, 47]. For the

first level of analysis (within meta-analyses), standard treatment effect size estimates were extracted from the primary outcome of each randomized trial, as described in Cohen [48], where a negative treatment effect size estimate implies a beneficial effect of the interventional group. The type of comparison (e.g., treatment, control) was classified based on the authors' classification of the comparison implemented in the meta-analysis reported in the review, and our classification was cross-checked by the principal assessor. Data from the RCTs in each of the selected meta-analyses were used. If a randomized trial was found in more than one meta-analysis, it was used once in the meta-analysis with the fewest number of studies evaluated. The raw data for each trial were obtained from each meta-analysis, and cross-checked with data reported in the primary trial. During this analysis process, the meta-analyses and RCTs were divided into two groups—those that adequately addressed the items (e.g., adequate allocation concealment) and those that did not adequately address the items (e.g., "no" or unclear allocation concealment) for each of the evaluated quality items.

Two treatment effect size estimates were calculated for each meta-analysis: one including all the studies that reported the characteristic of interest (e.g., allocation concealment), and one for studies that did not report that characteristic. A negative difference in treatment effect size implied that trials with the characteristic of interest (e.g., adequate allocation concealment) had a more favorable effect for the tested intervention group. Inverse-variance random-effects meta-analysis was used to derive pooled treatment effect size estimates for each of the meta-analyses. The DerSimonian and Laird estimate of variance was then calculated to determine heterogeneity among the RCTs [47]. A meta-regression technique was used for each of the meta-analyses to derive the difference between pooled treatment effect size estimates of the studies with and without the characteristic of research interest.

The second level of analysis (among the meta-analyses) entailed pooling the results of the previous analysis (combined differences from all meta-analyses) to

describe the treatment effect size estimate in each trial component across all meta-analyses. Treatment effect size estimates were combined at this stage, using inverse-variance random-effects meta-analysis [47], to account for between-meta analysis heterogeneity. The DerSimonian and Laird estimate of variance was calculated to determine the heterogeneity between meta-analyses [47], while all p-values were two-sided.

To determine the impact of potential additional factors (dental specialty, type of outcome, magnitude of treatment effect estimate, effect of heterogeneity of meta-analysis on treatment effect size estimates), the meta-epidemiological analysis was stratified. Interaction tests accompanying these stratified analyses were based on Z-scores according to the following factors: magnitude of treatment effect estimate (large, if $\leq$ -0.5, vs. small, if > -0.5), heterogeneity of meta-analysis (high, if $\tau^2 \geq 0.06$, vs. low, if $\tau^2 < 0.06$; this cut-off roughly corresponds to a difference of 1 between the largest and the smallest treatment effect size estimate), type of outcome (objective vs. subjective), and dental speciality (dental public health vs. other interventions or periodontal vs. other interventions). This meta-epidemiological approach was performed with STATA statistical software version 14 (College Station, TX: StataCorp LP).

6.3. Results

6.3.1. Characteristics of selected meta-analyses and randomized trials

From 1408 reviews included in the Oral Health Database of Systematic Reviews [49] (which included dental, oral, and craniofacial reviews published between 1955 and 2014), 1256 records were excluded based on the information provided in the title or abstract. The remaining 152 full-text reports were retrieved for a more detailed evaluation, of which 64 systematic reviews with meta-analyses

fulfilled the eligibility criteria. The complete list of excluded records and reasons for exclusion are available upon request. Ultimately, 64 meta-analyses including 540 randomized clinical trials analyzing 137,957 patients contributed to this investigation.

The 64 chosen therapeutic systematic reviews with meta-analyses were published between 2002 and 2014 (median year of publication: 2010; IQR: 2006, 2012), of which 34.4% were Cochrane reviews (n = 22). The chosen meta-analyses included a median of six trials (interquartile range [IQR] 6–10), published in the fields of periodontics and implantology [50-85], dental public health and pediatric dentistry [86-95], oral medicine and pathology [96-106], oral and maxillofacial surgery [107-110], orthodontics and dentofacial orthopedics [111, 112], and restorative dentistry [113].

Nearly one third of the trials were placebo-controlled (n = 204; 37.8%), and two thirds of those examined involved nonsurgical (n = 370; 68.5%) or nondrug (n = 359; 66.5%) interventions. One fifth of the trials were multicenter trials, where the majority of the trials used parallel design (n = 372; 68.9%), and one quarter used split-mouth design (n = 126; 23.3%). **Table 6.1** provides further details on the characteristics of the chosen meta-analyses.

6.3.2. Impact of inadequate sequence generation on treatment effect size estimate

Fifty-two meta-analyses, including 467 trials involving 133,055 patients, provided information for this meta-epidemiological analysis. **Figure 6.1a** displays a forest plot of the difference in treatment effect estimates between randomized trials with adequate and inadequate sequence generation. A positive value (> 0) across meta-analyses indicated that trials with inadequate sequence generation had inflated treatment effect size estimates compared with trials that used appropriate sequence generation. Results of the analysis showed that trials with inadequate sequence generation had significantly larger treatment effect size estimates (difference in effect

size = 0.13, 95% confidence interval 0.01 to 0.25, p = 0.037). This result means that treatment effect size estimates were 0.13 larger in trials with inadequate sequence generation than treatment effect size estimates in trials with adequate sequence generation.

The results of the stratified analyses showed that differences in treatment effect size estimates between trials with adequate and inadequate sequence generation were significant ($p < 0.001$) in meta-analyses with a large treatment benefit in overall meta-analysis, but not in meta-analyses with a small treatment benefit. However, none of the other factors considered (heterogeneity of meta-analysis, type of outcome, dental specialty) had a statistically significant interaction (see **Figure 6.1b)**.

6.3.3. Impact of inadequate allocation concealment on treatment effect size estimate

Thirty-nine meta-analyses, including 345 trials involving 110,797 patients, provided information for this meta-epidemiological analysis. **Figure 6.2a** shows a forest plot of the difference in treatment effect size estimates between randomized trials with adequate and inadequate allocation concealment. A positive value (> 0) across meta-analyses indicated that trials with inadequate allocation concealment inflated the treatment effect size estimate when compared to trials with adequate allocation concealment. Our meta-epidemiological results showed that treatment effect size estimates were significantly larger in RCTs with inadequate allocation concealment than treatment effect size estimates in RCTs with adequate allocation concealment (difference in treatment effect size estimate = 0.15, 95% confidence interval 0.02 to 0.27, p = 0.022).

The results of the stratified analyses showed that differences in treatment effect size estimates between trials with adequate and inadequate allocation concealment were significant ($p < 0.001$) in meta-analyses with a large treatment benefit in overall meta-analysis, but not in meta-analyses with a small treatment benefit. However, the

impact of allocation concealment on treatment effect size estimate stratified by other considered factors (heterogeneity of meta-analysis, type of outcome, dental specialty) was not statistically significant (see **Figure 6.2b)**.

6.3.4. Impact of baseline comparability on treatment effect size estimate

Thirty-two meta-analyses, including 310 trials involving 121,213 patients, provided information for this meta-epidemiological analysis. **Figure 6.3a** shows a forest plot of the difference in treatment effect size estimates between RCTs with balanced and imbalanced baseline characteristics. Results of the analysis showed there was no statistically significant difference between treatment effect size estimates in RCTs with balanced or imbalanced baseline characteristics (difference in effect size = 0.01, 95% confidence interval -0.09 to 0.12, p = 0.804). The results of the stratified analyses showed that none of the stratifying factors had a statistically significant interaction (see **Figure 6.3b)**.

6.4. Discussion

To our knowledge, this investigation is the first large meta-epidemiological study conducted in the domain of dental, oral, and craniofacial research, and one of the very few meta-epidemiological studies conducted in any medical field that examines continuous outcomes and/or employs adequate sample size to examine the impact of selection bias on treatment effect size estimates in RCTs. Thus, it provides evidence that is novel and of high priority and interest to researchers and methodologists, particularly in the field of oral health research.

A potential limitation of previous meta-epidemiological studies is that many of these studies are “underpowered,” this might have led to nonsignificant findings that are not true reflections of potential associations between a trial’s quality and its treatment effect size estimate [36]. We included a large number of meta-analyses,

several trial designs (including the split-mouth design), and all dental specialties; this should increase the statistical power and precision of the analysis, and assure the generalizability of the results. The number of trials assessed in our study was two to three times larger than that used in preceding meta-epidemiological studies conducted in other medical fields. Furthermore, the strict methodology applied to data collection and data analysis, based on previous meta-epidemiological work by our research group, addressed potentially limiting factors associated with this type of methodological research.

Our study showed that over two thirds of the trials did not clearly report either sequence generation and/or allocation concealment, with 67.6%, 84.8%, and 22.2% of trials judged as having "unclear" bias in sequence generation, allocation concealment, and baseline comparability, respectively. While findings in the sequence generation domain were in agreement with findings of recent methodology studies [30, 41], the proportion of trials that did not clearly report allocation concealment (84%) was higher than that reported in two recent studies by Armijo-Olivo et al., in the field of rehabilitation medicine [41] (71.8%, n = 393), and Hartling et al., in the field of pediatrics [45] (78.8%, n = 287). It was recently reported [29] that the majority of trial protocols with unclear allocation concealment also had unclear allocation concealment in the published trials, and if only studies with appropriate allocation concealment were included in reviews, nearly two thirds of the conclusions would have lost the beneficial effects of the intervention. This is concerning because of the increased potential for bias due to limitations in study protocols in oral health trials; therefore, potential clinical decision-making policy in dental practice may be compromised.

Our study determined that (1) treatment effect size estimates were 0.13 larger in trials with inadequate sequence generation than in trials with adequate sequence generation, and (2) treatment effect size estimates were 0.15 larger in trials with inadequate allocation concealment than in trials with adequate allocation concealment.

Thus, inadequate sequence generation and inadequate allocation concealment inflated treatment effect size estimates by about 1/5 to 1/4 of the common treatment effect size estimate reported in oral health research [114], such as clinical outcomes in periodontology [44]. However, baseline comparability was not associated with inflated or underestimated treatment effect size estimates in our study.

With respect to the direction and the magnitude of the treatment effect size estimate, our results agree with those in studies that showed that inadequate sequence generation could exaggerate treatment effect size estimates by 51% [21], 36% [115], and 11% [2, 6], depending on the medical field examined. Similarly, inadequate allocation concealment has been associated with increased treatment effect size estimates of 52% [21], 34% [46], and 10% [22], compared with adequately concealed trials. However, this association was not confirmed in some studies [23, 24]. Previous reports examining the effects of inadequate sequence generation and/or inadequate allocation concealment, have been restricted to RCTs in specific medical areas, such as pediatrics [45], low-back pain [25], osteoarthritis [34], and physical therapy [41]. These reports defined allocation concealment according to the Schulz tool for allocation concealment [2, 46, 116] and sequence generation according to the Jadad scale [6, 115, 117], the Cochrane Handbook [8, 118], or the Cochrane Collaboration's risk of bias tool [27, 41, 45].

While the above-mentioned studies [2, 6, 21, 23, 24, 115] assessed dichotomous outcomes, two recent studies [41, 45] examined the association between treatment effect size estimate and inadequate sequence generation and inadequate allocation concealment, using continuous outcomes. One of these studies [45] included 287 pediatric trials from 17 meta-analyses and found no significant difference in treatment effect size estimates between trials that employed adequate or inadequate sequence generation, and no significant difference in treatment effect size estimates between trials that employed adequate or inadequate allocation concealment.

Another study [119] assessed 275 physical therapy trials included in 22 meta-analyses, and showed a significant trend toward an exaggeration of treatment effect size estimates in trials with inadequate allocation concealment compared to trials with adequate allocation concealment, while there was no difference in treatment effect estimates in trials with adequate or inadequate sequence generation. However, while a meta-epidemiological analysis requires a large number of meta-analyses and trials, this was not the case in the majority of the above-mentioned studies. These inconsistent findings might be attributed to the use of different statistical approaches (e.g., logistic regression, weighted regression, or the Bayesian model [24, 120]); the assessment of different types of interventions, outcomes, and populations [36]; and the improper inclusion of trials with comparable active interventions (where identification of the direction of treatment effect is difficult). Improper inclusion of trials with comparable active interventions leads to inaccuracy when analyzing differences in treatment effect size estimates.

Our results showing an insignificant influence of imbalance in the baseline characteristics on treatment effect size estimates are in agreement with two studies [47, 123] that examined associations between baseline comparability and treatment effect size estimates. Two other studies [24, 25] showed no significant differences in treatment effect estimates based on potential bias related to an imbalance in baseline characteristics. One study [24] assessed baseline comparability in 256 pediatric trials from 24 meta-analyses in numerous medical domains (surgery, pediatrics, cardiovascular disease, and infectious diseases); the other study [25] assessed baseline comparability in 216 randomized trials included in 15 Cochrane reviews in the field of back pain. The insignificant association between imbalances in baseline characteristics and treatment effect size estimates found in our study and the above-mentioned reports [24, 25] could be due to the inclusion of only randomized trials, where baseline imbalances arise accidentally or by chance, and thus should not conceptually affect treatment effect size estimates [20, 36]. Randomized trial authors should report

baseline characteristics of patients allocated to each intervention, and judge whether a specific characteristic that was imbalanced among interventions has impacted the trial's findings [121].

6.4.1. Strengths and limitations of the study

This meta-epidemiological study provides an empirical analysis of the association between treatment effect size estimates and selection bias, in the domain of oral health research. The study has several limitations. First, the empirical evidence examined published studies only (bias was based on reported methodological characteristics), and did not evaluate the actual conduct of the RCTs. Accordingly, data extraction and analyses were based on the information given by the authors in the published reports. This approach, although widely used, limits the identification of actual bias if trial authors do not adequately report study elements. For example, evidence from a methodology study [122] showed that adequate allocation concealment and adequate sequence generation were reported in 18% and 21%, respectively, of trial publications only, and in 44% and 36%, respectively, of trial protocols. Second, the authors of the trials were not contacted for missing data given that a large proportion of the trials were published before the year 2005 when corresponding authors' information was not provided or not up-to-date.

Certain levels of heterogeneity are expected in any meta-epidemiological examination of the impact of bias on treatment effect size estimates. Such studies analyse numerous entities (meta-analysis, trials, and participants) that have a distinct potential for heterogeneity [36]. By applying in this study a cautious methodology to data collection and analysis, and by assembling a large number of meta-analyses and trials, the study power was increased and heterogeneity was reduced. Data analysis was restricted to trials in which the direction of the expected treatment effect was clear, including trials involving a control or placebo intervention. This procedure

reduced expected heterogeneity and confounding factors in the analyses, allowing for the detection of significant effects of methodological characteristics. The inclusion of trials with only comparably active interventions (e.g., comparison of two different types of dental implants) would potentially lead to countering treatment effects, eventually cancelling differences in treatment effect size estimates.

This study did not assess the (likely) effects of interactions with other design biases. Such an assessment would have to include a multivariate analysis with a larger number of meta-analyses and trials [123]. Future meta-epidemiological assembling of a greater number of meta-analyses and trials by synthesizing results from different disciplines and datasets should take other design biases into account.

6.5. Conclusions

Significant differences in treatment effect size estimates were identified in oral health trials based on the adequacy or inadequacy of sequence generation and the adequacy or inadequacy of allocation concealment. Trials with inadequate sequence generation or inadequate allocation concealment reported significantly larger treatment effect size estimates compared to trials that employed adequate sequence generation or adequate allocation concealment. Treatment effect sizes estimates were 0.13 and 0.15 larger in trials with inadequate sequence generation and inadequate allocation concealment, respectively, than in trials with adequate sequence generation and adequate allocation concealment, respectively. The baseline comparability was not associated with inflated or underestimated treatment effect size estimates.

Based on this evidence, authors of systematic reviews may consider excluding trials (conducted in the domains of dental, oral, and craniofacial research) with inadequate sequence generation and/or inadequate allocation concealment from meta-analyses, or should perform sensitivity analyses based on the adequacy of sequence

generation and the adequacy of allocation concealment in trials. Because of the expected impact of bias on treatment effect size estimates, dental journal editors and reviewers should insist on adequate sequence generation and adequate allocation concealment in the conduct and reporting of RCTs submitted for publication.

Table 6.1. Details of the meta-analyses included in the study

Title of Meta-Analysis	Author and Year	Primary Dental Specialty	Outcome	Outcome Type	Comparison	No. of Trials
Full-mouth disinfection for the treatment of adult chronic periodontitis	Eberhard 2008 [50]	Periodontics	Bleeding on probing	Subjective	Full-mouth scaling vs. control	5
Treatment of gingival recession with coronally advanced flap procedures	Cairo 2008 [51]	Periodontics	Gingival recession	Subjective	Coronally advanced flap plus enamel matrix derivative vs. coronally advanced flap	5
Effectiveness of systemic amoxicillin /metronidazole as an adjunctive therapy to full-mouth scaling and root planing in the treatment of aggressive periodontitis	Sgolastra 2012 [52]	Periodontics	Clinical attachment levels	Subjective	Full-mouth scaling plus combined amoxicillin-metronidazole vs. full-mouth scaling	6
Absorbable collagen membranes for periodontal regeneration	Stoecklin 2013 [53]	Periodontics	Clinical attachment levels	Subjective	Collagen membranes vs. control	11
An evaluation of bioactive glass in the treatment of periodontal defects	Sohrabi 2012 [54]	Periodontics	Clinical attachment levels	Subjective	Bioactive glass vs. control	14
Platform switching for marginal bone preservation around dental implants	Atieh 2010 [55]	Implantology	Marginal bone level	Subjective	Platform switch vs. platform match	5
Peri-implant marginal bone level	Annibali 2012 [56]	Implantology	Marginal bone level	Subjective	Platform switch vs. platform match	5
Is platelet concentrate advantageous for the surgical treatment of periodontal diseases?	Del Fabbro 2011 [57]	Periodontics	Clinical attachment levels	Subjective	Platelet-rich plasma vs. control	10
The effectiveness of a toothpaste containing triclosan and polyvinyl-methyl ether maleic acid copolymer in improving plaque control and gingival health	Davies 2004 [58]	Periodontics	Gingival index	Subjective	Toothpaste vs. control	6
Scaling and root planing treatment for periodontitis to reduce preterm birth and low birth weight	Kim 2012 [59]	Periodontics	Birth weight	Objective	Periodontal treatment vs. control	6
Lasers for the treatment of dentin hypersensitivity	Sgolastra 2013 [96]	Oral medicine and pathology	Pain	Subjective	Laser vs. placebo	13

Combinations of topical fluoride (toothpastes, mouth rinses, gels, varnishes) versus single topical fluoride for preventing dental caries in children and adolescents	Marinho 2004 [86]	Dental public health and pediatric dentistry	The decayed, missing, and filled surfaces (DMFS) index	Subjective	Fluoride toothpaste plus mouth rinse (or gel) vs. fluoride toothpaste	6
Enamel matrix derivative (Emdogain®) for periodontal tissue regeneration in intrabony defects	Esposito 2009 [61]	Periodontics	Probing attachment level	Subjective	Emdogain vs. control	9
Guided tissue regeneration for periodontal infra-bony defects	Needleman 2006 [60]	Periodontics	Attachment gain	Subjective	Guided tissue regeneration vs. control	13
The efficacy of dental floss in addition to a toothbrush on plaque and parameters of gingival inflammation	Berchier 2008 [62]	Periodontics	Gingival index	Subjective	Floss plus toothbrushing vs. toothbrushing only	5
Dentin hypersensitivity and oxalates	Cunha-Cruz 2011 [106]	Oral medicine and pathology	Dentin hypersensitivity	Subjective	Oxalate vs. placebo or no-treatment	11
The efficacy of 0.12% chlorhexidine mouth rinse compared with 0.2% on plaque accumulation and periodontal parameters	Berchier 2010 [63]	Periodontics	Plaque index	Subjective	0.12% vs. 0.2% chlorhexidine mouth rinse	7
Fluoride toothpastes of different concentrations for preventing dental caries in children and adolescents	Walsh 2010 [87]	Dental public health and pediatric dentistry	The decayed, missing, and filled surfaces (DMFS) index	Subjective	Fluoride toothpaste vs. placebo	19
A systematic review with meta-analysis of the effect of low-level laser therapy (LLLT) in cancer therapy-induced oral mucositis	Bjordal 2011 [97]	Oral medicine and pathology	Oral mucositis severity	Subjective	low-level laser therapy vs. placebo	6
Emergency management of acute apical periodontitis in the permanent dentition	Sutherland 2003 [64]	Periodontics	Pain	Subjective	Treatment vs control	5
The effectiveness of splint therapy in patients with temporomandibular disorders	Ebrahim 2012 [98]	Oral medicine and pathology	Pain	Subjective	Splint therapy vs. minimal/ no-treatment	10
Fluoride varnishes for preventing dental caries in children and adolescents	Marinho 2002 [95]	Dental public health and pediatric dentistry	The decayed, missing, and filled surfaces (DMFS) index	Subjective	Fluoride varnish vs. placebo or no treatment	7
A review of the effects of stannous fluoride on gingivitis	Paraskevas 2006 [65]	Periodontics	Gingival index	Subjective	SnF2 dentifrices vs. NaF	6

Manual versus powered toothbrushing for oral health	Robinson 2005 [90]	Dental public health	Gingival index	Subjective	Side to side powered toothbrushes vs. manual toothbrushes	8
Fluoride toothpastes for preventing dental caries in children and adolescents	Marinho 2003 (2) [89]	Dental public health and pediatric dentistry	The decayed, missing, and filled teeth (DMFT) index	Subjective	Fluoride toothpaste vs. placebo	40
Oral hygiene care for critically ill patients to prevent ventilator-associated pneumonia	Shi 2013 [99]	Oral medicine and pathology	Duration of ventilation	Objective	Chlorhexidine vs. placebo or usual care	6
In-office treatment for dentin hypersensitivity	Lin 2013 [100]	Oral medicine and pathology	Pain	Subjective	Physical occlusion (e.g., Pumice paste) vs. chemical occlusion (e.g., fluorides oxalates)	6
Triclosan/copolymer containing toothpastes for oral health	Riley 2013 [66]	Periodontics	Plaque index	Subjective	Triclosan or copolymer vs. control	10
Interventions for the management of dry mouth: topical therapies	Furness 2011 [101]	Oral medicine and pathology	Mouth dryness scale	Subjective	Saliva substitutes A vs. B	6
Guided tissue regeneration for the treatment of periodontal intrabony and furcation defects	Murphy 2003 [67]	Periodontics	Clinical attachment levels	Subjective	Open flap debridement vs. guided tissue regeneration with barrier	6
Can subepithelial connective tissue grafts be considered the gold standard procedure in the treatment of Miller Class I and II recession-type defects?	Chambrone 2008 [68]	Periodontics	Clinical attachment levels	Subjective	Guided tissue regeneration and membrane vs. subepithelial connective tissue graft	7
Surgical protocols for ridge preservation after tooth extraction	Vignoletti 2012 [107]	Oral and maxillofacial surgery	Marginal bone level	Subjective	Surgical procedure vs. control	7
Home-based chemically-induced whitening of teeth in adults	Hasson 2006 [113]	Restorative dentistry	Color	Subjective	Whitening product vs. whitening product - Colorimetric	6
The effect of flapless surgery on implant survival and marginal bone level	Lin 2013 [69]	Implantology	Marginal bone level	Subjective	flapless vs. flap procedures	5
Impact of implant support for mandibular dentures on satisfaction, oral and general health-related quality of life	Emami 2009 [70]	Implantology	Patient satisfaction	Subjective	Mandibular implant overdentures vs. conventional dentures	6

Systematic review on the effect of rinsing with povidone-iodine during nonsurgical periodontal therapy	Sahrmann 2010 [71]	Periodontics	Periodontal probing depth	Subjective	Povidone-iodine rinsing vs. control	5
Potassium containing toothpastes for dentine hypersensitivity	Poulsen 2006 [102]	Oral medicine and pathology	Tactile	Subjective	Potassium nitrate (no fluoride) vs. placebo (no potassium nitrate plus/-fluoride)	5
Sedation of children undergoing dental treatment	Lourenço-Matharu 2012 [91]	Pediatric dentistry	Houpt/other behavioral score	Subjective	Sedatives vs. placebo	6
Meta-analysis of local tetracycline in treating chronic periodontitis	Pavia 2003 [72]	Periodontics	Periodontal probing depth	Subjective	Tetracycline vs. placebo	6
Efficacy of periodontal treatment on glycaemic control in diabetic patients	Darré 2008 [73]	Periodontics	Glycated haemoglobin HbA1c	Objective	Periodontal treatment vs. control	6
Surgical Techniques for the removal of mandibular wisdom teeth	Coulthard 2014 [110]	Oral and maxillofacial surgery	Swelling	Subjective	Primary vs. secondary wound closure	6
Psychological treatment of dental anxiety among adults	Wide Boman 2013 [103]	Oral medicine and pathology	Dental anxiety scale	Subjective	Behavioral therapy vs. control	5
Systemic interventions for recurrent aphthous stomatitis (mouth ulcers)	Brocklehurst 2012 [104]	Oral medicine and pathology	Pain	Subjective	Intervention versus control	5
Psychosocial interventions for the management of chronic orofacial pain	Aggarwal 2011 [105]	Oral medicine and pathology	Pain	Subjective	Any psychosocial intervention vs. usual care	7
Interventions for replacing missing teeth: different types of dental implants	Esposito 2007 [74]	Implantology	Bone level	Subjective	Treatment vs. another treatment	10
Flossing for the management of periodontal diseases and dental caries in adults	Sambunjak 2011 [75]	Periodontics	Gingival index	Subjective	Toothbrushing plus flossing vs. toothbrushing alone	6
Efficacy and co-morbidity of oral appliances in the treatment of obstructive sleep apnea-hypopnea	Hoekema 2004 [111]	Orthodontics and dentofacial orthopedics	Apnea-Hypopnea index	Objective	Mandibular repositioning appliance vs. continuous positive airway pressure	6
Adjunctive photodynamic therapy to non-surgical treatment of chronic periodontitis	Sgolastra 2013 (2) [76]	Periodontics	Clinical attachment level	Subjective	Scaling root planing plus antimicrobial photodynamic therapy vs. scaling root planing	11

Fluoride mouth rinses for preventing dental caries in children and adolescents	Marinho 2003 [92]	Dental public health and pediatric dentistry	The decayed, missing, and filled surfaces (DMFS) index	Subjective	Fluoride mouth rinse vs. placebo or no-treatment	26
The use of enamel matrix derivative alone versus in combination with bone grafts to treat patients with periodontal intrabony defects	Li 2012 [77]	Periodontics	Clinical attachment levels	Subjective	Enamel matrix derivative plus bone grafts vs. enamel matrix derivative alone	5
Treatment of periodontitis improves the atherosclerotic profile	Teeuw 2014 [78]	Periodontics	hsCRP levels	Objective	Periodontal treatment vs. control	13
The effect of cetylpyridinium chloride-containing mouth rinses as adjuncts to toothbrushing on plaque and parameters of gingival inflammation	Haps 2008 [79]	Periodontics	Plaque index	Subjective	Cetylpyridinium chloride mouth rinses vs. brushing only	6
Different powered toothbrushes for plaque control and gingival health	Deacon 2010 [80]	Periodontics	Gingival index	Subjective	Side to side vs. rotation oscillation	6
Treatment of class II molar furcation involvement	Kinaia 2011 [81]	Periodontics	Bone level	Subjective	Non-resorbable vs. resorbable membranes	5
The long-term effect of a mouth rinse containing essential oils on dental plaque and gingivitis	Stoeken 2007 [82]	Periodontics	Plaque index	Subjective	Mouth rinse containing essential oils rinse vs. control	6
Fluoride gels for preventing dental caries in children and adolescents	Marinho 2002 (2) [88]	Dental public health and pediatric dentistry	The decayed, missing, and filled surfaces (DMFS) index	Subjective	Fluoride gel vs. placebo or no treatment	19
Corticosteroids reduce postoperative morbidity after third molar surgery	Markiewicz 2008 [108]	Oral and maxillofacial surgery	Late edema	Subjective	Corticosteroids vs. placebo	6
Pharmacological management of pain during orthodontic treatment	Angelopoulou 2012 [112]	Orthodontics and dentofacial orthopedics	Pain	Subjective	Ibuprofen vs. placebo	6
Interventions for replacing missing teeth: different times for loading dental implants	Esposito 2013 [83]	Implantology	Marginal bone level	Subjective	Immediate vs. conventional loading	8
Fluoride toothpaste efficacy and safety in children younger than 6 years	Wright 2014 [93]	Dental public health and pediatric dentistry	The decayed, missing, and filled teeth (DMFT) index	Subjective	Fluoride toothpastes vs. control	8

The efficacy of bone replacement grafts in the treatment of periodontal osseous defects	Reynolds 2003 [84]	Periodontics	Bone level	Subjective	Bone replacement grafts vs. open flap debridement defects	18
Secondary versus primary closure techniques for preventing postoperative complications following removal of impacted mandibular third molars	Carrasco-Labra 2012 [109]	Oral and maxillofacial surgery	Pain	Subjective	Secondary vs primary closure technique	7
Evidence that periodontal treatment improves diabetes outcomes	Engebretson 2013 [85]	Periodontics	Glycated haemoglobin HbA1c	Objective	Scaling and root planing vs. non-treatment	5
Primary prevention of dental erosion by calcium and fluoride	Zini 2014 [94]	Dental public health	Dental erosion prevention	Objective	Calcium vs. water	6

Table 6.2. Methods of sequence generation in the randomized trials

Sequence Generation Methods	No. (%) of 540 trials
Computer Software	86 (15.9)
Coin Tossing	41 (7.6)
Random Number table	29 (5.4)
Shuffling cards or envelopes	5 (0.9)
Drawing of lots	5 (0.9)
Throwing a dice	3 (0.6)
Lottery	3 (0.6)
Day of admission	2 (0.4)
Other	5 (0.9)
Unclear/Not reported	361 (66.9)
Total	540 (100)

Table 6.3. Reporting and timing of sequence generation in the randomized trials, N (%)

Reporting of randomization	**No**	**Yes**
In title	461 (85.4)	79 (14.6)
In abstract	211 (39.1)	329 (60.9)
In method	30 (5.6)	510 (94.4)
Randomization timing		
After consent	171 (31.7)	
After baseline measures	63 (11.7)	
Before baseline measures	3 (0.6)	
Before consent	2 (0.4)	
Unclear or not reported	301 (55.7)	

Table 6.4. Methods of allocation concealment in the randomized trials

Allocation Concealment Methods	No. (%) of 540 trials
Third party only	18 (3.3)
Pharmacy controlled	13 (2.4)
Sequentially numbered, sealed and opaque envelopes	12 (2.2)
Sealed (or opaque) envelopes only	9 (1.7)
Sequentially numbered sealed, opaque envelopes opened by third	6 (1.1)
Central allocation using telephone	5 (0.9)
Sealed envelopes open by third party	5 (0.9)
Envelops only	5 (0.9)
Open random allocation schedule (list of random numbers)	4 (0.7)
Other	4 (0.7)
Not reported	459 (85)
Total	540 (100)

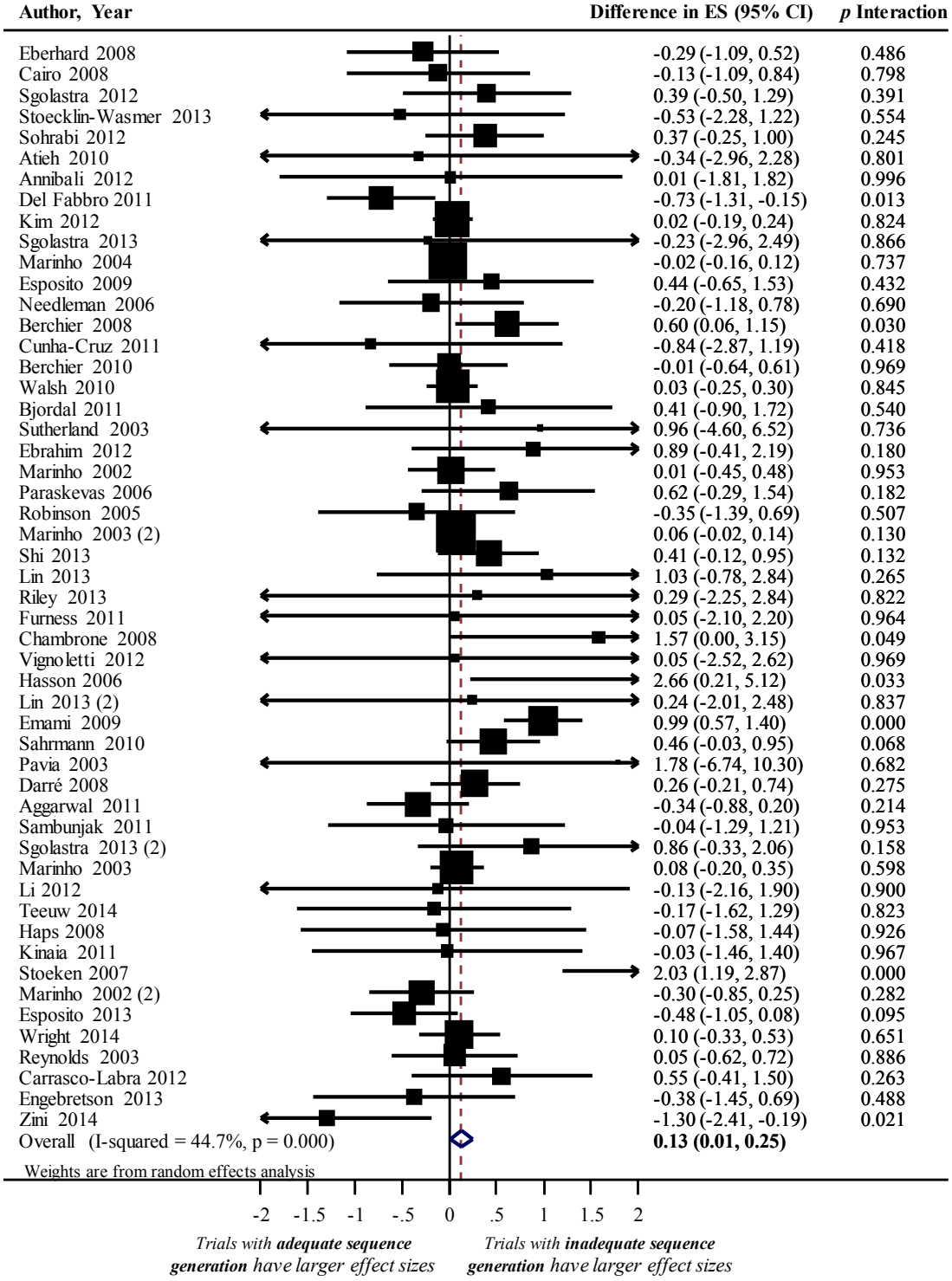

Figure 6.1a. Difference in treatment effect size (ES) estimate between trials with adequate and inadequate sequence generation. A positive value (more than zero) across meta-analyses indicates that trials with inadequate sequence generation exaggerate the treatment effect sizes when compared with trials with adequate sequence generation.

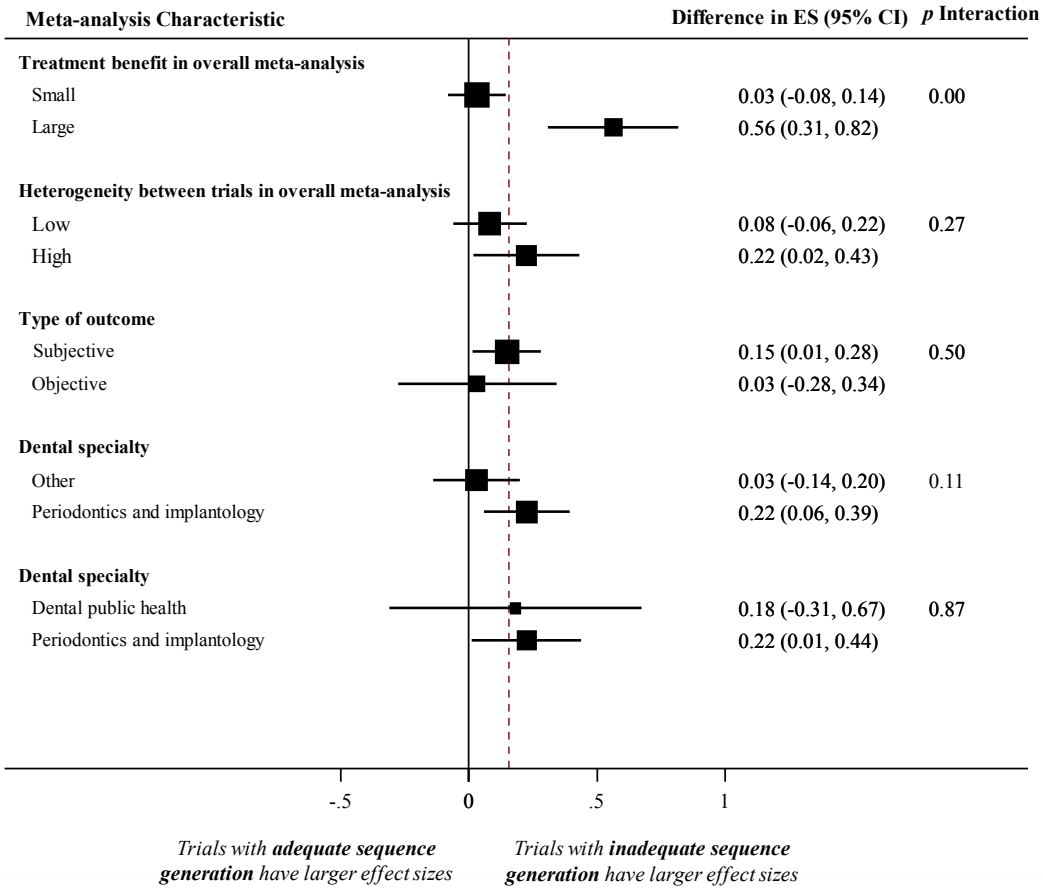

Figure 6.1b. Forest plot of the difference in treatment effect size (ES) estimate between trials with adequate and inadequate sequence generation stratified by meta-analyses characteristics.

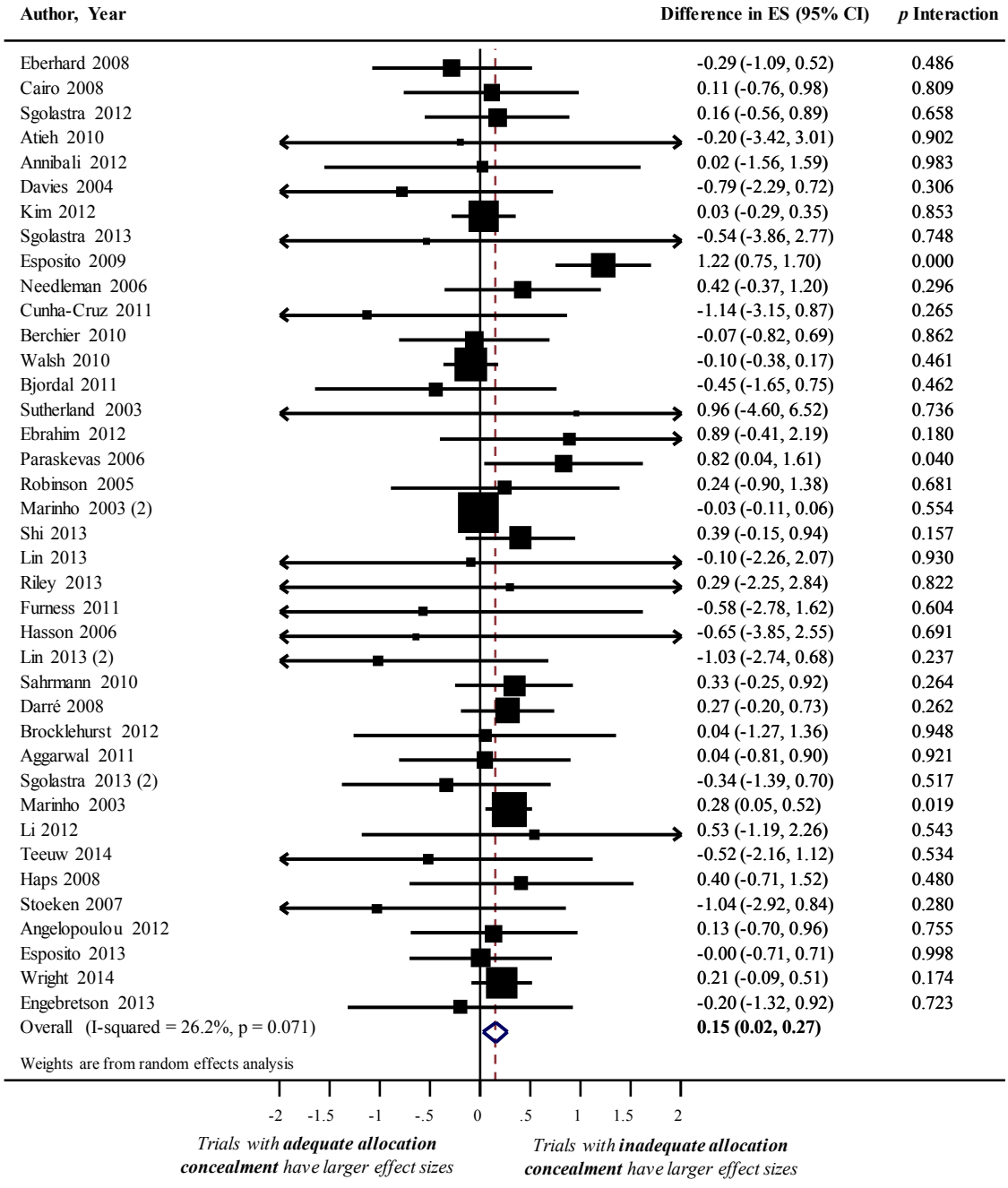

Figure 6.2a. Difference in treatment effect size (ES) estimate between trials with adequate and inadequate allocation concealment. A positive value (more than zero) across meta-analyses indicates that trials with inadequate allocation concealment exaggerate the treatment effect sizes when compared with trials with adequate allocation concealment.

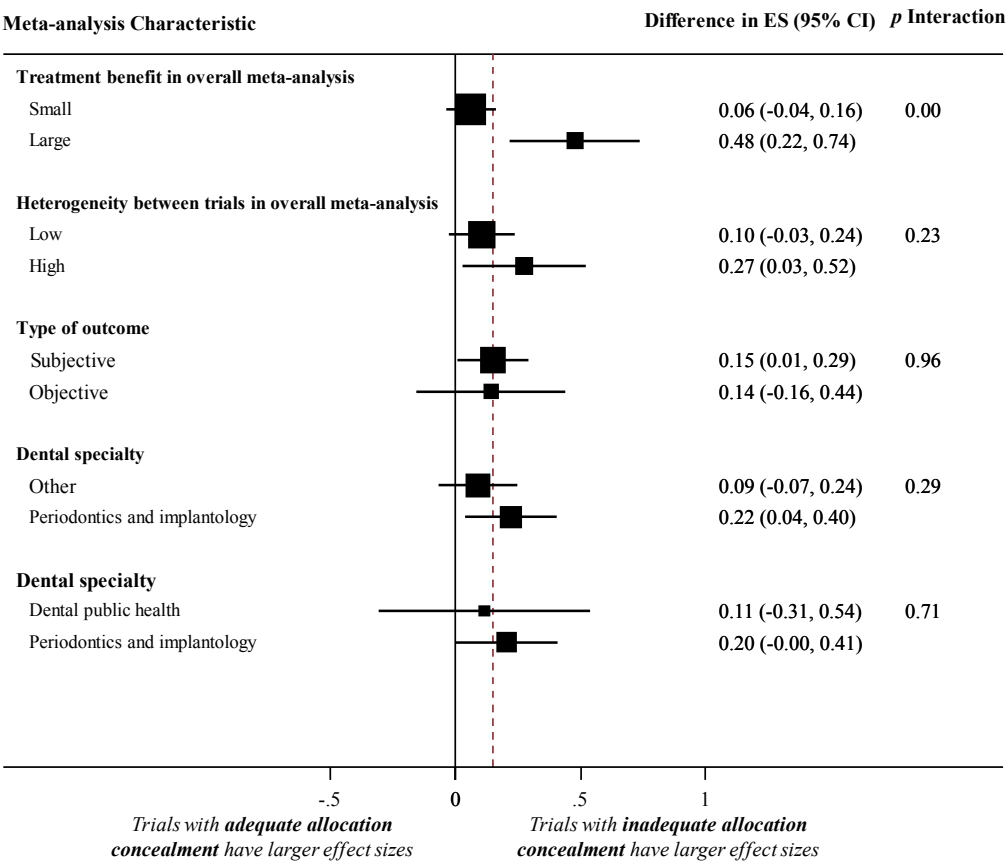

Figure 6.2b. Forest plot of the difference in treatment effect size (ES) estimate between trials with adequate and inadequate allocation concealment stratified by meta-analyses characteristics.

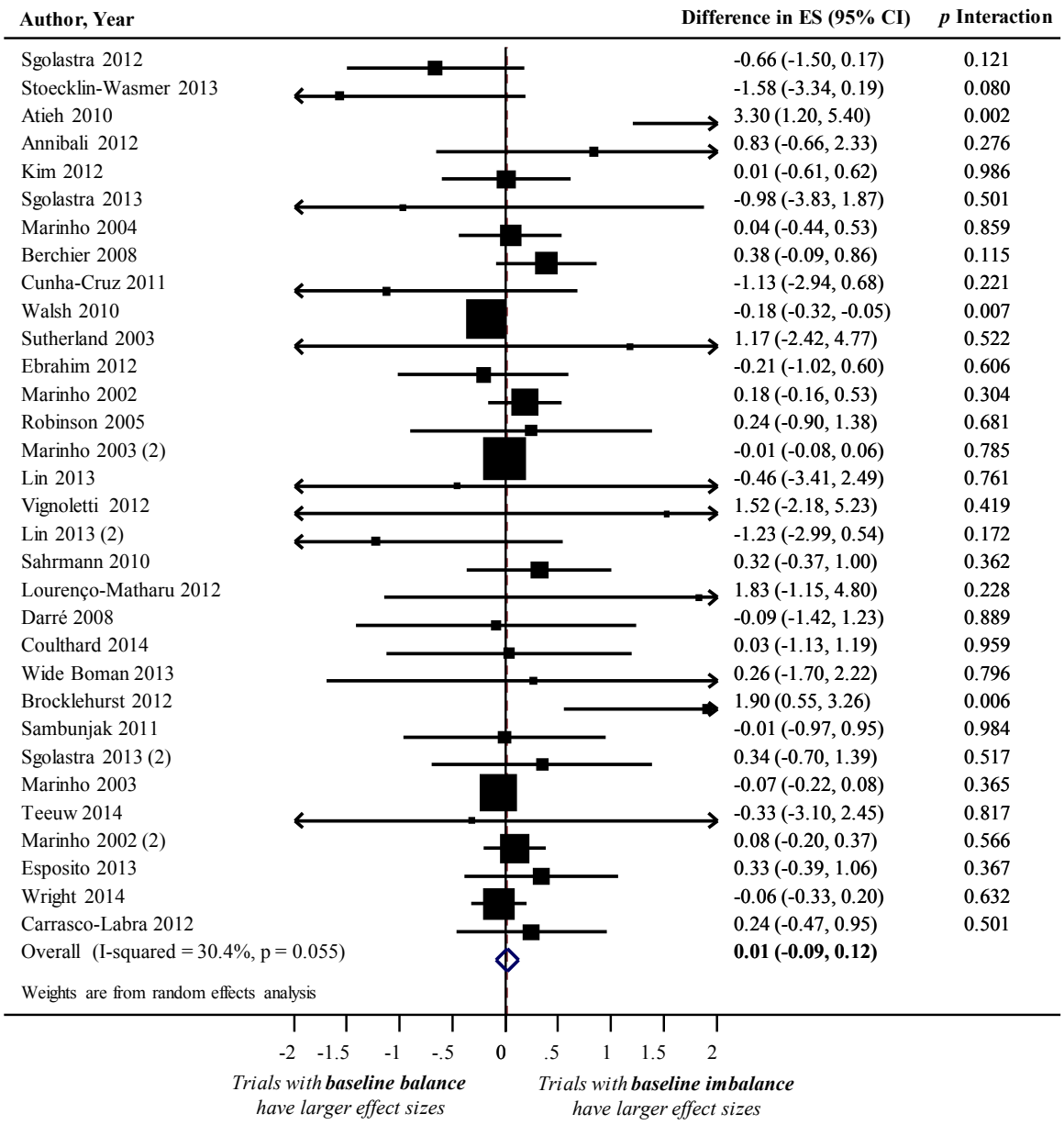

Figure 6.3a. Difference in treatment effect size (ES) estimate between trials with imbalance and balance of baseline characteristics. A positive value (more than zero) across meta-analyses indicates that trials with baseline imbalance exaggerate the treatment effect sizes when compared with trials with balance of baseline characteristics.

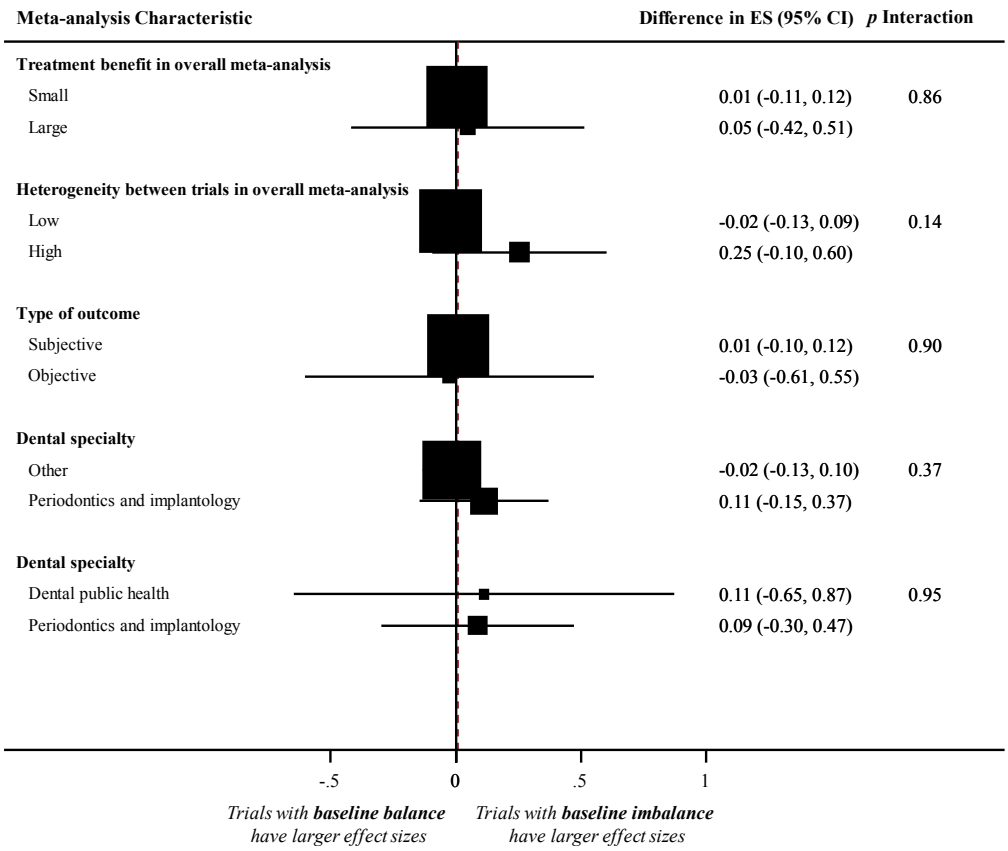

Figure 6.3b. Forest plot of the difference in treatment effect size (ES) estimate between trials with imbalance and balance of baseline characteristics stratified by meta-analyses characteristics.

References

[1] Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001; 134:663-94.

[2] Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998; 352:609-13.

[3] Needleman I, Moher D, Altman DG, Schulz KF, Moles DR, Worthington H. Improving the clarity and transparency of reporting health research: a shared obligation and responsibility. J Dent Res 2008; 87:894-5.

[4] Needleman I, Worthington H, Moher D, Schulz K, Altman DG. Improving the completeness and transparency of reports of randomized trials in oral health: the CONSORT statement. Am J Dent 2008; 21:7-12.

[5] Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010; 63:e1-37.

[6] Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. Health Technol Assess 1999; 3:i-iv, 1-98.

[7] Kahan BC, Rehal S, Cro S. Risk of selection bias in randomised trials. Trials 2015; 16:405.

[8] Higgins JP, Green S, eds. Cochrane Handbook For Systematic Reviews of Interventions version 5.0.0. Cochrane Collaboration 2008.

[9] Knottnerus JA, Tugwell P. Selection-related bias, an ongoing concern in doing and

publishing research. J Clin Epidemiol 2014; 67:1057-8.

[10] Berger VW, Exner DV. Detecting selection bias in randomized clinical trials. Contr Clin Trials 1999; 20:319-27.

[11] Schulz KF. Randomized controlled trials. Clin Obstet Gynecol 1998; 41:245-56.

[12] Pandis N, Polychronopoulou A, Eliades T. Randomization in clinical trials in orthodontics: its significance in research design and methods to achieve it. Eur J Orthod 2011; 33:684-90.

[13] Pandis N. Randomization. Part 3: allocation concealment and randomization implementation. Am J Orthod Dentofacial Orthop 2012; 141:126-8.

[14] Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. Lancet 2002; 359:515-9.

[15] Pandis N. Randomization. Part 1: sequence generation. Am J Orthod Dentofacial Orthop 2011; 140:747-8.

[16] Schulz KF, Altman DG, Moher D. Allocation concealment in clinical trials. JAMA 2002; 288:2406-7; author reply 8-9.

[17] Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. Lancet 2002; 359:614-8.

[18] Doig GS, Simpson F. Randomization and allocation concealment: a practical guide for researchers. J Crit Care 2005; 20:187-91.

[19] Altman DG, Dore CJ. Baseline comparisons in clinical trials. Lancet 1990; 335:1476.

[20] de Boer MR, Waterlander WE, Kuijper LD, Steenhuis IH, Twisk JW. Testing for baseline differences in randomized controlled trials: an unhealthy research behavior

that is hard to eradicate. Int J Behav Nutr Phys Act 2015; 12:4.

[21] Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2001; 135:982-9.

[22] Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gotzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int J Epidemiol 2007; 36:847-57.

[23] Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999; 282:1054-60.

[24] Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 2002; 287:2973-82.

[25] van Tulder MW, Suttorp M, Morton S, Bouter LM, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine 2009; 34:1685-92.

[26] Cook DA, West CP. Conducting systematic reviews in medical education: A stepwise approach. Med Educ 2012; 46:943-52.

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

[28] Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998; 352:609-13.

[29] Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gotzsche PC.

Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int J Epidemiol 2007; 36:847-57.

[30] Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med 2012; 157:429-38.

[31] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273:408-12.

[32] 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; 336:601-5.

[33] Sterne JAC, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med 2002; 21:1513-24.

[34] Nuesch E, Reichenbach S, Trelle S, Rutjes AWS, Liewald K, Sterchi R, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: A meta-epidemiologic study. Arthritis Rheum 2009; 61:1633-41.

[35] Van Tulder MW, Suttorp M, Morton S, Bouter LM, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine 2009; 34:1685-92.

[36] Berkman ND, Santaguida PL, Viswanathan M, Morton SC. The empirical evidence of bias in trials measuring treatment differences. Rockville (MD), Agency for Healthcare Research and Quality (US); 2014 Sep. Report No.:14-EHC050-EF.

[37] Faggion CM, Jr., Wu YC, Scheidgen M, Tu YK. Effect of risk of bias on the

effect size of meta-analytic estimates in randomized controlled trials in periodontology and implant dentistry. PloS One 2015; 10:e0139030.

[38] Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, Flores-Mir C. Methodological characteristics and treatment effect sizes in oral health randomised controlled trials: Is there a relationship? Protocol for a meta-epidemiological study. BMJ Open 2014; 4:e004527.

[39] American Dental Association. Specialty Definitions. http://www.ada.org/495.aspx (accessed March 2016).

[40] Armijo-Olivo S, Ospina M, da Costa BR, Egger M, Saltaji H, Fuentes CJ, et al. Poor reliability between cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. PloS One 2014; 9(5): e96920.

[41] Armijo-Olivo S, Saltaji H, da Costa BR, Fuentes J, Ha C, Cummings GG. What is the influence of randomisation sequence generation and allocation concealment on treatment effects of physical therapy trials? A meta-epidemiological study. BMJ Open 2015; 5:e008562.

[42] 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; 336:601-5.

[43] Hempel S, Suttorp M, Miles J. Empirical evidence of associations between trial quality and effect size. Rockville (MD), Agency for Healthcare Research and Quality (US); 2011 Jun. Report No.:11-EHC045-EF.

[44] Fenwick J, Needleman IG, Moles DR. The effect of bias on the magnitude of clinical outcomes in periodontology: a pilot study. J Clin Periodontol 2008; 35:775-82.

[45] Hartling L, Hamm MP, Fernandes RM, Dryden DM, Vandermeer B. Quantifying bias in randomized controlled trials in child health: a meta-epidemiological study. PloS One 2014; 9:e88008.

[46] Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med 2002; 21:1513-24.

[47] DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: An update. Contemp Clin Trials 2007; 28:105-14.

[48] Cohen J. Chapter 1: The Concepts of Power Analysis. In: Cohen J, editor. Statistical Power Analysis for the Behavioral Sciences. Second ed. Hillsdale, New Jersey: Academic Press, Inc; 1988. p. 1-17.

[49] Saltaji H, Cummings GG, Armijo-Olivo S, Major MP, Amin M, Major PW, et al. A descriptive analysis of oral health systematic reviews published 1991-2012: cross sectional study. PloS One 2013 8:e74545.

[50] Eberhard J, Jepsen S, Jervoe-Storm PM, Needleman I, Worthington HV. Full-mouth disinfection for the treatment of adult chronic periodontitis. Cochrane Database Syst Rev 2008: CD004622.

[51] Cairo F, Pagliaro U, Nieri M. Treatment of gingival recession with coronally advanced flap procedures: a systematic review. J Clin Periodontol 2008; 35:136-62.

[52] Sgolastra F, Petrucci A, Gatto R, Monaco A. Effectiveness of systemic amoxicillin/metronidazole as an adjunctive therapy to full-mouth scaling and root planing in the treatment of aggressive periodontitis: a systematic review and meta-analysis. J Periodontol 2012; 83:731-43.

[53] Stoecklin-Wasmer C, Rutjes AW, da Costa BR, Salvi GE, Juni P, Sculean A.

Absorbable collagen membranes for periodontal regeneration: a systematic review. J Dent Res 2013; 92:773-81.

[54] 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 Contr Clin Trials. J Periodontol 2012; 83:453-64.

[55] Atieh MA, Ibrahim HM, Atieh AH. Platform switching for marginal bone preservation around dental implants: a systematic review and meta-analysis. J Periodontol 2010; 81:1350-66.

[56] 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; 39:1097-113.

[57] Del Fabbro M, Bortolin M, Taschieri S, Weinstein R. Is platelet concentrate advantageous for the surgical treatment of periodontal diseases? A systematic review and meta-analysis. J Periodontol 2011; 82:1100-11.

[58] Davies RM, Ellwood RP, Davies GM. The effectiveness of a toothpaste containing triclosan and polyvinyl-methyl ether maleic acid copolymer in improving plaque control and gingival health: a systematic review. J Clin Periodontol 2004; 31:1029-33.

[59] Kim AJ, Lo AJ, Pullin DA, Thornton-Johnson DS, Karimbux NY. Scaling and root planing treatment for periodontitis to reduce preterm birth and low birth weight: a systematic review and meta-analysis of randomized controlled trials. J Periodontol 2012; 83:1508-19.

[60] Needleman IG, Worthington HV, Giedrys-Leeper E, Tucker RJ. Guided tissue regeneration for periodontal infra-bony defects. Cochrane Database Syst Rev 2006:

CD001724.

[61] 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.

[62] Berchier CE, Slot DE, Haps S, Van der Weijden GA. The efficacy of dental floss in addition to a toothbrush on plaque and parameters of gingival inflammation: a systematic review. Int J Dent Hyg 2008; 6:265-79.

[63] Berchier CE, Slot DE, Van der Weijden GA. The efficacy of 0.12% chlorhexidine mouthrinse compared with 0.2% on plaque accumulation and periodontal parameters: a systematic review. J Clin Periodontol 2010; 37:829-39.

[64] Sutherland S, Matthews DC. Emergency management of acute apical periodontitis in the permanent dentition: a systematic review of the literature. J Can Dent Assoc 2003; 69:160.

[65] Paraskevas S, van der Weijden GA. A review of the effects of stannous fluoride on gingivitis. J Clin Periodontol 2006; 33:1-13.

[66] Riley P, Lamont T. Triclosan/copolymer containing toothpastes for oral health. Cochrane Database Syst Rev 2013; 12:CD010514.

[67] Murphy KG, Gunsolley JC. Guided tissue regeneration for the treatment of periodontal intrabony and furcation defects. A systematic review. Ann Periodontol 2003; 8:266-302.

[68] Chambrone L, Chambrone D, Pustiglioni FE, Chambrone LA, Lima LA. Can subepithelial connective tissue grafts be considered the gold standard procedure in the treatment of Miller Class I and II recession-type defects? J Dent 2008; 36:659-71.

[69] Lin GH, Chan HL, Bashutski JD, Oh TJ, Wang HL. The effect of flapless surgery

on implant survival and marginal bone level: a systematic review and meta-analysis. J Periodontol 2014; 85:e91-103.

[70] Emami E, Heydecke G, Rompre 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; 20:533-44.

[71] Sahrmann P, Puhan MA, Attin T, Schmidlin PR. Systematic review on the effect of rinsing with povidone-iodine during nonsurgical periodontal therapy. J Periodontal Res 2010; 45:153-64.

[72] Pavia M, Nobile CG, Angelillo IF. Meta-analysis of local tetracycline in treating chronic periodontitis. J Periodontol 2003; 74:916-32.

[73] Darre L, Vergnes JN, Gourdy P, Sixou M. Efficacy of periodontal treatment on glycaemic control in diabetic patients: A meta-analysis of interventional studies. Diabetes Metab 2008; 34:497-506.

[74] Esposito M, Murray-Curtis L, Grusovin MG, Coulthard P, Worthington HV. Interventions for replacing missing teeth: different types of dental implants. Cochrane Database Syst Rev 2007: CD003815.

[75] 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.

[76] 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; 40:514-26.

[77] Li W, Xiao L, Hu J. The use of enamel matrix derivative alone versus in

combination with bone grafts to treat patients with periodontal intrabony defects: a meta-analysis. J Am Dent Assoc 2012; 143:e46-56.

[78] Teeuw WJ, Slot DE, Susanto H, Gerdes VE, Abbas F, D'Aiuto F, et al. Treatment of periodontitis improves the atherosclerotic profile: a systematic review and meta-analysis. J Clin Periodontol 2014; 41:70-9.

[79] Haps S, Slot DE, Berchier CE, Van der Weijden GA. The effect of cetylpyridinium chloride-containing mouth rinses as adjuncts to toothbrushing on plaque and parameters of gingival inflammation: a systematic review. Int J Dent Hyg 2008; 6:290-303.

[80] Deacon SA, Glenny AM, 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.

[81] Kinaia BM, Steiger J, Neely AL, Shah M, Bhola M. Treatment of Class II molar furcation involvement: meta-analyses of reentry results. J Periodontol 2011; 82:413-28.

[82] Stoeken JE, Paraskevas S, van der Weijden GA. The long-term effect of a mouthrinse containing essential oils on dental plaque and gingivitis: a systematic review. J Periodontol 2007; 78:1218-28.

[83] Esposito M, Grusovin MG, Maghaireh H, Worthington HV. Interventions for replacing missing teeth: different times for loading dental implants. Cochrane Database Syst Rev 2013; 3:CD003878.

[84] Reynolds MA, Aichelmann-Reidy ME, Branch-Mays GL, Gunsolley JC. The efficacy of bone replacement grafts in the treatment of periodontal osseous defects. A systematic review. Ann Periodontol 2003; 8:227-65.

[85] Engebretson S, Kocher T. Evidence that periodontal treatment improves diabetes outcomes: a systematic review and meta-analysis. J Periodontol 2013; 84:S153-69.

[86] Marinho VC, Higgins JP, Sheiham A, Logan S. Combinations of topical fluoride (toothpastes, mouthrinses, gels, varnishes) versus single topical fluoride for preventing dental caries in children and adolescents. Cochrane Database Syst Rev 2004: CD002781.

[87] Walsh T, Worthington HV, Glenny AM, 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.

[88] Marinho VC, Higgins JP, Logan S, Sheiham A. Fluoride gels for preventing dental caries in children and adolescents. Cochrane Database Syst Rev 2002: CD002280.

[89] Marinho VC, Higgins JP, Sheiham A, Logan S. Fluoride toothpastes for preventing dental caries in children and adolescents. Cochrane Database Syst Rev 2003: CD002278.

[90] Robinson PG, Deacon SA, Deery C, Heanue M, Walmsley AD, Worthington HV, et al. Manual versus powered toothbrushing for oral health. Cochrane Database Syst Rev 2005: CD002281.

[91] Lourenco-Matharu L, Ashley PF, Furness S. Sedation of children undergoing dental treatment. Cochrane Database Syst Rev 2012; 3:CD003877.

[92] Marinho VC, Higgins JP, Logan S, Sheiham A. Fluoride mouthrinses for preventing dental caries in children and adolescents. Cochrane Database Syst Rev 2003: CD002284.

[93] Wright JT, Hanson N, Ristic H, Whall CW, Estrich CG, Zentz RR. Fluoride

toothpaste efficacy and safety in children younger than 6 years: a systematic review. J Am Dent Assoc 2014; 145:182-9.

[94] Zini A, Krivoroutski Y, Vered Y. Primary prevention of dental erosion by calcium and fluoride: a systematic review. Int J Dent Hyg 2014; 12:17-24.

[95] Marinho VC, Higgins JP, Logan S, Sheiham A. Fluoride varnishes for preventing dental caries in children and adolescents. Cochrane Database Syst Rev 2002: CD002279.

[96] Sgolastra F, Petrucci A, Severino M, Gatto R, Monaco A. Lasers for the treatment of dentin hypersensitivity: a meta-analysis. J Dent Res 2013; 92:492-9.

[97] Bjordal JM, Bensadoun RJ, Tuner J, Frigo L, Gjerde K, Lopes-Martins RA. A systematic review with meta-analysis of the effect of low-level laser therapy (LLLT) in cancer therapy-induced oral mucositis. Support Care Cancer 2011; 19:1069-77.

[98] Ebrahim S, Montoya L, Busse JW, Carrasco-Labra A, Guyatt GH. The effectiveness of splint therapy in patients with temporomandibular disorders: a systematic review and meta-analysis. J Am Dent Assoc 2012; 143:847-57.

[99] 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.

[100] Lin PY, Cheng YW, Chu CY, Chien KL, Lin CP, Tu YK. In-office treatment for dentin hypersensitivity: a systematic review and network meta-analysis. J Clin Periodontol 2013; 40:53-64.

[101] 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.

[102] Poulsen S, Errboe M, Lescay Mevil Y, Glenny AM. Potassium containing toothpastes for dentine hypersensitivity. Cochrane Database Syst Rev 2006: CD001476.

[103] Wide Boman U, Carlsson V, Westin M, Hakeberg M. Psychological treatment of dental anxiety among adults: a systematic review. Eur J Oral Sci 2013; 121:225-34.

[104] Brocklehurst P, Tickle M, Glenny AM, Lewis MA, Pemberton MN, Taylor J, et al. Systemic interventions for recurrent aphthous stomatitis (mouth ulcers). Cochrane Database Syst Rev 2012; 9:CD005411.

[105] 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.

[106] Cunha-Cruz J, Stout JR, Heaton LJ, Wataha JC. Dentin hypersensitivity and oxalates: a systematic review. J Dent Res 2011; 90:304-10.

[107] Vignoletti F, Matesanz P, Rodrigo D, Figuero E, Martin C, Sanz M. Surgical protocols for ridge preservation after tooth extraction. A systematic review. Clin Oral Implants Res 2012; 23 Suppl 5:22-38.

[108] Markiewicz MR, Brady MF, Ding EL, Dodson TB. Corticosteroids reduce postoperative morbidity after third molar surgery: a systematic review and meta-analysis. J Oral Maxillofac Surg 2008; 66:1881-94.

[109] 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:e441-57.

[110] Coulthard P, Bailey E, Esposito M, Furness S, Renton TF, Worthington HV. Surgical techniques for the removal of mandibular wisdom teeth. Cochrane Database Syst Rev 2014; 7:CD004345.

[111] Hoekema A, Stegenga B, De Bont LG. Efficacy and co-morbidity of oral appliances in the treatment of obstructive sleep apnea-hypopnea: a systematic review. Crit Rev Oral Biol Med 2004; 15:137-55.

[112] Angelopoulou MV, Vlachou V, Halazonetis DJ. Pharmacological management of pain during orthodontic treatment: a meta-analysis. Orthod Craniofac Res 2012; 15:71-83.

[113] Hasson H, Ismail AI, Neiva G. Home-based chemically-induced whitening of teeth in adults. Cochrane Database Syst Rev 2006: CD006202.

[114] Pandis N. The effect size. Am J Orthod Dentofacial Orthop 2012; 142:739-40.

[115] Linde K, Scholz M, Ramirez G, Clausius N, Melchart D, Jonas WB. Impact of study quality on outcome in placebo-controlled trials of homeopathy. J Clin Epidemiol 1999; 52:631-6.

[116] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273:408-12.

[117] Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Contr Clin Trials 1996; 17:1-12.

[118] Mhaskar R, Djulbegovic B, Magazin A, Soares HP, Kumar A. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. J Clin Epidemiol 2012; 65:602-9.

[119] Armijo-Olivo S, Fuentes J, da Costa BR, Saltaji H, Ha C, Cummings GG. Blinding in physical therapy trials and its association with treatment effects: A meta-epidemiological study. Am J Phys Med Rehabil 2016 May 4.

[120] Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. Stat Med 2007; 26:2745-58.

[121] Elkins MR. Assessing baseline comparability in randomised trials. J Physiother 2015; 61:228-30.

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

[123] Hempel S, Miles J, Suttorp MJ, Wang Z, Johnsen B, Morton S, et al. Detection of associations between trial quality and effect sizes. Rockville (MD), Agency for Healthcare Research and Quality (US); 2012 Jan. Report No.:12-EHC010-EF.

Chapter 7

Influence of blinding on treatment effect size estimate among randomized trials in dentistry

7.1. Background

As evidence-based practice has grown over the past two decades, there has been a consistent generation of new randomized trials and systematic reviews in medicine and dentistry. Currently, thousands of trials and meta-analyses of these trials are published every year to guide healthcare professionals in their evidence-based decisions in clinical practice. In the field of dentistry alone, nearly 50 new clinical trials and 20 systematic reviews are published every month [1-3]. These trials and systematic reviews, in turn, support much of the treatment modalities and treatment recommendations in dental practice based on the current best-identified evidence. While randomized controlled trials (RCTs), the building blocks of systematic reviews and meta-analyses, are considered to provide reliable evidence for dental decision making, RCTs are susceptible to bias (underestimation or overestimation of treatment effect size) due to limitations in their design, conduct, and reporting [4, 5]. For results and outcomes of RCTs to be generalizable and valid to specific patient subsets, they need to be properly designed, carefully conducted, and accurately reported to a standard that warrants the implementation of their results [4, 6].

Blinding (or "masking") has been recognized as an important criterion for reaching a high methodological quality, particularly with respect to internal validity of randomized trials [7]. Blinding is broadly used in a trial to prevent performance bias (blinding of participants and care providers) and detection bias (blinding of assessors) [8-10]. Blinding can be applied at numerous levels of a trial, including participants, outcome assessors, care providers, data analysts, or other personnel. Thus, several terms (e.g., single-, double-, or triple-blind) have been used to describe blinding types

[6, 11, 12]. However, the use of these terms has been inconsistent among research groups, and this contributes to conceptual and operational ambiguity. While appropriate blinding can reduce performance and detection biases, it is not always feasible to apply blinding in a trial, particularly in a randomized trial that involves surgical or device interventions such as oral surgery and orthodontics, as participants are often aware of the type of intervention they are receiving. The appropriateness of blinding depends on factors such as the type of outcome examined (e.g., objective vs. subjective) [10] and the type of intervention applied (e.g., surgical vs. drug). For example, it is more difficult to implement blinding in RCTs of surgical interventions than to implement blinding in RCTs of drug interventions in which trial investigators can use placebo medications to attain adequate blinding [13].

Published meta-epidemiological studies focused on the blinding domain have found that there are potential associations between treatment effect size estimates and blinding of participants [14-19], care providers [15-17, 19], assessors [15, 17-21], and "double blinding" [22, 23]. While those meta-epidemiological investigations were conducted within numerous health fields, the value of the conclusions may be limited, based on numerous factors, when generalized to other healthcare fields. These factors include a failure to evaluate continuous outcomes because of a preference for assessing dichotomous outcomes [15, 21, 23] (which was very common in these investigations), the emergence of inconsistent methodological findings associated with treatment effect size estimates [15, 17, 22], and the trial being "underpowered" [24] by lacking adequate sample size, which is needed to properly quantify bias in randomized trails. More notably, the meta-epidemiological studies reported that the extent of bias in the treatment effect size estimate associated with blinding varied across different medical fields as well as across different types of intervention [17, 24].

To date, no meta-epidemiological study has examined the bias related to blinding in randomized trials within any oral health subspecialties or scope of practice

in dentistry. Therefore, it is unclear whether the previously mentioned conclusions hold true in the field of oral health research where blinding is sometimes difficult or not feasible, especially in oral health trials involving surgical or device interventions, such as orthodontic trials.

Thus, our specific research questions were: (1) Do oral health randomized trials with adequate blinding of participants, outcome assessors, and health care providers yield different treatment effect size estimates than trials with lack of blinding? (2) Do specific nonmethodological meta-analysis characteristics (e.g., dental specialty, type of treatment, type of outcome [objective vs. subjective], magnitude of the treatment effect size estimate, heterogeneity of meta-analysis) modify the association between blinding and treatment effect size estimate? The findings generated from this work will inform current initiatives for developing and disseminating future research frameworks for the conduct and reporting of oral health randomized trials.

7.2. Methods and Analysis

This study is part of a large meta-epidemiological study that investigates the association between methodological characteristics and treatment effect size estimates in oral health RCTs. The protocol for this meta-epidemiological study was registered on PROSPERO (CRD42014014070), and published *a priori* [25].

7.2.1. Literature search

We conducted a comprehensive search of the literature deposited in six electronic databases (PubMed, MEDLINE, EMBASE, ISI Web of Science, Evidence-Based Medicine Reviews–Cochrane Database of Systematic Reviews, and Health STAR) from database inception to May 2014. The search strategy was planned with the assistance of a health sciences information specialist and included a combination of index terms and keywords related to systematic reviews and oral health. The search

strategy for each database can be found in **Appendix A.3**. We also searched the American Dental Association (ADA)–Evidence-based Dentistry database [26] and hand-searched the reference lists of potentially relevant studies identified in the main search, which focused on the quality of systematic reviews in oral health. We did not restrict the searches to English language nor did we limit them by other means.

7.2.2. Inclusion and exclusion criteria

Two independent reviewers (H.S., M.A.) with dental research and clinical backgrounds screened the titles and abstracts retrieved by the search strategy. Only abstracts deemed to potentially fulfill the inclusion criteria were selected. The full text of relevant reports that lacked sufficient information in the abstract were also retrieved before a final decision was made. The same assessors independently determined the final eligibility of full texts; discrepancies were resolved through consensus.

We included meta-analyses if they met the following predefined eligibility criteria: (1) the meta-analyses was in the field of oral health research and examined a therapeutic intervention related to treatment, prevention, or rehabilitation of dental, oral, or craniofacial diseases [27, 28]; and (2) the meta-analyses examined at least one continuous outcome and included a minimum of five randomized trials with quantitative data of treatment effect size estimates. We subsequently selected RCTs included in the selected meta-analyses that met the following predefined eligibility criteria: (1) the design was reported to be an RCT [33] where findings were reported in a way that allowed for calculation of treatment effect size estimate; (2) the comparison was between an intervention versus a placebo, there was no treatment control, or standard care (trials with a comparison of one active intervention versus another active intervention were excluded); and (3) the trials examined a therapeutic intervention related to a dental specialty recognized by the American Dental Association (ADA) [28].

7.2.3. Data extraction

A panel of five reviewers from diverse health research areas (H.S., C.H., J.S., J.F., S.A-O.) carried out the data extraction. To ensure consistency during the data extraction phase, one of the team members performed the reviewer training process where the review panel evaluated 10 randomized trials not included in the final set of trials and then discussed these to achieve consistency. A similar reviewer training process was conducted in other studies performed by the same research team [29, 30]. Once agreement was achieved regarding data extraction, protocol, and interpretation, the data extraction phase was completed. Data extraction was performed in duplicate, that is, two assessors independently carried out data extraction, with consensus meetings employed to resolve any disagreement. One assessor with an oral health research background (H.S.) performed complete data extraction (n = 540, 100%) while another assessor (either C.H., J.S., or J.F.) with a medical (nonoral health) research background acted as a second assessor. The two assessors conferred with a third assessor (S.A-O.) if an agreement could not be reached, to achieve complete consensus. Only consensus data were used for statistical analyses. A structured and pilot-tested data extraction template designed in a Microsoft Office Access database was used for data extraction.

The primary outcome reported for each review was identified as the primary outcome for our analysis. Alternatively, the primary outcome for the analysis was determined as the outcome associated with the meta-analysis that involved the largest number of trials (in case the review's primary outcome was binary, not clearly stated, or the quantitative analysis associated with the outcome included less than five trials). Details from each of the included randomized trials and meta-analyses were extracted for the primary outcome of the review; the following elements were extracted: means, standard deviations, sample sizes, publication year, dental specialty (e.g., dental public health, endodontics, periodontics, oral medicine and oral pathology, oral and maxillofacial surgery, prosthodontics and restorative dentistry, orthodontics and dentofacial orthopaedics, and pediatric dentistry), primary outcome assessed, type of

comparison in a review, number of included trials in a review, trial design (e.g., parallel, split-mouth, crossover, and factorial), type of outcome in a trial (e.g., drug vs. nondrug or subjective vs. objective [23]), and number of centers in a trial (e.g., multicenter vs. single center). To classify the type of comparison, the classification of the comparison implemented in the quantitative analysis reported in the review (e.g., treatment vs. control) was used.

To assess the risk of bias associated with blinding in the selected randomized trials, we applied nine blinding-based criteria (see **Table 7.1**), namely: patient blinding (blinding of participants allocated to interventions), assessor blinding (blinding of data collectors), care-provider blinding (blinding of dental clinicians and/or therapists who provided the interventions), investigator blinding (blinding of the principal investigator), statistician blinding (blinding of the data analyst), double blinding (we considered a study as double blinded when blinding of both patients and assessors was judged as having a low risk of bias), study described as "double blind" (by trial investigators), triple blinding (we considered a study as triple blinded when blinding of patients, assessors, and care providers were achieved), and the propriety of blinding (blinding that was properly implemented within the trial's components according to the primary outcome).

We scored each of these items following the definitions and methods for each of these criteria in the guidelines of the quality assessment tools that were found to be valid and most commonly used in health research [7, 31-38]. We established our evaluation based on the chosen primary outcome of analysis, and employed a 3-ordinal scoring scheme comprised of "high, unclear, low" risk of bias [39] for two of the domains (patient blinding and assessor blinding) and "yes, no, unclear" [7] for the other five domains. **Table 7.1** provides further details of the definitions of the blinding-based criteria used in the study.

Moreover, we assessed whether each individual component of a trial

(participants, assessors, care-providers, statisticians, or investigators) would be blinded to study measurements: random assignment, hypothesis, details of interventions, outcome measures, and outcome analysis.

7.2.4. Data analysis

To describe the blinding in the randomized trials selected, we conducted descriptive analyses including proportions and percentages of study elements. To examine whether dental randomized trials with adequate blinding reported different treatment effect size estimates than trials with lack of blinding, we conducted a two-level analysis using a meta-meta-analytic approach with a random-effects model following guidelines established by Sterne et al. [40]. This type of analysis is reported to be the most effective to address our research question, given that the methodological approach used for our meta-epidemiological analysis takes into account heterogeneity between randomized trials, within meta-analyses in a first step, and among meta-analyses in a second step [41, 42]. We obtained raw data for each trial from each meta-analysis and cross-checked the numbers with the data reported in the primary trial.

For the "within meta-analyses level" (first level analysis), we obtained a standardized treatment effect size estimate for the primary outcome of each randomized trial, as outlined by Cohen [43]. A negative treatment effect size estimate entailed a favourable effect of the tested intervention. We obtained data from each selected randomized trial and meta-analysis. We considered a trial if it was included in more than one meta-analysis, only once (from the meta-analysis with the fewer number of trials). We divided included trials, for each meta-analysis and each randomized trial component, into two groups according to the relevant quality criterion (e.g., participant blinding, assessor blinding, care-provider blinding, double blinding, triple blinding)—those that adequately addressed the criterion and those that did not ("no" or "unclear"). We calculated two treatment effect size estimates for each

meta-analysis: the first corresponded to the pooled treatment effect size estimate from trials including the characteristic of interest (e.g., patient blinding) and the second corresponded to the pooled treatment effect size estimate from trials where the characteristic of interest (e.g., no or unclear patient blinding) was not met. We conducted inverse-variance random-effects meta-analysis to derive pooled treatment effect size estimates for each meta-analysis, and calculated the DerSimonian and Laird estimates of variance to determine heterogeneity between randomized trials [41]. Thus, for each meta-analysis, we used meta-regression approaches to derive the difference between pooled estimates from trials with and without the characteristic of interest. A negative difference in treatment effect size estimate implied that trials with the blinding-based item yielded a more favourable treatment effect size estimate for the tested intervention.

For the "among meta-analyses level" (second level analysis), we pooled findings of the previous analysis (combined differences from all meta-analyses) to describe the effect of each trial's component across all meta-analyses. We combined treatment effect size estimates at this stage using inverse-variance random-effects meta-analysis [41] to account for between-meta analysis heterogeneity, and calculated the DerSimonian and Laird estimates of variance to determine heterogeneity between meta-analyses [41]. All p-values were two-sided.

To examine whether specific characteristics modify the associations between blinding and the treatment effect size estimate, we stratified the analyses with interaction tests based on Z scores according to the following factors: type of outcome (objective vs. subjective), dental speciality (periodontal vs. other interventions, or dental public health vs. other interventions), magnitude of the treatment effect size estimate (small, if > -0.5 vs. large, if ≤ -0.5), and heterogeneity of the meta-analysis (low if $\tau^2 < 0.06$ vs. high if $\tau^2 \geq 0.06$; the cut-off of $\tau^2 = 0.06$ roughly amounts to a difference between the largest and the smallest treatment effect size estimate, where the smallest treatment effect size estimate = 1). We performed all analyses using

STATA statistical software version 14 (College Station, TX: StataCorp LP). The analysis was conducted by the principal assessor who was trained, and supervised by a team member with vast experience in analyses of meta-epidemiological studies.

We calculated the sample size according to recommendations in Hempel et al. [44] and Berkman et al. [24], given that sample size calculation for these types of studies is not clearly established in the scientific literature. It was reported that previous meta-epidemiological investigations had inadequate sample sizes, and were therefore labeled "underpowered" [24]. From previous meta-epidemiological investigations [30, 45, 46], we anticipated obtaining a difference in treatment effect size estimate of at least 0.15 (SE = 0.087) between trials with and without quality criteria [45]. This magnitude of difference in treatment effect size estimate has been claimed to resemble nearly 1/4 to 1/2 of classic treatment effect size estimates for interventions in fields similar to the field of dentistry [45]. Accordingly, we planned a sample size of nearly 500 randomized trials included in 60 systematic reviews to demonstrate such a meaningful difference. This is approximately two to three times the number of trials included in previously published meta-epidemiological investigations [18, 45, 47].

7.3. Results

7.3.1. Characteristics of selected systematic reviews and included randomized trials

The updated database of dental, oral, and craniofacial systematic reviews [45] included 1408 records (published between 1991 and 2014) of which 152 systematic reviews with meta-analyses were judged to be potentially relevant; of these, 64 (32 Cochrane and 32 non-Cochrane reviews) satisfied the eligibility criteria for the present report. The complete list of excluded reviews is available upon request.

Overall, the chosen meta-analyses were published between 2002 and 2014 (median year of publication: 2010; interquartile range [IQR] 2006–2012), while the

median number of trials included in the meta-analyses was six (IQR 6–10). A total of 540 trials analyzing 137,957 patients were considered for this study. The meta-analyses examined a therapeutic intervention related to the fields of periodontics (36 reviews; 271 trials), dental public health and pediatric dentistry (10 reviews; 145 trials), oral medicine and pathology (11 reviews; 80 trials), oral and maxillofacial surgery (4 reviews; 26 trials), orthodontics and dentofacial orthopedics (2 reviews; 12 trials), and restorative dentistry (1 review; 6 trials). Approximately one-fifth of the trials, were multicenter trials, with nearly one-third of the trials placebo-controlled (n = 204; 37.8%), and two-thirds of the trials examined were nondrug (n = 359; 66.5%) or nonsurgical (n = 370; 68.5%) interventions. The majority of trials used parallel design (n = 372; 68.9%), and one-quarter used split-mouth design (n = 126; 23.3%). **Table 6.1** (chapter 6) contains a complete list and characteristics of the chosen meta-analyses.

7.3.2. Blinding in dental randomized trials

Blinding of patients was judged as adequate (low risk of bias) in 71.5% (n = 386) of the trials, and blinding of the outcome assessment was judged as adequate (low risk of bias) in 59.4% of the trials. Blinding of both patients and assessors was judged as adequate in 72.8% of the trials (n = 273), and 76.5% (n = 117) of the trials were assessed as adequate with respect to blinding of patients, assessors, and care-providers. Blinding of the assessor was reported in 59.4% of the trials (n = 321), while blinding of patients was unclear/not reported in nearly half of the trials (n = 279; 51.7%). Two-thirds of trials were not described as double-blind (n = 358; 66.3%). The method of blinding was appropriate in 53% of the trials (n = 286), while blinding of the principal investigator and statistician was unclear/not reported in the vast majority of trials. **Tables 7.2 and 7.3** provide details of the blinding of individual components (participants, assessors, principal investigators, care-providers, and statisticians), and the level of blinding (random assignment, hypothesis, details of intervention, and data analysis) in RCTs of oral health interventions.

7.3.3. Impact of patient blinding on treatment effect size estimate

Figure 7.1a displays a forest plot of the difference in treatment effect size estimates between trials with the presence and lack of patient blinding. Twenty-eight meta-analyses, including 275 trials that analyzed 109,753 patients, provided information for this meta-epidemiological analysis. Results of the analysis showed that trials with inadequate patient blinding had significantly larger treatment effect size estimates (difference in treatment effect size estimates = 0.12, 95% confidence interval 0.00 to 0.23, p = 0.046). This result implies that treatment effect size estimates were 0.12 larger in trials with lack of patient blinding. However, the impact of patient blinding on treatment effect size estimates stratified by other characteristics of meta-analyses (heterogeneity of meta-analysis, type of outcome, and dental speciality) was not statistically significant for any of the characteristics (see **Figure 7.1b)**.

7.3.4. Impact of assessor blinding on treatment effect size estimate

Figure 7.2a displays a forest plot of the difference in treatment effect size estimates between trials with a presence and a lack of assessor blinding. Forty-four meta-analyses, including 408 trials that analyzed 119,282 patients, provided information for this meta-epidemiological analysis. Although assessor blinding was not associated with a statistically significant difference in treatment effect size estimate, trials with lack of assessor blinding tended to inflate treatment effect size estimates when compared with trials with a presence of assessor blinding (difference in treatment effect size estimate = 0.06, 95% confidence interval -0.06 to 0.18, p = 0.316). A positive value (more than zero) across meta-analyses would indicate that a lack of assessor blinding inflated the treatment effect size estimate. The results of the stratified analyses show that none of the meta-analyses characteristics had a statistically significant interaction with the treatment effect size estimate (see **Figure 7.2b)**.

7.3.5. Impact of care-provider blinding on treatment effect size estimate

Figure 7.3a displays a forest plot of the difference in treatment effect size estimates between randomized trials with a presence and a lack of care-provider blinding. Eighteen meta-analyses, including 408 trials that analyzed 109,383 patients, provided information for this meta-epidemiological analysis. Care-provider blinding was not associated with a statistically significant difference in treatment effect size estimate (difference in treatment effect size estimate = 0.02, 95% confidence interval -0.04 to 0.09, p = 0.509). A positive value (more than zero) across meta-analyses would indicate that the lack of care-provider blinding inflated the treatment effect size estimate. The results of the stratified analyses show that none of the meta-analyses characteristics had a statistically significant interaction with the treatment effect size estimate (see **Figure 7.3b)**.

7.3.6. Impact of principal-investigator blinding on treatment effect size estimate

Figure 7.4a displays a forest plot of the difference in treatment effect size estimates between randomized trials with a presence and a lack of principal-investigator blinding. Eighteen meta-analyses, including 162 trials that analyzed 59,757 patients, provided information for this meta-epidemiological analysis. Principal-investigator blinding was not associated with a statistically significant difference in treatment effect size estimate (difference in treatment effect size estimate = -0.02, 95% confidence interval -0.10 to 0.06, p = 0.641). A positive value (more than zero) across meta-analyses would indicate that a lack of principal-investigator blinding inflated the treatment effect size estimate. The results of the stratified analyses show that none of the meta-analyses characteristics had a statistically significant interaction with the treatment effect size estimate (see **Figure 7.4b**).

7.3.7. Impact of data-analyst blinding on treatment effect size estimate

Due to the small number of trials with adequate blinding of the data-analyst, meta-epidemiological analysis of the data could not be performed for this criterion.

7.3.8. Impact of describing a trial as "double-blind" on treatment effect size estimate

Figure 7.5a displays a forest plot of the difference in treatment effect size estimates between randomized trials with and without reporting "double blinding." Twenty-eight meta-analyses, including 294 trials that analyzed 111,052 patients, provided information for this meta-epidemiological analysis. Trials not described as double-blind tended to exaggerate treatment effect size estimates compared to trials described as double-blind. However, differences in treatment effect size estimates were not statistically significant (difference in treatment effect size estimate = 0.09, 95% confidence interval -0.05 to 0.22, p = 0.203). A positive value (> 0) across meta-analyses indicated that failure to report "double blinding" inflated the treatment effect size estimate. The results of stratified analyses showed that none of the meta-analyses characteristics had a statistically significant interaction with the treatment effect size estimate (see **Figure 7.5b)**.

7.3.9. Impact of blinding of both patients and assessors (double blinding) on treatment effect size estimate

Figure 7.6a shows a forest plot of the difference in treatment effect size estimates in randomized trials with and without blinding of both patients and assessors. Nineteen meta-analyses, including 224 trials that analyzed 106,716 patients, provided information for this meta-epidemiological analysis. Meta-epidemiological results showed a statistically significant difference between the treatment effect size estimate in randomized trials that implemented patient and assessor blinding (double blinding) (difference in treatment effect size estimate = 0.19, 95% confidence interval 0.06 to 0.32, p = 0.004) and the treatment effect size estimate in randomized trials that did not employ blinding. Treatment effect size estimates were significantly larger (0.19) in trials with lack of blinding of both patients and assessors than treatment effect size estimates in trials that blinded patients and assessors. However, the impact of blinding of both patients and assessors on treatment effect size estimates stratified

by examined characteristics of meta-analyses was not statistically significant for any of the characteristics (see **Figure 7.6b)**.

7.3.10. Impact of blinding of patients, assessors, and care-providers (triple blinding) on treatment effect size estimate

Figure 7.7a shows a forest plot of the difference in treatment effect size estimates between randomized trials with the presence and lack of blinding of patients, assessors, and care-providers (triple blinding). Ten meta-analyses, including 151 trials that analyzed 99,293 patients, provided information for this meta-epidemiological analysis. Results of the analysis showed that trials that did not implement patient, assessor, and care-provider blinding had significantly larger treatment effect size estimates (difference in treatment effect size estimate = 0.14, 95% confidence interval 0.03 to 0.25, p = 0.013) than trials that implemented blinding of those three components. These results imply that treatment effect size estimates were 0.14 larger in trials with lack of blinding of patients, assessors, and care-providers than in trials that implemented blinding of those three components. However, results of the stratified analyses show that none of the examined meta-analyses characteristics had a statistically significant interaction with the treatment effect size estimate (see **Figure 7.7b)**.

7.3.11. Impact of using an appropriate method of blinding on treatment effect size estimate

Figure 7.8a shows a forest plot of the difference in treatment effect size estimates between trials with the presence and lack of an appropriate method of blinding. Forty meta-analyses provided information for this meta-epidemiological analysis. The presence of an appropriate method of blinding was not associated with a statistically significant difference in treatment effect size estimate, trials that lacked an appropriate method of blinding tended to inflate treatment effect size estimates compared to trials with the presence of an appropriate method of blinding (difference

in treatment effect size estimate = 0.06, 95% confidence interval -0.06 to 0.18, p = 0.325). A positive value (more than zero) across meta-analyses would have indicated that lack of an appropriate method of blinding inflated the treatment effect size estimate. The results of the stratified analyses showed that differences in treatment effect size estimates between trials with the presence or lack of appropriate blinding were significant ($p < 0.02$) in meta-analyses with a large treatment benefit in overall meta-analysis, but not in meta-analyses with a small treatment benefit. However, none of the other considered factors (heterogeneity of meta-analysis, type of outcome, and dental specialty) had a statistically significant interaction with the treatment effect size estimate (see **Figure 7.8b)**.

7.4. Discussion

Our investigation provides empirical evidence of the impact of bias associated with nine blinding-based criteria (related to patient, assessor, care-provider, and principal-investigator blinding) on the treatment effect size estimate. This analysis is of important to methodologists and researchers in the fields of dental, oral, and craniofacial research. To our knowledge, this study is the first meta-epidemiological study conducted in any medical or dental field that examines continuous outcomes of the impact of blinding of both patients and assessors (double blinding; trial conduct) and of patients, assessors, and care-providers (triple blinding; trial conduct) on treatment effect size estimates in randomized trials.

Our study shows significant differences in treatment effect size estimates in oral health trials based on different types of blinding. For example, trials with lack of patient and assessor blinding had significantly larger treatment effect size estimates compared to trials without lack of patient and/or assessor blinding. Patient blinding and assessor blinding were associated with inflated treatment effect size estimates (significant at the level of patient blinding), while care-provider and principal-investigator blinding were not related to inflated treatment effect size estimates.

Interestingly, blinding of both assessors and patients was found to be associated with the largest overestimation in treatment effect size (0.19). This measured magnitude of bias represents approximately 1/3 of common treatment effect size estimates reported in oral health research [48], such as clinical outcomes in periodontology [46]. The fact that treatment effect size estimates in oral health trials may have been biased due to lack of blinding is concerning, as clinical decision making related to recommended dental treatments and modalities may therefore not be based on valid findings.

The stratified analyses showed that the extent of bias associated with a lack of blinding was not significantly associated with any of the other factors considered at the meta-analysis level. This agrees with a recent study conducted in the area of physical therapy, and is contrary to other meta-epidemiological studies [49], which showed that trials with subjective outcomes exaggerated treatment effect size estimates compared to trials with objective outcomes. This could be due to having a small number of trials with objective outcomes in our study, or to differences between interventions in different medical disciplines.

Reports examining the impact of lack of blinding of patient, therapist, or assessor on treatment effect size estimate were conducted in particular medical fields such as physical therapy [19], thrombosis and cardiovascular disease [15, 21], pediatrics [18], osteoarthritis [45], and low-back pain [16]. The studies reported inconsistent findings. The treatment effect size estimate was smaller in trials that employed patient blinding [15] or assessor blinding [20, 21] in some studies, whereas in other studies the treatment effect size estimate was smaller in trials that lacked patient [17] or assessor blinding [15]. However, an association between the treatment effect size estimate and the presence or lack of blinding was not confirmed in some studies [16, 45]. Furthermore, while the definition of double blinding varied largely among the meta-epidimiological studies with respect to the level of blinding (patient, assessor, and care-provider blinding), a lack of double blinding was found to be associated with exaggerated treatment effect size estimates in general [22, 23, 50]. The

inconsistent findings might be due to the examination of different types of outcome, intervention, and population, to the implementation of different definitions of quality assessment, and to the application of various statistical and modeling approaches [24]. For example, Schulz et al. [50] applied a multiple logistic regression model to analyze data on binary outcomes from 250 trials included in 33 meta-analyses; the definition of double blinding was based on whether the trial's conduct claimed to be a double-blind. Egger et al. [22] defined "double blinding" based on whether the trial was described as double-blind, or included at least assessor blinding; the study analyzed data from 304 trials included in 39 meta-analyses with binary outcomes in several medical fields (infectious diseases, neurology, among others).

Two recent studies [18, 19] that examined the association between lack of blinding of patient, therapist, or assessor, and treatment effect size estimate using continuous outcomes, also reported inconsistent findings. One study assessed the adequacy of patient and assessor blinding in 287 pediatric trials from 17 meta-analyses [18], and showed no significant difference in treatment effect size estimates between studies, based on potential bias related to lack of blinding. Another study assessed 165 physical therapy trials included in 17 meta-analyses and found that trials with a lack of patient or assessor blinding tended to underestimate treatment effect size when compared with trials with appropriate blinding (although, the differences were nonstatistically significant) [19]. It should be noted that in both of the above-mentioned studies, the lack of significant results might be accounted for by the small number of trials, the precision of the analyses performed, and/or the examination of interventions where blinding is not crucial or fundamental (i.e., outcomes are objective or automated with no assessor involvement).

Because the concept of blinding is implemented at multiple levels of a trial (e.g., patients, assessors, care providers, data analysts, investigators), there is confusion when describing the level of blinding implemented. For example, "double blinding" or "triple blinding" may refer to the blinding at any two or three of the

previous levels. Failure to clearly report the levels that such terms refer to leads to confusion. Investigators of randomized trials conducted in the field of dentistry need to implement blinding of patients, assessors, care providers, data analysts, and other personnel when applicable, and explicitly report on mechanisms used to achieve and assure successful blinding, as recommended by the Consolidated Standards of Reporting Trials (CONSORT) statement. In addition, investigators of randomized trials should state the levels (e.g., patients, assessors, care providers) and components (e.g. allocation, outcomes assessed, details of interventions) they are referring to when using the terms "double" or "triple" in a blinded trial. Alternatively, they should avoid using the terms "double" or "triple" blind trial when reporting trial findings, and report who was blinded and to what components blinding was achieved, so the reader can evaluate the potential associated bias. As well, editors and peer reviewers of dental journals should require authors of randomized trials to adhere to the CONSORT guidelines and insist on adequate conduct and reporting of blinding in submitted randomized trials.

When we examined the association between double blinding and treatment effect size estimate, we performed the analysis on two different criteria: reporting of "double blinding" as a term in a trial, and actual conduct of blinding of both assessors and patients. Haahr and Hróbjartsson [51], who examined a random sample of RCTs from the Cochrane Central Register of Controlled Trials, suggested that it is incorrect to assume blinding of a trial participant based only on the term "double blind." The study found that the blinding of patients, care providers, and assessors was clearly described in only three (2%) out of 200 blinded randomized trials, while 56% of the trials failed to describe the blinding status of any individual involved in a trial. That study concluded that either patients, care providers, or assessors were not blinded in one of five "double blind" randomized trials. Another trial study [52] showed that adequate reporting of blinding was common in some medical journals, and that inadequate reporting of blinding does not necessarily entail a lack of actual blinding.

For example, it was reported that randomized trial authors frequently use blinding, although they fail to describe its methods. For instance, authors of RCTs failed to report the blinding status of patients in 26% of trials, and patients were actually blinded in 20% of trials in which patients were not reported to be blinded. Similar results were found in a recent study by Kahan et al. [53], who indicated that blinding of outcome assessors is uncommonly used and inadequately reported in a cohort of 258 trials published in four high-impact medical journals.

An implication that can be drawn from our meta-epidemiological work is that authors of systematic reviews of oral health interventions should consider excluding dental trials with lack of blinding from meta-analyses, or at least perform sensitivity analyses on included trials based on the adequacy of blinding. In all instances, authors should consider the likely level of bias associated with reported (or unreported) blinding status when interpreting the findings of a quantitative analysis.

The above-mentioned implications should be considered with caution, particularly in oral health trials involving surgical or device interventions (such as orthodontic trials) where patient blinding is not feasible; in this case, informing patients with details of the intervention is required, and sometimes ethically compulsory. While these trials are prone to biases, particularly when the trials examine self-reported outcomes, implementation of blinding in the conduct of these trials is often unacceptable for ethical and practical reasons. For example, in the case of trials comparing surgical interventions to nonsurgical interventions (e.g., comparison of surgical removal of wisdom teeth versus retention or conservative management), patients and surgeons cannot be blinded. However, trialists may consider using "expertise-based" trial design, whereby patients are allocated to multiple surgeons and each surgeon performs a single treatment [54]. While this design helps to minimize performance bias related to surgeon blinding, it does not ensure patient blinding [55]. Furthermore, in trials where patients cannot be blinded (e.g., comparison of manual versus electric toothbrushing), trialists may consider

using objective outcomes that have established validity and reliability [54]. When blinding is feasible, trialists should consider blinding as many trial components (participants, assessors, care-providers, statisticians, investigators) as ethically and practically possible.

7.5. Conclusions

We found significant differences in treatment effect size estimates between oral health trials based on lack of patient and assessor blinding; Trials that lacked patient and assessor blinding had significantly larger treatment effect size estimates. Treatment effect size estimates were 0.19 and 0.14 larger in trials with lack of blinding of both patients and assessors (double blinding) and blinding of patients, assessors, and care-providers (triple blinding). No significant differences were identified in other blinding criteria. Based on this evidence, investigators of systematic reviews conducted in dental, oral, and craniofacial trials should perform sensitivity analyses based on the adequacy of blinding in included trials. The potential impact of blinding on bias in treatment effect size estimates suggests that dental journal editors and reviewers should insist on adequate blinding (when feasible) with respect to trial conduct and reporting, in published trials' reports.

Table 7.1. Guidelines for quality assessment of included trials [7, 31, 56-62]

Items /Definitions	Yes	No	Unclear
Performance Bias			
Patient blinding [39]: Was knowledge of the allocated intervention adequately prevented during the study? "Blinding of patients is a must when outcomes are subjective or self-reported. When Outcomes are measured by an assessor, then assessors should be blinded to group allocation. When Outcomes are automated (there is no assessor involved) then, blinding of participants or assessors is not an issue."	Any one of the following: No blinding or incomplete blinding, but the review authors judge that the outcome is not likely to be influenced by lack of blinding (Automated outcome or administrative); Blinding of participants and key study personnel ensured, and unlikely that the blinding could have been broken; Objectives automatized outcomes coming from databases or hospital register office.	Any one of the following: No blinding or incomplete blinding, and the outcome is likely to be influenced by lack of blinding; Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding.	Any one of the following: Insufficient information to permit judgement of 'Low risk' or 'High risk'; The study did not address the issue of blinding.
Blinded therapist/care-provider	The study describes in the title, abstract, or text that the therapists/care-providers were blinded. The blinding was appropriate.	The study describes in the title, abstract, or text that the therapists/care-providers were not blinded, or because of the nature of the intervention (e.g., exercise prescription or supervision, etc.), the therapist could not be blinded.	There is insufficient information to permit a judgment.
Blinded principal-investigator	The study describes in the title, abstract, or text that the investigator was blinded. The blinding was appropriate.	The study describes in the title, abstract, or text that the investigator was not blinded.	There is insufficient information to permit a judgment.
Blinded statistician	The study describes in the title, abstract, or text that the statistician was blinded. The blinding was appropriate.	The study describes in the title, abstract, or text that the statistician was not blinded.	There is insufficient information to permit a judgment.
Detection Bias			
Assessor blinding [39]: Was knowledge of the allocated intervention adequately prevented during the study? Detection bias due to knowledge of the allocated interventions by outcome assessors.	Any one of the following: No blinding of outcome assessment, but the review authors judge that the outcome measurement is not likely to be influenced by lack of blinding; Blinding of outcome assessment ensured, and unlikely that the blinding could have been broken.	Any one of the following: No blinding of outcome assessment, and the outcome measurement is likely to be influenced by lack of blinding; Blinding of outcome assessment, but likely that the blinding could have been broken and the outcome measurement is	Any one of the following: Insufficient information to permit judgement of 'Low risk' or 'High risk'; The study did not address the issue of blinding.

		likely to be influenced by lack of blinding.	
Detection/Performance Bias			
Double blinding (i.e., blinding of both patients and assessors) [39]	Both patient blinding and assessor blinding were judged as having low risk of bias.	Both patient blinding and assessor blinding were judged as having high risk of bias	Both patient blinding and assessor blinding were judged as having unclear risk of bias.
Study described as double blind	"Double blind" is the description in the study related to "blindness." Also, it should be stated that neither the person doing the assessments nor the study participants could identify the intervention being assessed.	Not described as double blind.	There is insufficient information to permit a judgment.
Triple blinding (i.e., blinding of patients, assessors, and caregivers)	Both patient blinding and assessor blinding were judged as having low risk of bias. Also, care-providers are blinded.	Both patient blinding and assessor blinding were judged as having high risk of bias. Also, care-providers are not blinded.	Both patient blinding and assessor blinding were judged as having unclear risk of bias. Also, care-providers are judged as "unclear".
The method of blinding was appropriate	The authors use the blinding method appropriately. Blinding of participants/patients is a "must" when outcomes are subjective or self-reported. When outcomes are measured by an assessor, the assessors should be blinded to group allocation. Also, score "completely done" when it is unlikely that the blinding could have been broken and the nonblinding of others is unlikely to introduce bias. No blinding, but the review authors judge that the outcome and the outcome measurement are not likely to be influenced by lack of blinding. Objectives automatized outcomes coming from databases or hospital register office.	There is no blinding or incomplete blinding is performed, and the outcome or outcome measurement is likely to be influenced by lack of blinding.	There is insufficient information to permit a judgment.

Table 7.2. Blinding in randomized trials of oral health interventions (N = 540)

Domain	Risk of Bias Assessment, N (%)		
	Low Risk	High Risk	Unclear Risk
Blinding of patients/participants	386 (71.5)	7 (1.3)	147 (27.2)
Blinding of assessors	321 (59.4)	16 (3.0)	203 (37.6)
Double blinding (blinding of both patients and assessors) †	273 (72.8)	7 (1.9)	95 (25.3)
Triple blinding (blinding of patients, assessors, and care-providers) ‡	117 (76.5)	7 (4.6)	29 (19.0)
Item	**Quality Assessment, N (%)**		
	Yes	No	Unclear/Not reported
Study described as double-blind	181 (33.5)	358 (66.3)	1 (0.2)
Blinding of assessors	321 (59.4)	16 (3.0)	203 (37.6)
Blinding of patients	192 (35.6)	69 (12.8)	279 (51.7)
Blinding of therapists/care-providers	134 (24.8)	356 (65.9)	50 (9.3)
Blinding of principal investigator	33 (6.1)	10 (1.9)	497 (92.0)
Blinding of data analyst	9 (1.7)	3 (0.6)	528 (97.8)
Method of blinding appropriate	286 (53)	17 (3.1)	237 (43.9)

† Does not equal 100 % for overall, as the item was not applicable in 165 trials.
‡ Does not equal 100 % for overall, as the item was not applicable in 387 trials.

Table 7.3. Type of blinding in randomized trials of oral health interventions (N = 540); N (%)

Component	Random allocation			Hypothesis			Details of intervention			Outcome assessment			Data analysis		
	Yes	No	Unclear/ NR	Yes	No	Unclear/ NR	Yes	No	Unclear/ NR	Yes	No	Unclear/ NR	Yes	No	Unclear/ NR
Participants	194 (35.93)	70 (12.96)	276 (51.11)	1 (0.19)	12 (2.22)	527 (97.59)	2 (0.37)	221 (40.93)	317 (58.70)	0 (0.0)	71 (13.15)	469 (86.85)	0 (0.0)	0 (0.0)	540 (100.0)
Assessors	322 (59.63)	15 (2.78)	203 (37.59)	1 (0.19)	11 (2.04)	528 (97.78)	8 (1.48)	37 (6.85)	495 (91.67)	NA	NA	NA	0 (0.0)	0 (0.0)	540 (100.0)
Principal Investigator	32 (5.93)	10 (1.85)	498 (92.22)	0 (0.0)	0 (0.0)	540 (100.00)	0 (0.0)	0 (0.0)	540 (100.0)	1 (0.19)	2 (0.37)	533 (99.44)	0 (0.0)	0 (0.0)	540 (100.0)
Care-providers	136 (25.19)	351 (65.00)	53 (9.81)	2 (0.37)	10 (1.85)	528 (97.78)	0 (0.0)	0 (0.0)	540 (100.0)	1 (0.19)	16 (2.96)	523 (96.85)	0 (0.0)	0 (0.0)	540 (100.0)
Statisticians	9 (1.67)	3 (0.56)	528 (97.78)	1 (0.19)	0 (0.0)	539 (99.81)	0 (0.0)	0 (0.0)	540 (100.0)	0 (0.0)	0 (0.0)	540 (100.0)	NA	NA	NA

NA, not applicable; NR, not reported.

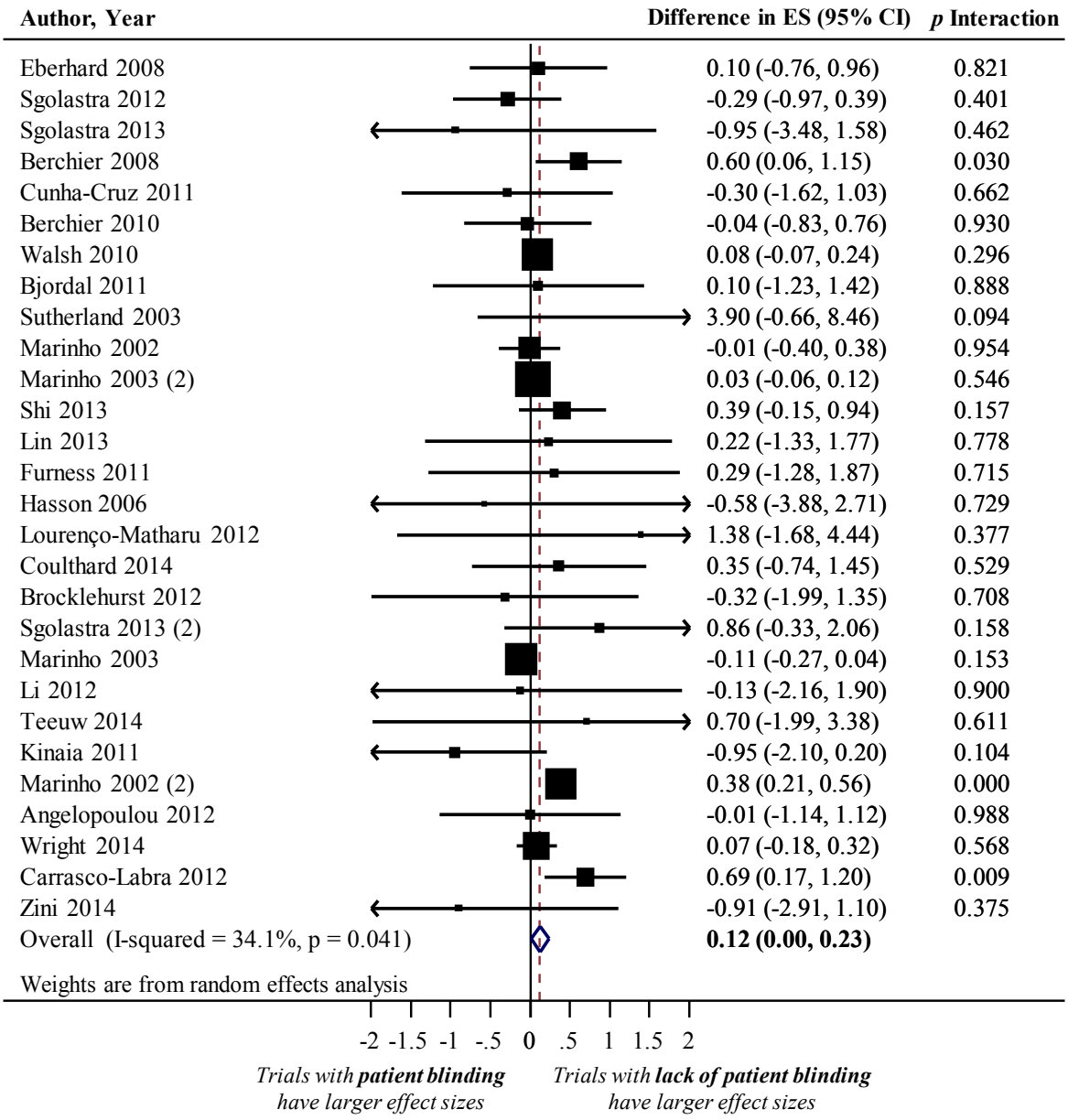

Figure 7.1a. Difference in treatment effect size (ES) estimate between trials with presence and lack of patient blinding. A positive value (more than zero) across meta-analyses indicates that treatment effect size estimates are larger in trials that lack patient blinding compared to trials with adequate patient blinding.

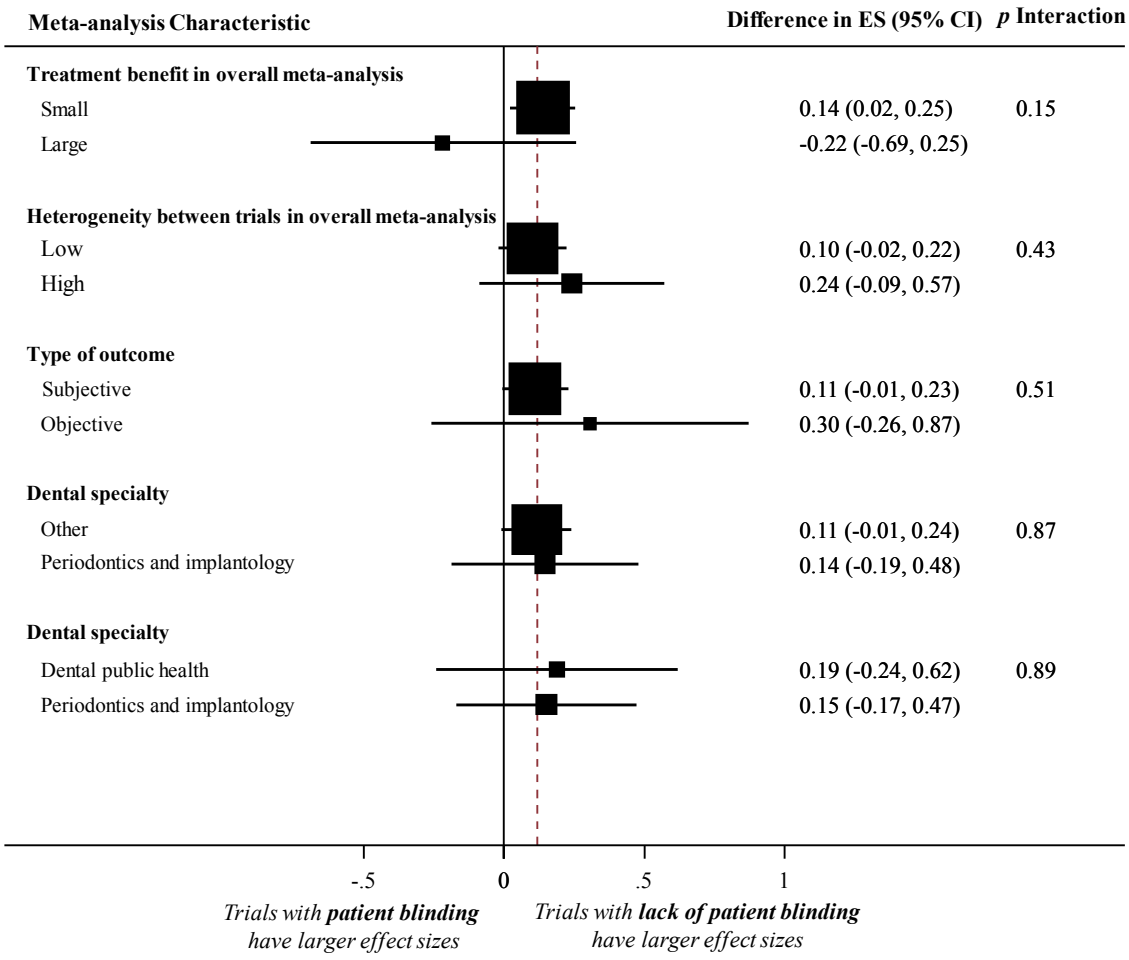

Figure 7.1b. Forest plot of the difference in treatment effect size (ES) estimate between trials with presence and lack of patient blinding stratified by meta-analyses characteristics.

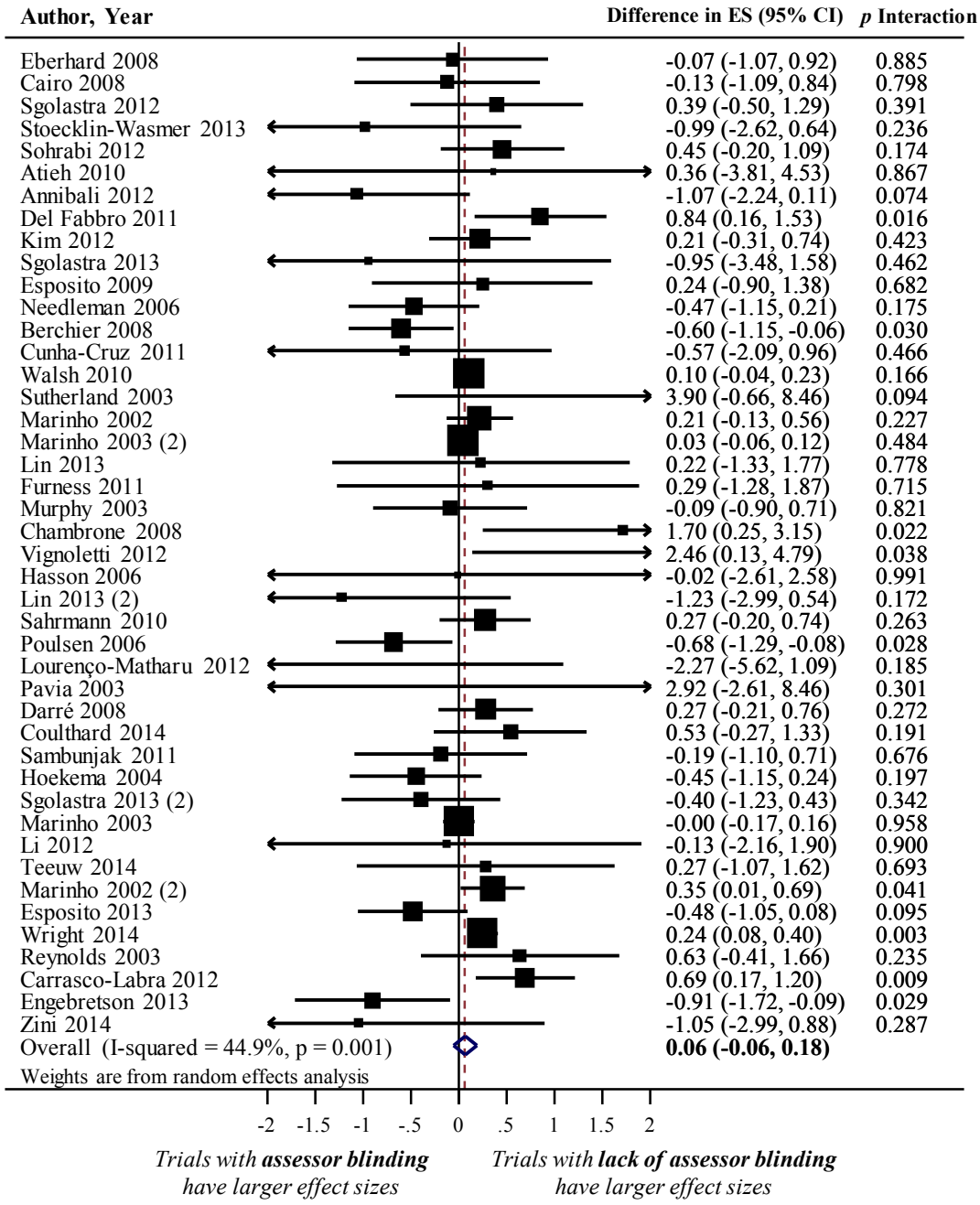

Figure 7.2a. Difference in treatment effect size (ES) estimate between trials with presence and lack of assessor blinding. A positive value (more than zero) across meta-analyses indicates that lack of assessor blinding inflates the treatment effect size when compared with trials with adequate assessor blinding.

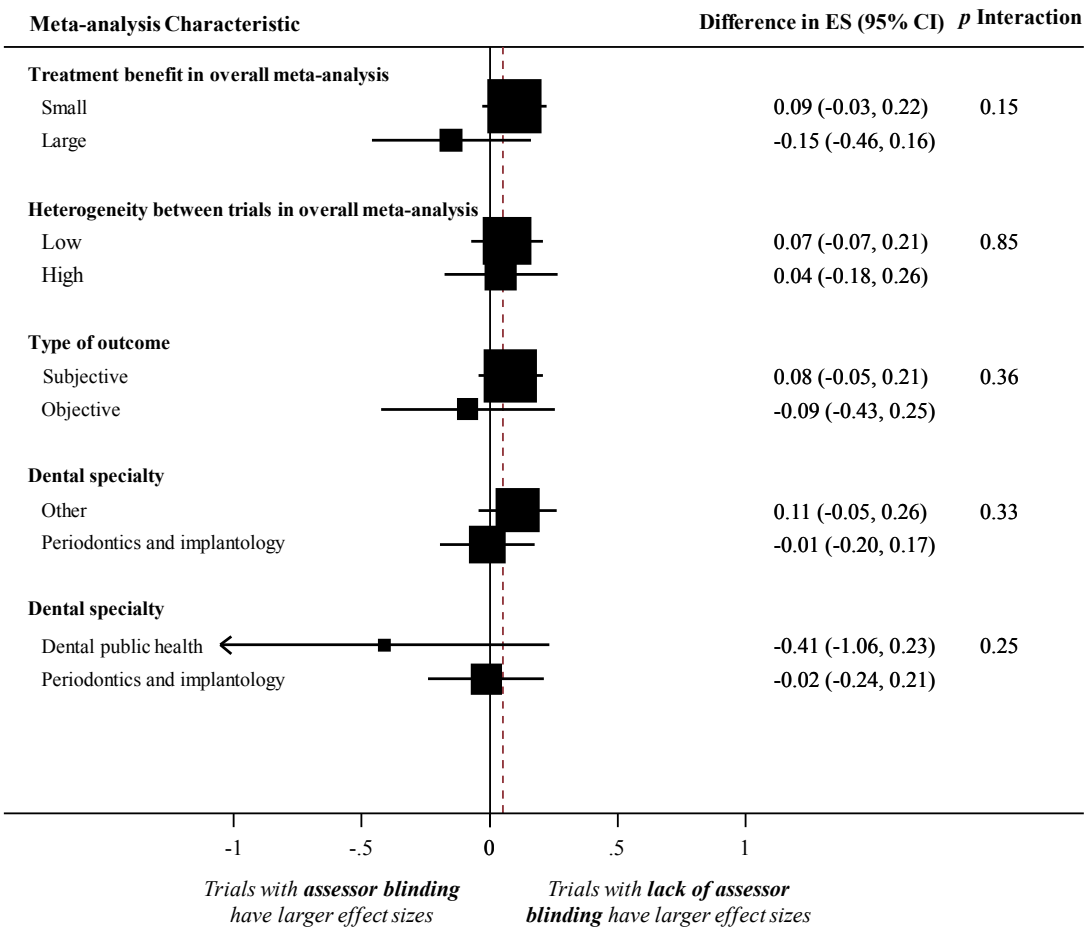

Figure 7.2b. Forest plot of the difference in treatment effect size (ES) estimate between trials with presence and lack of assessor blinding stratified by meta-analyses characteristics.

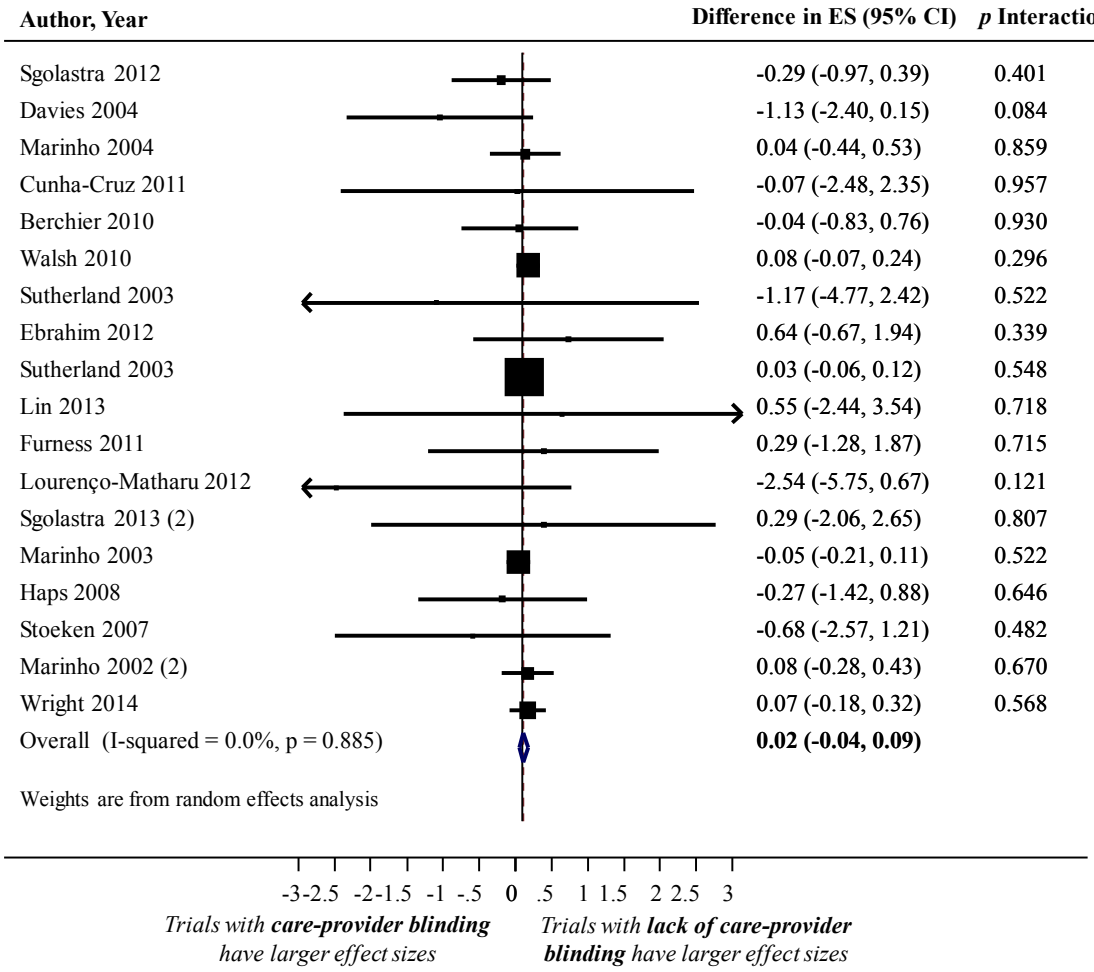

Figure 7.3a. Difference in treatment effect size (ES) estimate between trials with presence and lack of care-provider blinding. A positive value (more than zero) across meta-analyses indicates that the lack of care-provider blinding inflates the treatment effect sizes when compared with trials with adequate care-provider blinding.

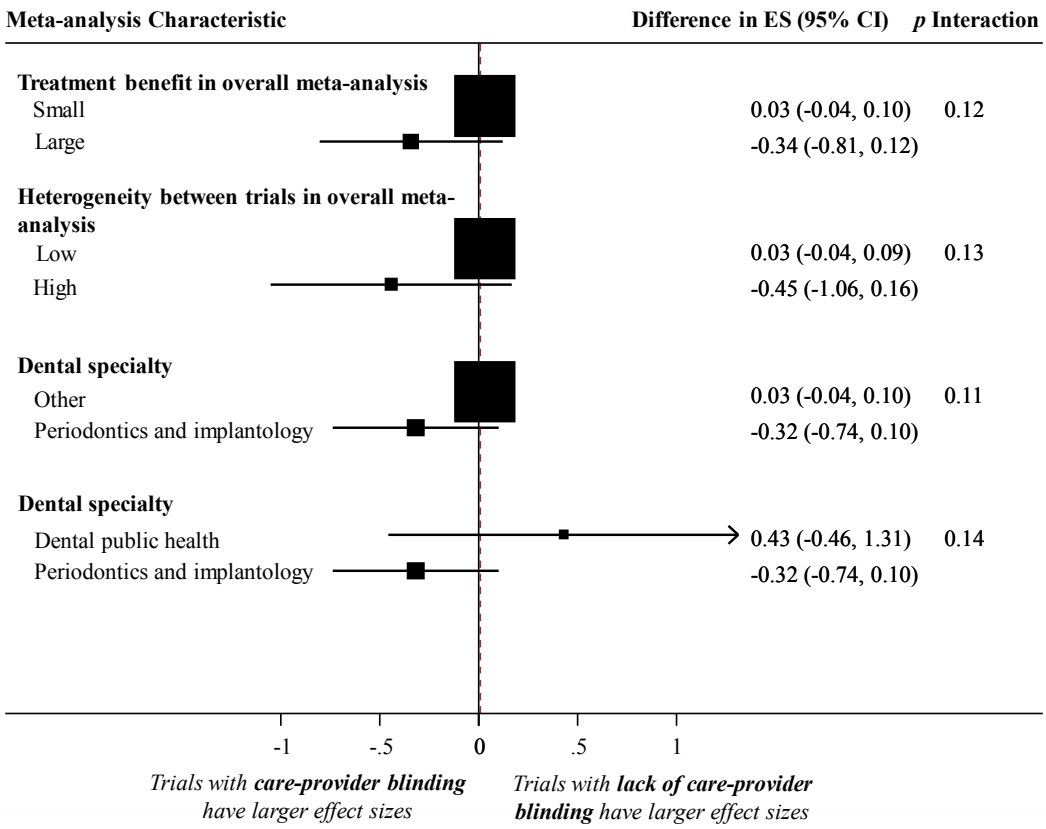

Figure 7.3b. Forest plot of the difference in treatment effect size (ES) estimate between trials with presence and lack of care-provider blinding stratified by meta-analyses characteristics.

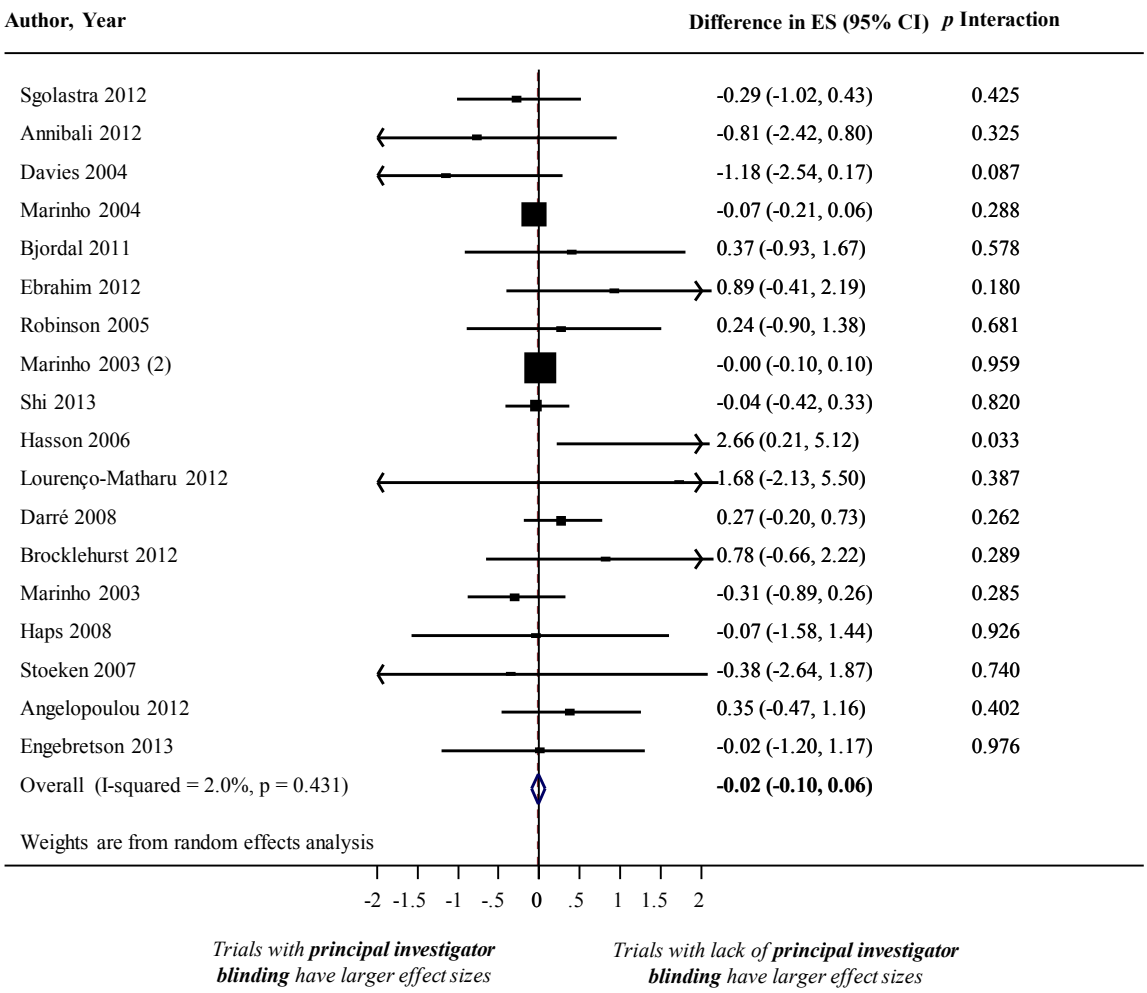

Figure 7.4a. Difference in treatment effect size (ES) estimate between trials with presence and lack of principal-investigator blinding. A positive value (more than zero) across meta-analyses indicates that the lack of principal-investigator blinding inflates the treatment effect sizes when compared with trials with adequate principal investigator blinding.

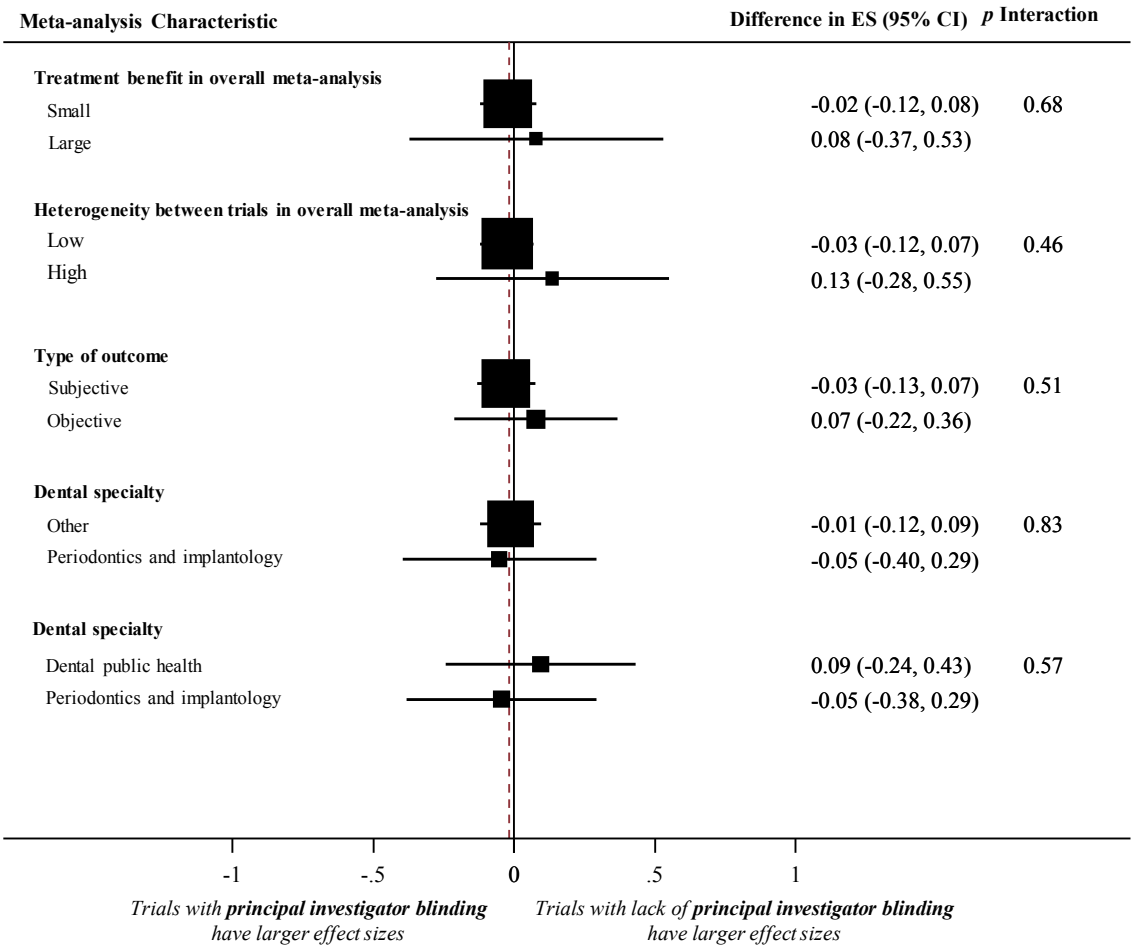

Figure 7.4b. Forest plot of the difference in treatment effect size (ES) estimate between trials with presence and lack of principal-investigator blinding stratified by meta-analyses characteristics.

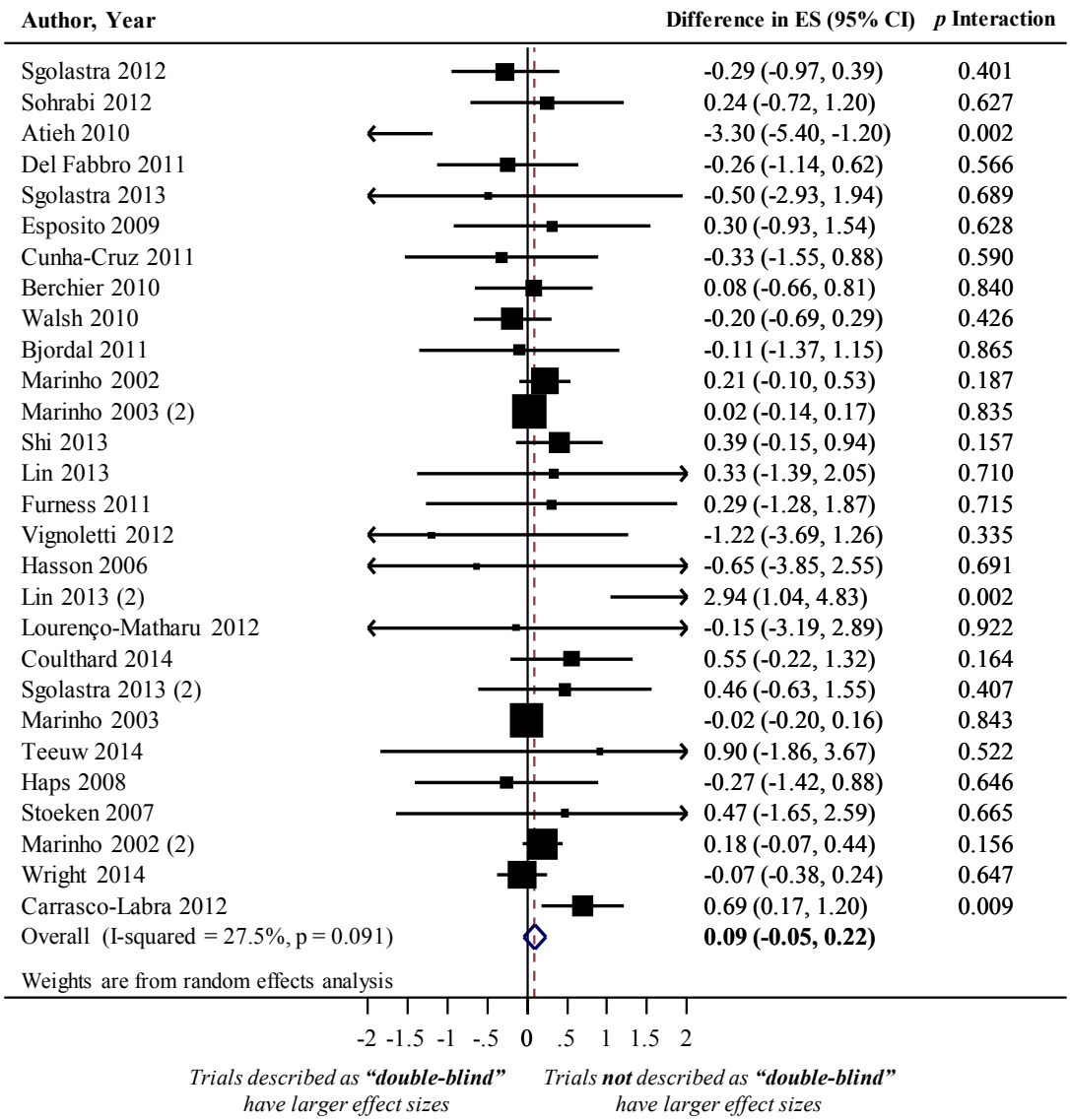

Figure 7.5a. Difference in treatment effect size (ES) estimate between trials with presence and lack of "double-blinded" description. A positive value (more than zero) across meta-analyses indicates that trials not described as "double-blinded" inflate the treatment effect size when compared with trials described as "double blinded."

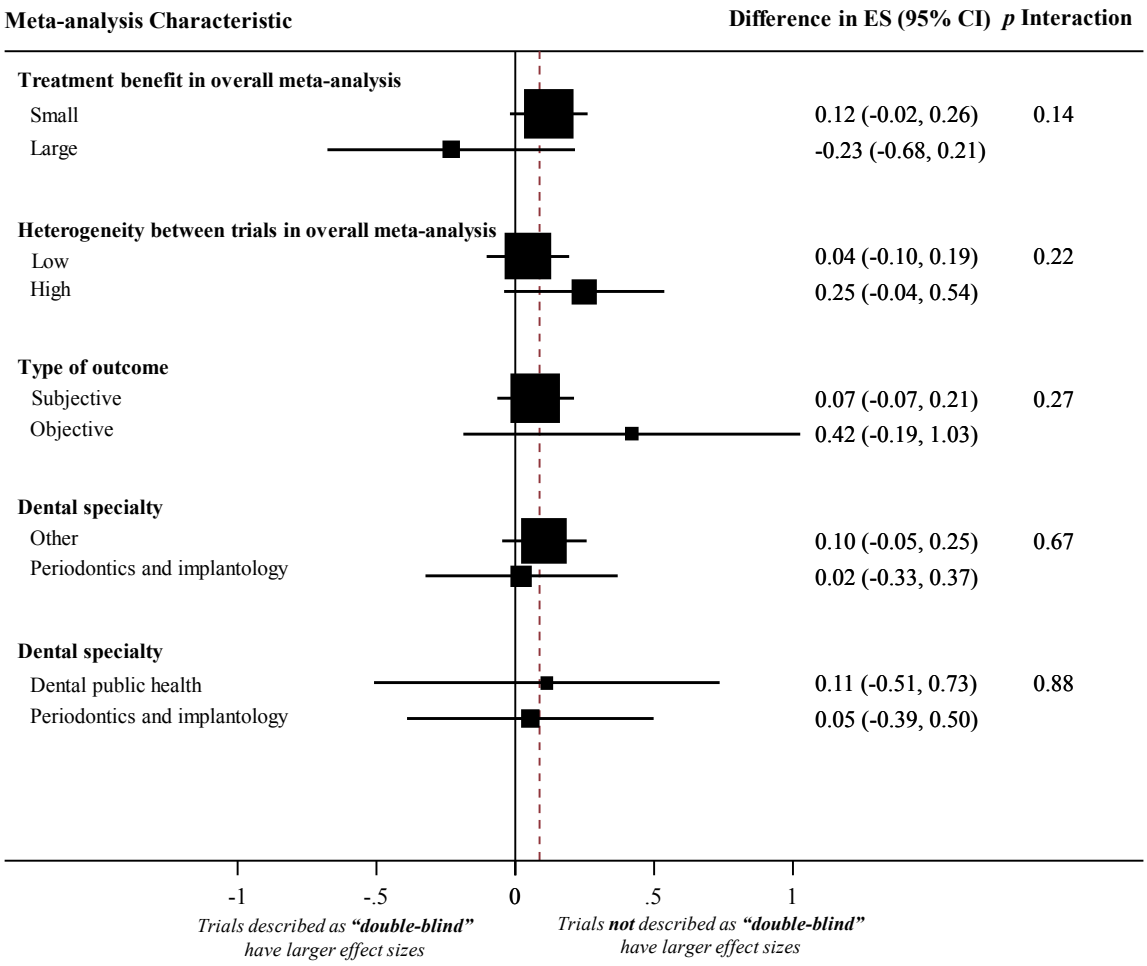

Figure 7.5b. Forest plot of the difference in treatment effect size (ES) estimate between trials with presence and lack of "double-blinded" description stratified by meta-analyses characteristics.

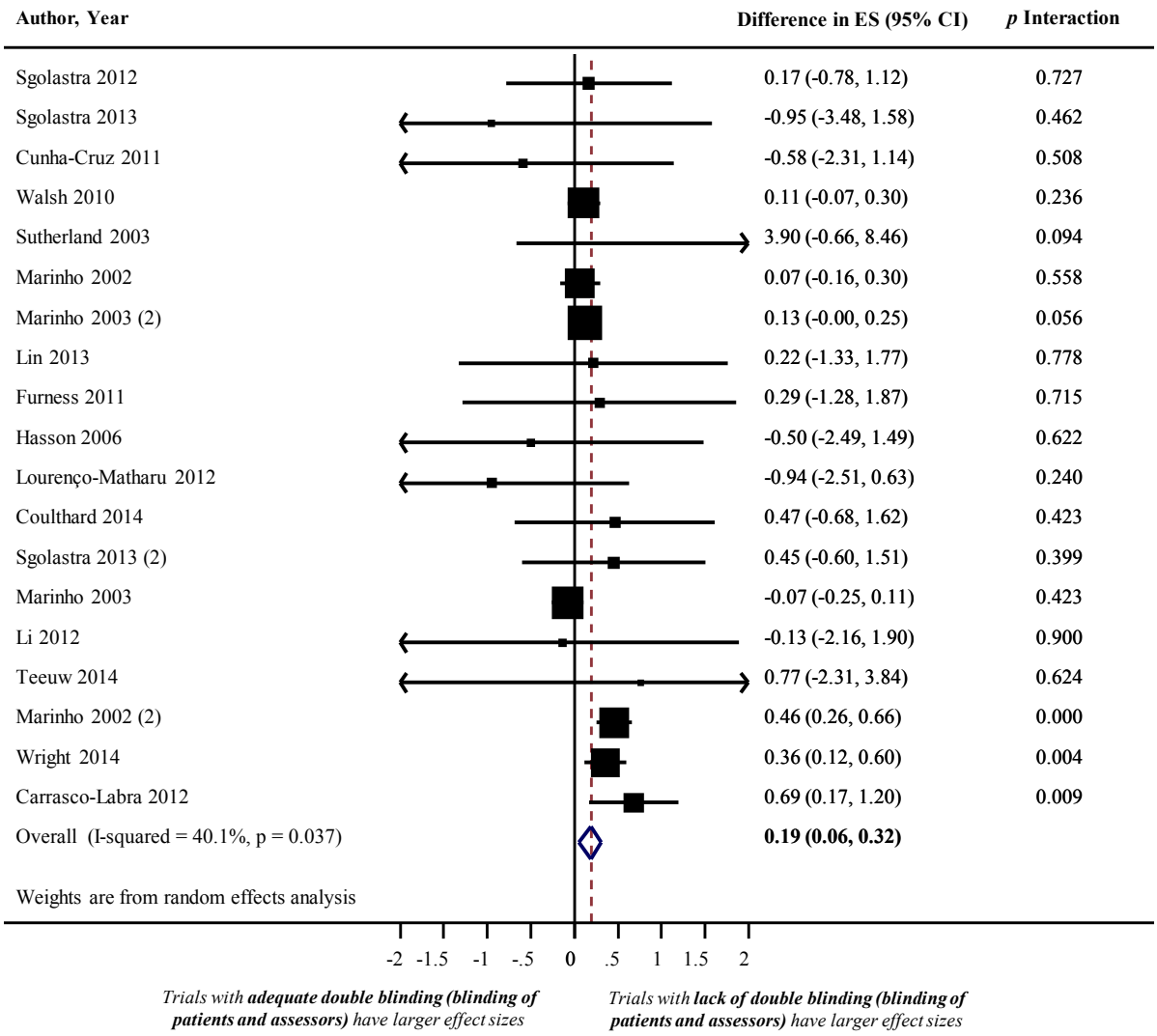

Figure 7.6a. Difference in treatment effect size (ES) estimate between trials with and without blinding of both patients and assessors. A positive value (more than zero) across meta-analyses indicates that lack of blinding of both patients and assessors inflates the treatment effect size when compared with trials with adequate blinding of patients and assessors

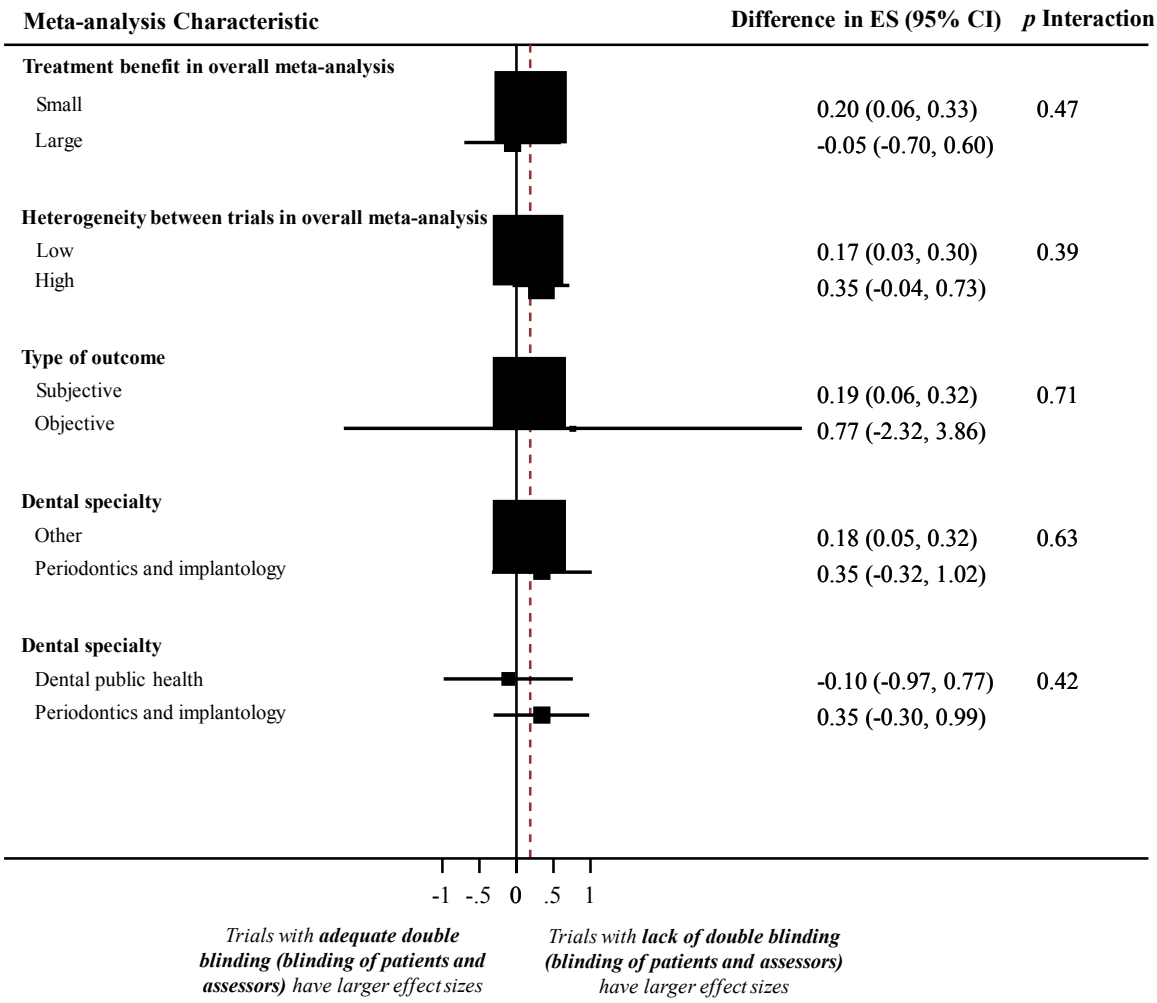

Figure 7.6b. Forest plot of the difference in treatment effect size (ES) estimate between trials with and without blinding of both patients and assessors stratified by meta-analyses characteristics.

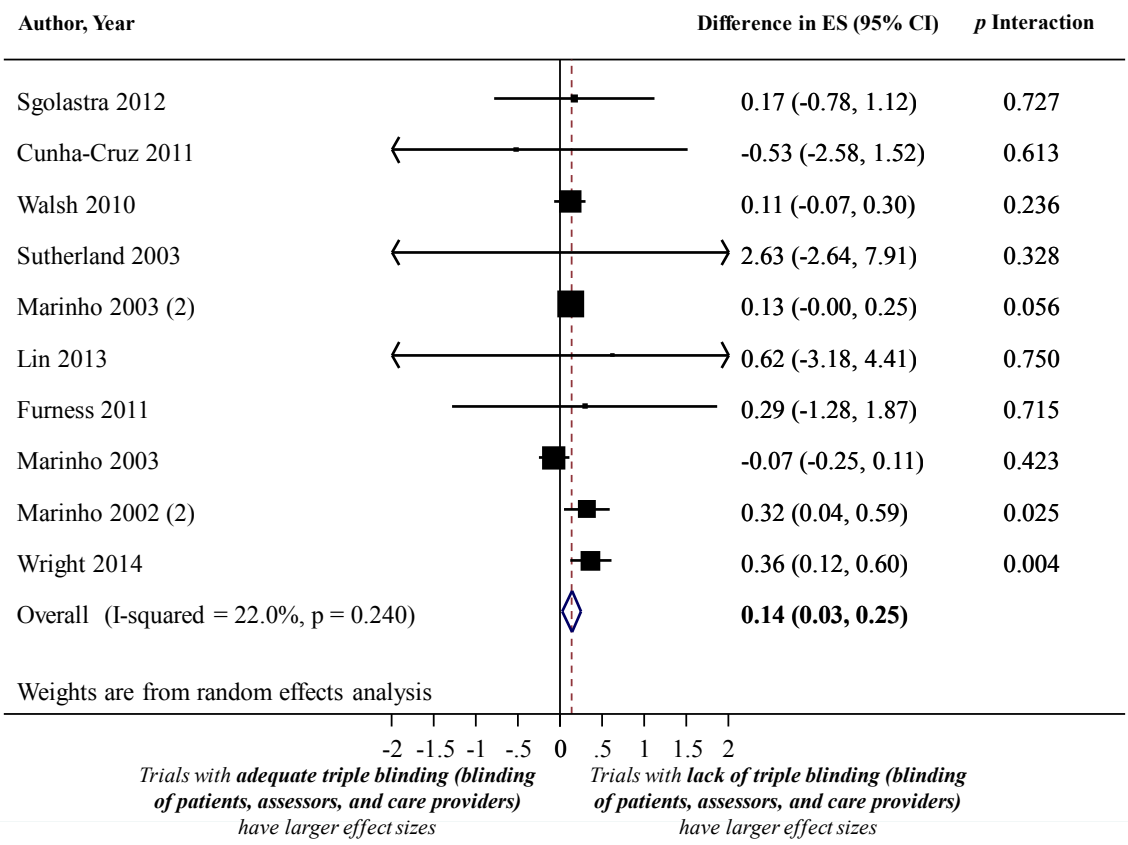

Figure 7.7a. Difference in treatment effect size (ES) estimate between trials with and without blinding of patients, assessors, and care providers. A positive value (more than zero) across meta-analyses indicates that lack of blinding of patients, assessors, and care providers inflates the treatment effect sizes when compared with trials adequately blinded in those 3 components

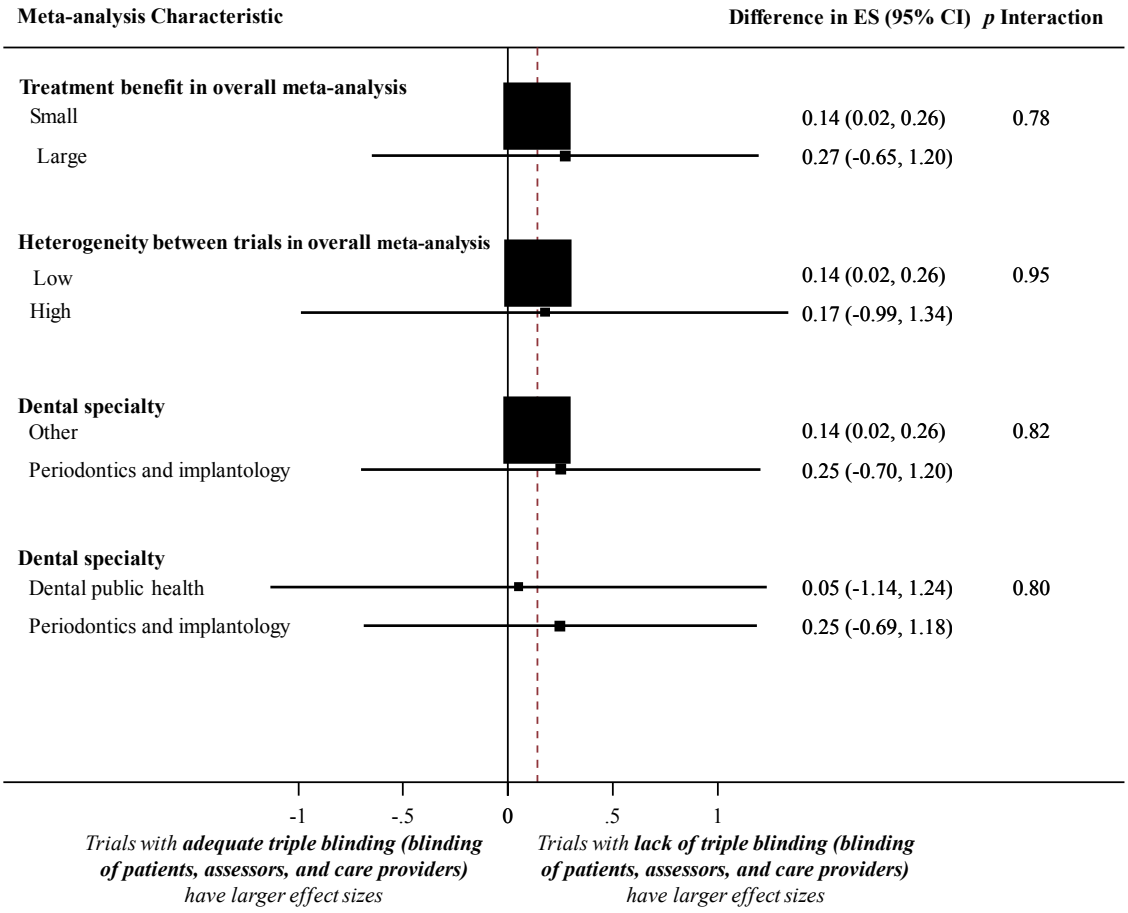

Figure 7.7b. Forest plot of the difference in treatment effect size (ES) estimate between trials with and without blinding of patients, assessors, and care providers stratified by meta-analyses characteristics.

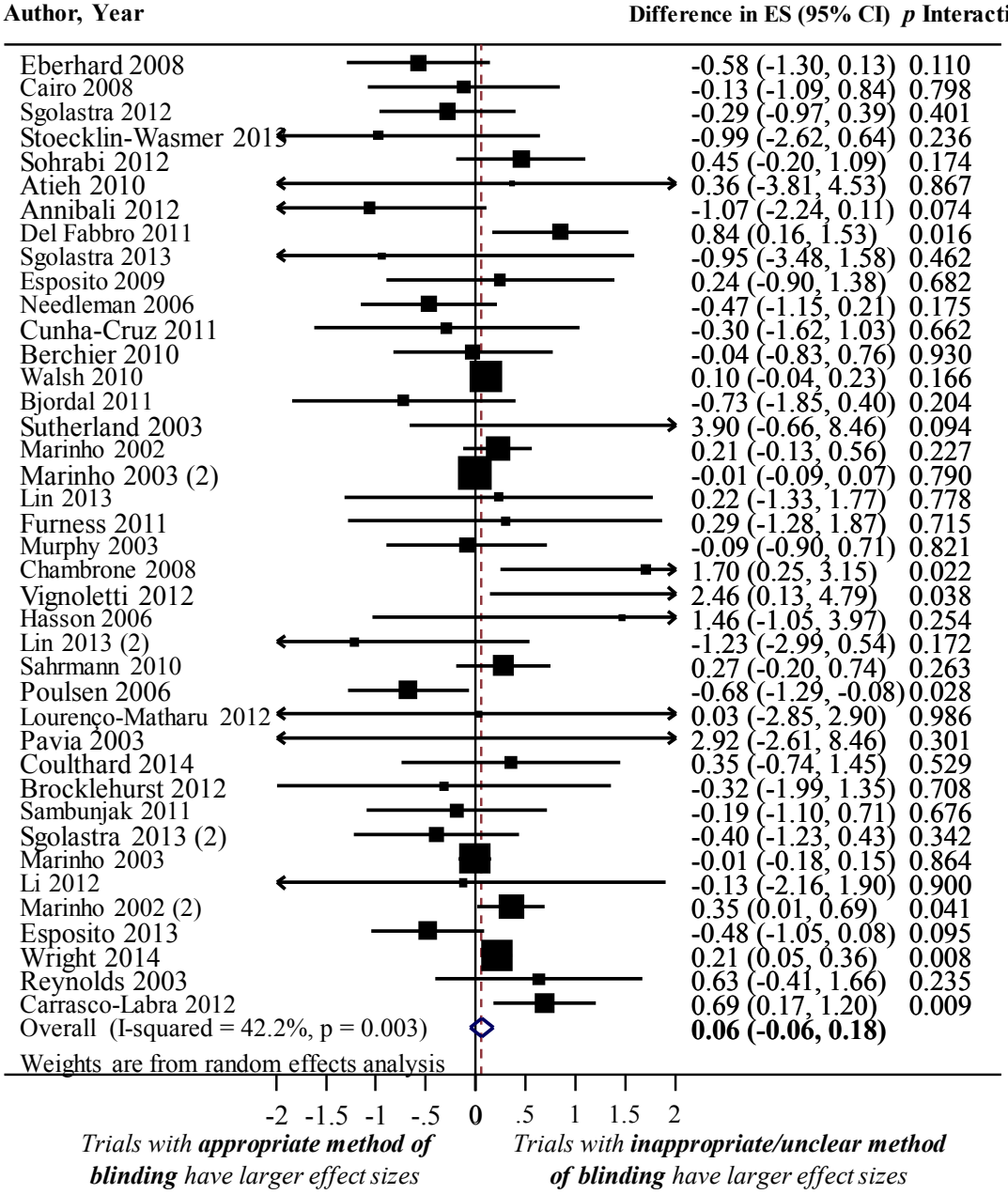

Figure 7.8a. Difference in treatment effect size (ES) estimate between trials with and appropriate method of blinding. A positive value (more than zero) across meta-analyses indicates that lack of an appropriate method of blinding inflates the treatment effect size.

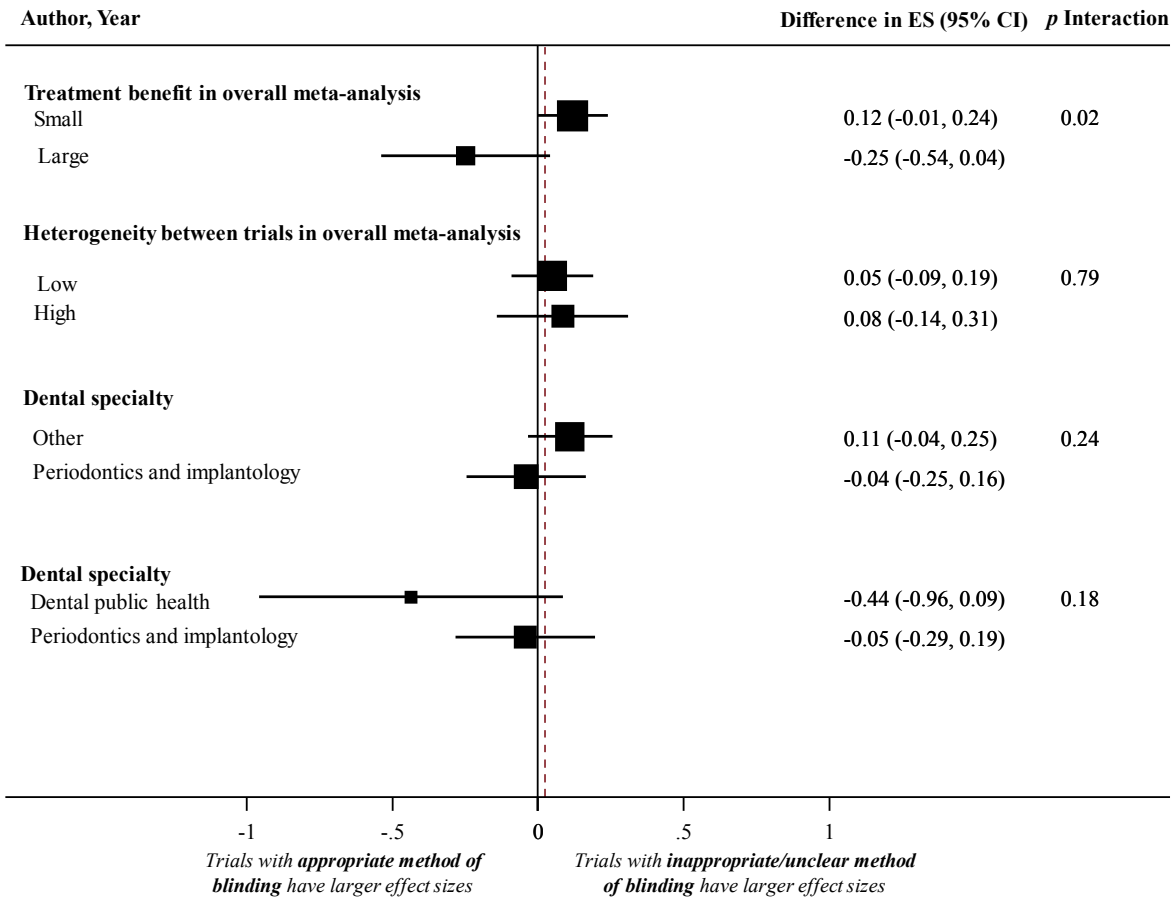

Figure 7.8b. Forest plot of the difference in treatment effect size (ES) estimate between trials with and without appropriate method of blinding stratified by meta-analyses characteristics.

References

[1] Needleman I, Moher D, Altman DG, Schulz KF, Moles DR, Worthington H. Improving the clarity and transparency of reporting health research: a shared obligation and responsibility. J Dent Res 2008; 87:894-5.

[2] Needleman I, Worthington H, Moher D, Schulz K, Altman DG. Improving the completeness and transparency of reports of randomized trials in oral health: the CONSORT statement. Am J Dent 2008; 21:7-12.

[3] Saltaji H, Cummings GG, Armijo-Olivo S, Major MP, Amin M, Major PW, et al. A descriptive analysis of oral health systematic reviews published 1991–2012: cross sectional study. PloS One 2013; 8:e74545.

[4] Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. Health Technol Assess 1999; 3:i-iv, 1-98.

[5] Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. J Am Podiatr Med Assoc 2001; 91:437-42.

[6] Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. Int J Surg 2011.

[7] Armijo-Olivo S, Fuentes J, Ospina M, Saltaji H, Hartling L. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: a descriptive analysis. BMC Med Res Methodol 2013; 13:116.

[8] Delgado-Rodriguez M, Llorca J. Bias. J Epidemiol Community Health 2004; 58:635-41.

[9] Gluud LL. Bias in clinical intervention research. Am J Epidemiol 2006; 163:493-501.

[10] Pandis N. Blinding or masking. Am J Orthod Dentofacial Orthop 2012; 141:389-90.

[11] Schulz KF, Grimes DA. Blinding in randomised trials: Hiding who got what. Lancet 2002; 359:696-700.

[12] Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med 2008; 148:295-309.

[13] McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. BMJ 2002; 324:1448-51.

[14] Nuesch E, Reichenbach S, Trelle S, Rutjes AW, Liewald K, Sterchi R, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. Arthritis Rheum 2009; 61:1633-41.

[15] Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 2002; 287:2973-82.

[16] van Tulder MW, Suttorp M, Morton S, Bouter LM, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine 2009; 34:1685-92.

[17] Hempel S, Miles J, Suttorp MJ, Wang Z, Johnsen B, Morton S, et al. Detection of associations between trial quality and effect sizes. Rockville (MD) 2012.

[18] Hartling L, Hamm MP, Fernandes RM, Dryden DM, Vandermeer B. Quantifying bias in randomized controlled trials in child health: a meta-epidemiological study. PloS One 2014; 9:e88008.

[19] Armijo-Olivo S, Fuentes J, da Costa BR, Saltaji H, Ha C, Cummings GG. Blinding in physical therapy trials and its association with treatment effects: A meta-epidemiological study. Am J Phys Med Rehabil 2016 May 4.

[20] Hrobjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ 2012; 344:e1119.

[21] Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999; 282:1054-60.

[22] Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 2003; 7:1-76.

[23] 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; 336:601-5.

[24] Berkman ND, Santaguida PL, Viswanathan M, Morton SC. The empirical evidence of bias in trials measuring treatment differences. Rockville (MD) 2014.

[25] Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, Flores-Mir C. Methodological characteristics and treatment effect sizes in oral health randomised controlled trials: Is there a relationship? Protocol for a meta-epidemiological study. BMJ Open 2014; 4:e004527.

[26] American Dental Association (ADA). Specialty Definitions 2015. Available at: http://www.ada.org/495.aspx. Accessed July 2015.

[27] Piromchai P, Thanaviratananich S, Laopaiboon M. Systemic antibiotics for chronic rhinosinusitis without nasal polyps in adults. Cochrane Database Syst Rev 2011: CD008233.

[28] Trees AH, Howe TE, Grant M, Gray HG. WITHDRAWN: Exercise for treating anterior cruciate ligament injuries in combination with collateral ligament and meniscal damage of the knee in adults. Cochrane Database Syst Rev 2011: CD005961.

[29] Armijo-Olivo S, Ospina M, da Costa BR, Egger M, Saltaji H, Fuentes CJ, et al. Poor reliability between cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. PloS One 2014; 9(5): e96920.

[30] Armijo-Olivo S, Saltaji H, da Costa BR, Fuentes J, Ha C, Cummings GG. What is the influence of randomisation sequence generation and allocation concealment on treatment effects of physical therapy trials? A meta-epidemiological study. BMJ Open 2015; 5:e008562.

[31] Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Contr Clin Trials 1996; 17:1-12.

[32] Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 1998; 51:1235-41.

[33] van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. Spine 2003; 28:1290-9.

[34] van Tulder MW, Assendelft WJ, Koes BW, Bouter LM. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for Spinal Disorders. Spine 1997; 22:2323-30.

[35] Moseley AM, Herbert RD, Sherrington C, Maher CG. Evidence for physiotherapy practice: a survey of the Physiotherapy Evidence Database (PEDro). Aust J Physiother 2002; 48:43-9.

[36] Sherrington C, Herbert RD, Maher CG, Moseley AM. PEDro. A database of randomized trials and systematic reviews in physiotherapy. Man Ther 2000; 5:223-6.

[37] Bizzini M, Childs JD, Piva SR, Delitto A. Systematic review of the quality of randomized controlled trials for patellofemoral pain syndrome. J Orthop Sports Phys Ther 2003; 33:4-20.

[38] Armijo-Olivo S, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. Phys Ther 2008; 88:156-75.

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

[40] Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med 2002; 21:1513-24.

[41] DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: An update. Contemp Clin Trials 2007; 28:105-14.

[42] Sterne JAC, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med 2002; 21:1513-24.

[43] Cohen J. Chapter 1: The Concepts of Power Analysis. In: Cohen J, editor. Statistical Power Analysis for the Behavioral Sciences. Second ed. Hillsdale, New Jersey: Academic Press, Inc. 1988. p. 1-17.

[44] Hempel S, Suttorp M, Miles J. Empirical evidence of associations between trial quality and effect size. Rockville (MD), Agency for Healthcare Research and Quality (US); 2011 Jun. Report No.:11-EHC045-EF.

[45] Nuesch E, Reichenbach S, Trelle S, Rutjes AWS, Liewald K, Sterchi R, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: A meta-epidemiologic study. Arthritis Rheum 2009; 61:1633-41.

[46] Fenwick J, Needleman IG, Moles DR. The effect of bias on the magnitude of clinical outcomes in periodontology: a pilot study. J Clin Periodontol 2008; 35:775-82.

[47] Van Tulder MW, Suttorp M, Morton S, Bouter LM, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine 2009; 34:1685-92.

[48] Pandis N. The effect size. Am J Orthod Dentofacial Orthop 2012; 142:739-40.

[49] 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; 336:601-5.

[50] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273:408-12.

[51] Haahr MT, Hrobjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. Clin Trials 2006; 3:360-5.

[52] Devereaux PJ, Choi PTL, El-Dika S, Bhandari M, Montori VM, Schünemann HJ, et al. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. J Clin Epidemiol 2004; 57:1232-6.

[53] Kahan BC, Rehal S, Cro S. Blinded outcome assessment was infrequently used and poorly reported in open trials. PloS One 2015; 10:e0131926.

[54] Karanicolas PJ, Farrokhyar F, Bhandari M. Practical tips for surgical research: blinding: who, what, when, why, how? Can J Surg 2010; 53:345-8.

[55] Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, et al. Need for expertise based randomised controlled trials. BMJ 2005; 330:88.

[56] De Vet HC, De Bie RA, Van Der Heijden GJMG, Verhagen AP, Sijpkes P, Knipschild PG. Systematic reviews on the basis of methodological criteria. Physiotherapy 1997; 83:284-9.

[57] Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 1998; 51:1235-41.

[58] Van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. Spine 2003; 28:1290-9.

[59] Van Tulder MW, Assendelft WJJ, Koes BW, Bouter LM, Bombardier C, Nachemson AL, et al. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for spinal disorders. Spine 1997; 22:2323-30.

[60] Moseley AM, Herbert RD, Sherrington C, Maher CG. Evidence for physiotherapy practice: a survey of the Physiotherapy Evidence Database (PEDro). Aust J Physiother 2002; 48:43-9.

[61] Sherrington C, Herbert RD, Maher CG, Moseley AM. PEDro. A database of randomized trials and systematic reviews in physiotherapy. Man Ther 2000; 5:223-6.

[62] Bizzini M, Childs JD, Piva SR. Systematic review of the quality of randomized controlled trials for patellofemoral pain syndrome. J Orthop Sports Phys Ther 2003; 33:4-20.

Chapter 8

Influence of performance bias, reporting bias, and attrition bias on treatment effect size estimates in randomized clinical trials of oral health interventions

8.1. Background

In the evidence-based practice approach, in order for the outcomes of randomized clinical trials (RCTs) to be generalizable to patient subsets, they need to be properly conducted and reported to a standard that enables the implementation of their results to improve existing treatments and procedures [1, 2]. Because RCTs are considered the highest grade of scientific evidence for the assessment of treatment intervention efficacy, recommendations that stem from these trials are used to guide policy decisions and clinical practice.

Methodological quality of reported RCTs [3-6] has uncovered various sources of bias (e.g., performance bias, attrition bias, and reporting bias). Performance bias arises due to differences between study groups with respect to delivered interventions and other related factors (e.g., dissimilarity of cointerventions and inadequacy of compliance to treatment protocol). Attrition bias occurs due to differences between study groups with respect to dropout from a trial that leads to "incomplete outcome data" [7]. Reporting bias occurs when selective outcome reporting (i.e., preferential reporting of beneficial findings) [7] occurs. Other sources of bias can be associated with "early" stopping of a trial for benefit (i.e., stopping a trial on the basis of finding beneficial treatment effect size estimates, rather than finalizing it as intended [8]), and "funding bias," or the inappropriate influence of funding on trial findings [7]. These biases have been found to influence the magnitude of treatment effect estimates in RCTs and can skew the overall conclusions of meta-analyses [2, 8, 9], leading to inappropriate treatment decisions

[10, 11].

Meta-epidemiological studies that evaluate the extent of bias in the treatment effect size estimate in a series of meta-analyses have shown a clear association between trial quality and treatment effect size estimate [12-14]. Poor methodological quality typically results in an exaggeration of the treatment effect size estimate [9, 15, 16], and manifests as poor compliance with treatment protocol [9], baseline imbalance [17], presence of cointerventions [18], and early trial stoppage [17]. Also, although industrial funding of a trial is not considered a methodological quality criterion, it was found to be associated with overestimation of treatment effect size [19]. Such misleading information can cause dental clinicians to make inappropriate clinical decisions, leading to ineffective treatment and poor outcomes [20]. However, not all the studies confirmed these associations [13, 17, 18].

Few meta-epidemiological studies address these factors within the field of medicine [9, 11-13, 21, 22], and some of the studies' conclusions may not be directly applicable to other healthcare fields. The assessment of reporting quality rather than methodologic quality [11, 12, 22], the placement of emphasis on only a few medical specialties, the limitation of assessment to a narrow set of quality items (generally sequence generation, allocation concealment, and blinding) [23], and the examination of dichotomous rather than continuous outcomes [9, 11, 12] has led to the emergence of inconsistent methodological findings associated with treatment effect size estimates [9, 11].The extent of bias in treatment effect size estimates associated with the aforementioned methodological deficiencies varies across different types of intervention [12, 15] research in medical and health research fields [18].

No meta-epidemiological study yet has evaluated the influence of performance bias, reporting bias, and attrition bias on treatment effect size estimates within the field of dentistry, nor in its nine subspecialties. Importantly, it is not clear

whether the previously mentioned conclusions from other heath areas hold true in the field of oral health research, given the many unique design characteristics employed in oral health research, such as: difficulty in blinding, the use of a broad spectrum of intervention modalities (surgical, nonsurgical, drug, nondrug) [24], various outcome measurements, split-mouth and crossover designs, and clustering effects [25, 26]. These unique design characteristics make RCTs in the field of oral health more challenging than RCTs in other healthcare fields.

Here, the following study attributes were examined to determine whether bias might have influenced treatment effect size estimates in RCTs of oral health interventions:

(1) methodological study characteristics such as performance bias (e.g., dissimilarity of cointerventions, indequacy of compliance to treatment protocol), attrition bias (e.g., the occurance of withdrawals in an analysis based on the intention-to-treat approach), and reporting bias (e.g., incomplete outcome reporting and inappropriate influence of funders);

(2) nonmethodological study characteristics (e.g., dental specialty, type of treatment, type of outcome [objective vs. subjective], heterogeneity of meta-analysis).

This work will impact the methodological quality of future oral health trials by providing new and valuable insights into bias in RCTs of oral health interventions. It could further assist guideline developers and policymakers in making informed decisions about the implementation of dental interventions.

8.2. Methods

This study was part of a large meta-epidemiological study that measured the impact of bias on treatment effect size estimates in RCTs of oral health interventions.

Chapters 6 and 7 present details of the methodology implemented in this study, including the study sample, data extraction, sample size calculation, and data analysis. Briefly, we conducted comprehensive searches of the literature in six electronic databases, from database inception to May 2014. We screened the titles and abstracts retrieved by the search strategy and determined the eligibility of the full texts. We included meta-analyses that examined a therapeutic intervention and included a minimum of five RCTs. We subsequently selected RCTs included in the selected meta-analyses in which the comparison was between an intervention versus a placebo, a no treatment control, or standard care.

A panel of five reviewers from diverse health research areas carried out the data extraction in duplicate, and two assessors independently carried out the data extraction. We extracted details from each included RCT and meta-analysis, for the primary outcome of the review, on the following elements were extracted: mean, standard deviation, sample size, publication year, dental specialty, primary outcome assessed, type of comparison in a review, number of included trials in a review, trial design, type of outcome in a trial, and number of centers in a trial. In chapters 6 and 7 we report measurements of bias associated with selection bias (sequence generation, allocation concealment, and baseline imbalance) and nine blinding-based criteria. In this chapter we employed a set of the remaining quality criteria that covered the following three types of bias [7, 27-29] (see **Table 8.1**):

1) Performance bias (compliance bias) can be avoided under the following conditions: similarity of cointerventions (i.e., interventions other than the trial intervention are similar to each other) and adequacy of compliance to treatment protocol (i.e., 80% adherence of participants to the treatment protocol and attendance at sessions as planned);

2) Reporting bias can be avoided under the following conditions: complete outcome reporting (i.e., primary and secondary outcomes presented in the methods section

are similar to outcomes reported in the results section [17]) and appropriate influence of funders;

3) Attrition bias can be avoided under the following conditions: reporting of the withdrawal/dropout rate, acceptability of the withdrawal/dropout rate (i.e., a dropout rate less than or equal to 20% is acceptable), obtaining of complete outcome data (i.e., the absence of missing data), analysis based on an intention-to-treat approach (i.e., patients are analyzed in the groups to which they were initially randomized), having less than 10% and 20% missing data when an intention-to-treat analysis is not performed.

These quality criteria were extracted from quality assessment tools that were reported to be valid and are commonly used in health research [30-37]. The criteria were selected based on preliminary work of the research team who identified criteria to assess the reporting quality and the methodological quality of the RCTs [29, 38]. Using the original tools as a guideline, definitions and methods were derived for each criterion. A three-part answering scheme (high, low, unclear [27]) was employed for two of the domains (selective outcome reporting and incomplete outcome data), and answers of "yes," "no," and "unclear" were employed for the other quality domains. **Table 8.1** provides further details of the definitions and methods used to assess the biases in the chosen RCTs.

We conducted a two-level data analysis using a meta-meta-analytic approach with a random-effects model, following guidelines established by Sterne et al. [39]. For the "within meta-analyses level," we obtained a standardized treatment effect size for the primary outcome of each trial, as outlined by Cohen [40]. For the "among meta-analyses level," we pooled findings of the previous analyses (combined differences from all meta-analyses) to describe the effect on the treatment effect size estimate of each component in each trial across all meta-analyses. We combined all treatment effect size estimates at this stage using inverse-variance

random-effects meta-analysis [41] to account for between-meta analyses heterogeneity, and calculated the DerSimonian and Laird estimate of variance [41]. To examine whether specific characteristics modified the associations, we stratified the analyses with interaction tests based on Z scores according to the following factors: type of outcome, dental speciality, magnitude of the treatment effect sizes, and heterogeneity of the meta-analysis. We performed all analyses using STATA statistical software version 14 (College Station, TX: StataCorp LP).

8.3. Results

Chapters 6 and 7 report detailed descriptions of characteristics of the selected reviews and a complete list of these reviews. Ultimately, 64 meta-analyses, including 540 randomized clinical trials, contributed to this study analysis.

8.3.1. Impact of balance in cointervention on the treatment effect size estimate

The similarity of cointervention was judged as adequate (at low risk of bias) in 40.2% of the trials (n = 217), while 59.8% (n = 323) of the trials were assessed as unclear. **Figure 8.1a** shows a forest plot of the difference in treatment effect size estimate between trials with similar and imbalanced cointerventions. Thirty-two meta-analyses, including 261 trials that analyzed 31,239 patients, provided information for this meta-epidemiological analysis. The analyses indicated that there was no statistically significant difference between treatment effect size estimates in trials with similarity or imbalances in cointerventions (difference in treatment effect size estimate = 0.08, at 95% confidence interval: -0.11 to 0.27; p = 0.417). A positive value (> 0) across meta-analyses would have indicated that the imbalance in cointerventions overestimated the treatment effect size. The results of the stratified analyses show that none of the meta-analyses characteristics had a statistically significant interaction with the treatment effect size estimate (**Figure 8.1b)**.

8.3.2. Impact of adequacy of compliance to treatment protocol on the treatment effect size estimate

Patient compliance to the treatment was judged as being at low risk of bias in 53.5% of the trials (n = 289), while 46.4% (n = 250) of the trials were assessed as unclear. **Figure 8.2a** displays a forest plot of the difference in treatment effect size estimate between trials with adequate and inadequate/unclear compliance to treatment protocol. Thirty-two meta-analyses, including 280 trials that analyzed 84,819 patients, provided information for this meta-epidemiological analysis. Although inadequacy of patient compliance to treatment was not associated with statistically significant differences in treatment effect size estimate, these trials tended to overestimate treatment effect size compared to trials with adequate compliance to treatment (difference in treatment effect size estimate = 0.10, at 95% confidence interval: -0.02 to 0.22; p = 0.114). A positive value (> 0) across meta-analyses would have indicated that inadequacy in patient treatment protocol compliance caused overestimates of the treatment effect size. The results of the stratified analyses show that none of the meta-analyses characteristics had a statistically significant interaction with the treatment effect size estimate (**Figure 8.2b)**.

8.3.3. Impact of selective or incomplete outcome reporting on the treatment effect size estimate

Incomplete outcome reporting was judged as being at low risk of bias in 96.1% (n = 519) of the trials, while 3% (n = 16) of the trials were judged as having an unclear risk of bias. **Figure 8.3a** displays a forest plot of the difference in treatment effect size estimate between trials with complete and incomplete outcome reporting. Fourteen meta-analyses, including 170 trials that analyzed 54,570 patients, provided information for this meta-epidemiological analysis. Incomplete outcome

reporting was not associated with a statistically significant difference in treatment effect size estimate (difference in treatment effect size estimate = 0.00, at 95% confidence interval: -0.28 to 0.29; p = 0.989). A positive value (> 0) across meta-analyses would have indicated that incomplete outcome reporting overestimated the treatment effect size. Results of the impact of complete outcome reporting on the treatment effect size estimate stratified by characteristics of meta-analyses show that none of the meta-analyses characteristics had a statistically significant interaction with the treatment effect size estimate (**Figure 8.3b**).

8.3.4. Impact of inappropriate influence of funding on the treatment effect size estimate

The influence of the trial sponsor was assessed as being unclear in 72.8% of the trials (n = 393), while it was assessed as appropriate in 16.7% (n = 90) of the trials. **Figure 8.4a** displays a forest plot of the difference in treatment effect size estimate between trials with appropriate and inappropriate influence of funders. Thirty-seven meta-analyses, including 328 trials that analyzed 85,934 patients, provided information for this meta-epidemiological analysis. The analyses revealed that trials with inappropriate influence of funders had significantly larger treatment effect size estimates; treatment effect size estimtes were 0.10 larger in trials with inappropriate influence of funders than in trials with appropriate influence of funders (95% confidence interval: 0.02 to 0.19; p = 0.017). A positive value (> 0) across meta-analyses indicated that a lack of appropriate influence of funders inflated the treatment effect size estimate.

The results of the stratified analyses showed that the difference in treatment effect size estimate between trials with appropriate and inappropriate influence of funders was significant (p = 0.02) in meta-analyses with a high level of heterogeneity between trials, but not in meta-analyses with a low level of heterogeneity between

trials. However, none of the other characteristics of meta-analyses had a statistically significant interaction with the treatment effect size estimate (**Figure 8.4b**).

8.3.5. Impact of reporting the withdrawal/dropout rate on the treatment effect size estimate

Withdrawals/dropouts were reported in the majority of trials (88.9%, n = 480), while 60 trials (11.1%) failed to report withdrawals/dropouts. Twenty-six meta-analyses, including 271 trials that analyzed 75,307 patients, provided information for this meta-epidemiological analysis. **Figure 8.5a** displays a forest plot of the difference in treatment effect size estimate between trials that reported dropouts and trials that failed to report dropouts. Results of the analysis showed that trials that failed to report withdrawals/dropouts had significantly larger treatment effect size estimates; treatment effect size estimates were 0.24 larger in trials that did not report withdrawals/dropouts than in trials that reported withdrawals/dropouts (at 95% confidence interval: 0.05 to 0.43; $p = 0.013$). A positive value (> 0) across meta-analyses indicates that a lack of appropriate influence of funding inflated the treatment effect size estimate.

The results of the stratified analyses show that the difference in treatment effect size estimate between trials that reported dropouts and trials that failed to report dropouts was significant ($p = 0.03$) in meta-analyses with a high level of heterogeneity between trials, but not in meta-analyses with a low level of heterogeneity between trials. Interestingly, the difference in treatment effect size estimate between trials that failed to report withdrawals/dropouts and trials that did report withdrawals/dropouts was also found to be significant ($p < 0.001$) in meta-analyses of periodontal trials, but not in meta-analyses of other dental interventions. However, type of outcome and treatment benefit in overall meta-analysis did not have a statistically significant interaction with the treatment effect size estimate

(**Figure 8.5b**).

8.3.6. Impact of an acceptable withdrawal/dropout rate on the treatment effect size estimate

The rate of withdrawal/dropout was judged to be acceptable (≤ 20%) in 73.1% of the trials (n = 395), while it was unacceptable (>20%) in 17.2% of trials (n = 93) (**Figure 8.6a)**. Results obtained from 41 meta-analyses, including 387 trials that analyzed 123,172 patients, showed that a withdrawal/dropout rate of ≤ 20% was not associated with a statistically significant difference in treatment effect size estimate though, trials with higher withdrawal/dropout rate tended to inflate the treatment effect size estimate (difference in treatment effect size estimate = 0.04, at 95% confidence interval: -0.06 to 0.14; p = 0.421). A positive value (> 0) across meta-analyses indicates that an unacceptable withdrawal/dropout rate inflated the treatment effect size estimate.

The results of the stratified analyses showed that the difference in treatment effect size estimate between trials with acceptable and unacceptable withdrawal/dropout rates was significant (p = 0.01) in meta-analyses with a high level of heterogeneity between trials, but not in meta-analyses with a low level of heterogeneity between trials. Also, the difference in treatment effect size estimate between trials with higher withdrawal/dropout rates compared to trials with acceptable/lower withdrawal/dropout rates was found to be significant (p < 0.001) in meta-analyses of periodontal trials, but not in meta-analyses of other dental interventions. However, type of outcome and treatment benefit in overall meta-analysis did not have a statistically significant interaction with the treatment effect size estimate (**Figure 8.6b**).

8.3.7. Impact of complete/incomplete outcome data on the treatment effect size estimate

The presence of "complete outcome data" was judged as being unclear in 28.1% (n = 152) of the trials, while 17.2% (n = 93) of the trials were judged as having a high risk of bias and 54.6% (n = 295) were judged as having a low risk of bias. Forty-nine meta-analyses, including 401 trials that analyzed 86,072 patients, provided information for this meta-epidemiological analysis. The analysis indicated that trials with incomplete outcome data did not have treatment effect size estimates that were significantly different from trials that provided complete outcome data (difference in treatment effect size estimate = -0.01, at 95% confidence interval: -0.16 to 0.14; p = 0.902) (**Figure 8.7a**). A negative value (< 0) across meta-analyses indicated that trials that reported complete outcome data had inflated treatment effect size estimates compared with trials with incomplete outcome data. The results of the stratified analyses show that none of the meta-analyses characteristics had a statistically significant interaction with the treatment effect size estimate (**Figure 8.7b)**.

We carried out a supplementary analysis after considering complete versus incomplete outcome reporting (removing trials with unclear obtaining of outcome reporting), given that a considerable number of trials judged as having a high risk of bias (n = 93) or low risk of bias (n = 295) were identified in this domain. Results of the analyses (22 meta-analyses, including 216 trials that analyzed 6 8,892 patients) showed that trials that lacked complete outcome data had larger treatment effect size estimates (**Figure 8.8a**) than trials that reported complete outcome data. While the difference was not statistically significant, treatment effect size estimates were 0.26 larger in trials that reported complete outcome data than in trials that did not (at 95% confidence interval: -0.06 to 0.58; p = 0.106). The results of the stratified analyses showed that differences in treatment effect size estimate between trials with complete versus incomplete outcome data were statistically significant in meta-analyses with a high level of heterogeneity between trials (p = 0.01), but not in meta-analyses with a low level of heterogeneity between trials; in meta-analyses with a

large treatment benefit in overall meta-analysis ($p < 0.001$), but not in meta-analyses with a small treatment benefit; and in meta-analyses that examined objective outcome ($p = 0.02$), but not in meta-analyses that examined subjective outcome. However, dental specialty did not have a statistically significant interaction with the treatment effect size estimate (**Figure 8.8b**).

8.3.8. Impact of missing data in the absence of an intention-to-treat analysis on the treatment effect size estimate

Thirty-nine meta-analyses, including 296 trials that analyzed 21,725 patients, provided information for this meta-epidemiological analysis. The results showed that trials that performed analyses based on an intention-to-treat approach (including those from trials with no missing data) did not have a statistically significant difference in treatment effect size estimate when compared with trials that did not perform an intention-to-treat analysis (difference in treatment effect size estimate = -0.01, at 95% confidence interval: -0.16 to 0.14; $p = 0.873$) (**Figure 8.9a**). The results of the stratified analyses show that differences in the treatment effect size estimate were significant ($p = 0.02$) in meta-analyses that examined objective outcome, but not in meta-analyses that examined subjective outcome. However, treatment benefit in overall meta-analyses and the heterogeneity between trials in the overall meta-analyses did not have a statistically significant interaction with the treatment effect size estimate (**Figure 8.9b**).

8.3.9. Impact of having more than 10% missing data in the absence of an intention-to-treat analysis on the treatment effect size estimate

While the difference was not statistically significant, the treatment effect size estimate was 0.07 larger in trial reports that had more than 10% missing data (when data analysis based on an intention-to-treat approach was not performed) (at 95% confidence interval: -0.08 to 0.22; $p = 0.374$). For this meta-epidemiological

analysis, 49 meta-analyses, including 413 trials that analyzed 86,276 patients, provided information for the analysis (**Figure 8.10a**). The results of the stratified analyses show that none of the meta-analyses characteristics had a statistically significant interaction with the treatment effect size estimate (**Figure 8.10b)**.

8.3.10. Impact of having more than 20% missing data in the absence of an intention-to-treat analysis on the treatment effect size estimate

While similar results were found when increasing the acceptable level of missing data to ≤ 20% (40 meta-analyses, including 382 trials that analyzed 122,523 patients, provided information for this meta-epidemiological analysis), the difference in treatment effect size estimate was higher (difference in treatment effect size estimate = 0.10, at 95% confidence interval: -0.01 to 0.22; p = 0.070) than in trials that had more than 20% missing data (where data analysis based on an intention-to-treat approach was not performed). Thus, trials that had more than 20% missing data (where data analysis based on an intention-to-treat approach was not performed) tended to inflate the treatment effect size estimates of 0.10 compared with trials that had ≤ 20% missing data (**Figure 8.11a**). The results of the stratified analyses showed that the difference in treatment effect size estimate between trials that had ≤20% missing data (or were conducted with an intention-to-treat analysis) and trials with >20% missing data (and were conducted without an intention-to-treat analysis) was significant ($p < 0.001$) in meta-analyses with a high level of heterogeneity between trials but not in meta-analyses with a low level of heterogeneity between trials, and in meta-analyses of periodontal trials but not in meta-analyses of other dental interventions. However, type of outcome and treatment benefit in overall meta-analysis did not have a statistically significant interaction with the treatment effect size estimate (**Figure 8.11b**).

8.4. Discussion

8.4.1. Main findings

This study was carried out in the domain of dentistry to discern whether RCTs with appropriate conduct and RCTs with inappropriate conduct yielded different treatment effect size estimates. A sample of 64 meta-analyses, including 540 RCTs conducted in dental, oral, and craniofacial domains between 1995 and 2013 were examined. Evidence was found of significant differences in treatment effect size estimates in oral health trials, depending on the appropriateness of funder influence and the reporting of withdrawal/dropout rates. While no differences in treatment effect size estimate were identified when analyzing the impact of balance in cointerventions, the adequacy of compliance to treatment protocol, analysis based on an intention-to-treat approach, and the obtaining of complete outcome data, the current study showed a consistent trend toward an overall overestimation of the treatment effect size. In contrast, selective outcome reporting and the presence of incomplete outcome data were not associated with over- or under-estimation of the treatment effect size estimate.

8.4.2. Comparison of this study with other studies

Influence of industrial funding

There is much debate in the literature about the impact and magnitude of sponsorship bias on treatment effect size estimates. The findings in this study support several studies [19, 42-44] that described a clear influence on trial results of industry-related interventions. Bias associated with an inappropriate influence of funding was found to increase the treatment effect size estimate by an average of 0.10; this magnitude of bias may represent one fifth of the treatment effect size estimate in RCTs involving some dental specialties [45]. While the aforementioned studies described numerous scenarios where the level of sponsorship could influence the design, conduct, and reporting of a clinical trial, other studies [17, 46, 47] did not

detect a significant influence of funding on the treatment effect size estimate. Possible explanations for the contradictory results are differences in assessment of and definition of sponsorship bias, type of sponsorship (e.g., pharmacological products and financial supports to trialists), and type of trial evaluated (e.g., placebo-controlled or active control). Bias in industry-sponsored oral heath trials can potentially benefit the sponsoring company and might lead to inappropriate treatment decisions. For example, a recently published report [48] examined the influence of industry sponsorship in 41 RCTs of dental implants and found that the likelihood of implant failure in sponsored RCTs was much lower than the likelihood of implant failure in nonsponsored RCTs. On the contrary, a recent network meta-analysis [49] assessed the impact of industry sponsorship on 114 dental restorative RCTs and found that material performance rankings did not differ on the basis of sponsorship. That study concluded that the influence of industry sponsorship on RCTs of restorative dentistry was "limited."

Influence of performance bias

Although the current study showed a trend toward an overestimation of treatment effect size when analyzing the influence of the similarity of cointerventions and the adequacy of compliance to treatment protocol, the results were not statistically significant. The results of the current study are generally in line with results reported by the authors of one [13] of the two studies [9, 13] that investigated the association between treatment effect size estimates and performance bias (related to a dissimilarity of cointerventions and an inadequacy of compliance to treatment protocol). The van Tulder et al. study [13] showed no significant difference in treatment effect size estimates between studies based on potential bias related to dissimilarity of cointerventions and inadequacy of compliance to treatment protocol. Contrary to the current study and van Tulder et al. [13], Moher et al. [9] found that trials with inadequacy of compliance to treatment protocol overestimated

treatment effect size, whereas trials with dissimilarity of cointerventions underestimated treatment effect size.

Influence of attrition bias

An intention-to-treat approach is often used to minimize attrition bias in RCTs [50]. However, findings analyzed from studies that examined the association between attrition bias and the treatment effect size estimate were inconsistent in terms of the direction and magnitude of the association. This conclusion was based on the definition of attrition and the outcomes evaluated. For example, studies that defined attrition bias as the number of dropouts from the intervention, found that trials with a higher dropout level had a larger treatment effect size estimate [51]. Nuesch et al. [52] assessed 167 trials (included in 14 meta-analyses) that investigated patients with osteoarthritis using pain as an outcome. The authors concluded that the magnitude and direction of bias associated with patient exclusion was "unpredictable" and that patient exclusion frequently inflated treatment effect size estimates. Hartling et al. [17] used the Cochrane risk of bias tool to define attrition bias in 287 child-health trials as "incomplete outcome data," and found no significant difference in treatment effect size estimate between trials based on attrition bias. That study concluded that investigators of reviews should not exclude RCTs from meta-analyses on the basis of attrition bias associated with "incomplete outcome data." In the current study, a trial's analysis based on an intention-to-treat approach was not found to affect the treatment effect size estimate. However, when we carried out the meta-epidemiological analysis for trials with missing data due to a dropout rate of more than 10% and 20%, the magnitude of the difference in treatment effect size estimate increased from 0.07 with a p value of 0.374 (for more than a 10% dropout rate) to 0.10 with p value of 0.070 (for more than a 20% dropout rate). Based on this finding, the acceptable dropout rate could be around 20%, and trials with dropout rates above this level would be expected to present an increase in bias.

Future meta-epidemiological studies should use other meta-epidemiological analyses to identify an exact cut-off ceiling for acceptable dropout levels in oral health trials.

While our study did not show a statistically significant difference in treatment effect size estimate based on an acceptable dropout/withdraw rate of 20%, a failure to report the dropout/withdrawal rate was found to be associated a 0.24 increase in the treatment effect size estimate; this magnitude represented one third to one half of the treatment effect size estimate observed in RCTs of many dental interventions [45]. Thus, in trials in the field of dentistry, a report of the dropout rate trial is more important than a dropout/withdraw rate of > 20% with respect to associated bias. Importantly, a failure to report the dropout rate in a trial might be attributed to patients' response to the examined intervention, especially if the frequency of and/or the motives for dropping out vary among intervention groups [53]. Therefore, reporting the dropout rate could be considered one of the stratifying factors when synthesizing evidence from meta-analyses, especially in dentistry.

Influence of reporting bias

Hartling et al. [17] found no significant association between selective outcome reporting and the treatment effect size estimate. However, previous studies [54, 55] found larger treatment effect size estimates in trials with a presence of selective outcome reporting than in trials with an absence of selective outcome reporting.

The direction and magnitude of the difference in treatment effect size estimate found in the current study agree with the findings in Hartling et al. [17], possibly because the current study relied, when defining selective outcome reporting, on the similarity of outcomes presented in methods, compared with outcomes reported in results. Furthermore, an assessment of selective outcome reporting was performed in the current study without evaluating the published protocols of trials.

Future investigations should evaluate the published protocols of the RCTs when assessing selective outcome reporting.

8.4.3. Limitations of the study

This meta-epidemiological study provides an empirical analysis of the association between treatment effect size estimate and bias in the domain of oral health research. The study had several limitations. First, the empirical evidence was derived from published studies only which could lead to the detection of bias based on reporting, it did not evaluate the actual conduct of the trials. Second, authors of the trials were not contacted for missing data since this procedure was impracticable; for example, some data was unattainable given that a large proportion of the trials were published before the year 2005 when information regarding the corresponding author was often not provided or not up-to-date. Another potential limitation was that data extraction and analyses were based on information provided by the authors of the published reports. This approach, although widely used, limits the identification of bias when study elements are not properly reported by trial authors.

Certain levels of heterogeneity are expected in the current meta-epidemiological study and are potentially rooted in any future meta-epidemiological examination of the impact of bias on treatment effect size estimates, given that this type of study is built on numerous entities of analysis (meta-analysis, trials, interventions, outcomes, and participants) that have a potential for heterogeneity [56]. By applying a cautious methodology to data collection and analysis and by assembling a large number of meta-analyses and trials in our study, we increased the study power and minimized heterogeneity. Further, data analysis was restricted to trials where the direction of expected treatment effect size estimate was clear, including trials involving controls or placebo interventions; this procedure minimized the heterogeneity and confounding factors in the analyses and advanced our ability to detect significant effects of methodological characteristics.

Finally, the current study did not assess the influence on our study of interactions with other design biases on the magnitude of treatment effect size. Such an assessment would require a multivariate analysis with a larger number of meta-analyses and trials [23]. A meta-epidemiological assembly of a larger number of meta-analyses and trials through the synthesis of results from different disciplines and datasets could take this matter into account.

8.5. Conclusions

This study detected significantly larger treatment effect size estimates in oral health trials that performed inadequate reporting of withdrawal/dropout rates, in trials that had a > 20% dropout rate that did not perform an analysis based on an intention-to-treat approach, and in trials that presented an inappropriate influence of funding than in trials that performed adequate conduct of the aforementioned quality items. A tendency toward exaggeration of treatment effect size estimate (although not statistically significant) was identified in trials that presented imbalances in cointerventions, inadequacy of compliance to treatment, incomplete outcome data, and a > 10% dropout rate without performing an intention-to-treat approach. In contrast, trials that presented an acceptable (≤ 20%) dropout rate, selective outcome reporting, and analysis based on an intention-to-treat approach were not associated with overestimated or underestimated treatment effect size.

Table 8.1. Guidelines for quality assessment of included trials [30, 31, 34, 37, 57-61]

Items /Definitions	Yes	No	Unclear
Performance Bias (or Compliance Bias)			
Cointerventions avoided or comparable. Cointerventions are interventions other than the treatment under study.	The authors state that subjects did not receive an additional intervention, or that cointerventions were balanced between treatment and control groups. Data about cointerventions are presented and comparable between treatment and control groups.	Subjects received additional interventions besides the intervention under study. The cointerventions were not balanced between treatment and control groups.	There is insufficient information to permit a judgment. **N/A:** Treatment and control groups did not receive an intervention in addition to the intervention under study.
Adequacy of compliance to treatment protocol: Compliance acceptable in all groups (80% acceptable).	There is ≥ 80% compliance in treatment and control groups. The control group might have to be "compliant" as well. For example, in an exercise intervention, the control group would have to comply by doing no exercise.	There is less than 80% compliance in treatment and control groups.	There is insufficient information to permit a judgment.
Reporting Bias			
Complete outcome reporting	Outcomes reported in methods section need to match those reported in results section. If 0-30% of the secondary outcomes are not reported score "Yes". Main outcome has to be included in both methods and results.	≥ 70% of secondary outcomes were unreported (combining methods or results sections).	There is insufficient information to permit a judgment, or if 30%-69% of outcomes are unreported.
Appropriate influence of funders	If a sponsor is acknowledged with clear statement regarding no involvement of a sponsor in trial conduct, data management /analysis, or co-authorship, OR funding is coming from a governmental agency or foundation, OR if sponsor is acknowledged only as providing equipment or drug for the study but no one of the authors is paid by the company or the company had nothing to do with designing or analyzing the trial.	If a sponsor is acknowledged with information provided that a co-author works for that company of that company was involved in conduct of the study.	There is insufficient information to permit a judgment, or if there is no mentioning of funding source.
Attrition Bias			
Report of withdrawals /dropouts	There is clear reporting of all withdrawals and dropouts. Generally, this is done by using a flowchart. 1. Number of drop outs	There is not reporting of all withdrawals and dropouts.	There is insufficient information to permit a judgment.

	2. If no withdrawals, should be stated in article.		
Adequacy of withdrawals /dropouts: Withdrawals/dropouts rate describe and acceptable (maximum 20% drop out rate)	The withdrawals/dropouts rate is less or equal than 20%, OR with multiple time points, at any point must be at least 85% patients included in analysis.	The withdrawals /dropouts rate is >20% when only one-point time is evaluated.	There is insufficient information to permit a judgment.
Obtaining of complete outcome data	No missing outcome data (If all patients were accounted for in the analysis), OR If the numbers and reasons for withdrawal/drop-out were described and comparable across groups and the authors performed intention-to-treat approach with ≤ 20% drop outs, then score low risk of bias, OR If the numbers and reasons for withdrawal/drop-out were described and comparable across groups but they did not perform an intention-to-treat approach and the dropout rate was less or equal than 10% drop outs.	No intention-to-treat approach, or intention-to-treat approach performed with > 20% drop outs.	No Intention-to-treat but >10% and ≤ 20% drop outs.
Analysis based on the intention-to-treat approach: Patients analyzed in the groups to which they were randomized.	When authors said that they use intention-to-treat approach and according to evaluation they analyzed the subjects as randomized; OR when author have no missing data this is assumed as intention-to-treat approach if no other protocol deviations occurred.	When authors did not use intention-to-treat approach or they said that they use intention-to-treat approach but according to evaluation of the study they did not analyzed as randomized. Per protocol or "As treated" analyses are not "intention-to-treat approach".	There is insufficient information to permit a judgment.
Having more than 10% missing data (without performing data analysis based on intention-to-treat approach)	When authors said that have more than 10% missing data AND analyzed the subjects as randomized (without implementing intention-to-treat approach).	When authors said that have 10% or less missing data or analyzed data based on intention-to-treat approach.	There is insufficient information to permit a judgment.
Having more than 20% missing data (without performing data analysis based on intention-to-treat approach)	When authors said that have more than 20% missing data AND analyzed the subjects as randomized (without implementing intention-to-treat approach)	When authors said that have 10% or less missing data or analyzed data based on intention-to-treat approach.	There is insufficient information to permit a judgment.

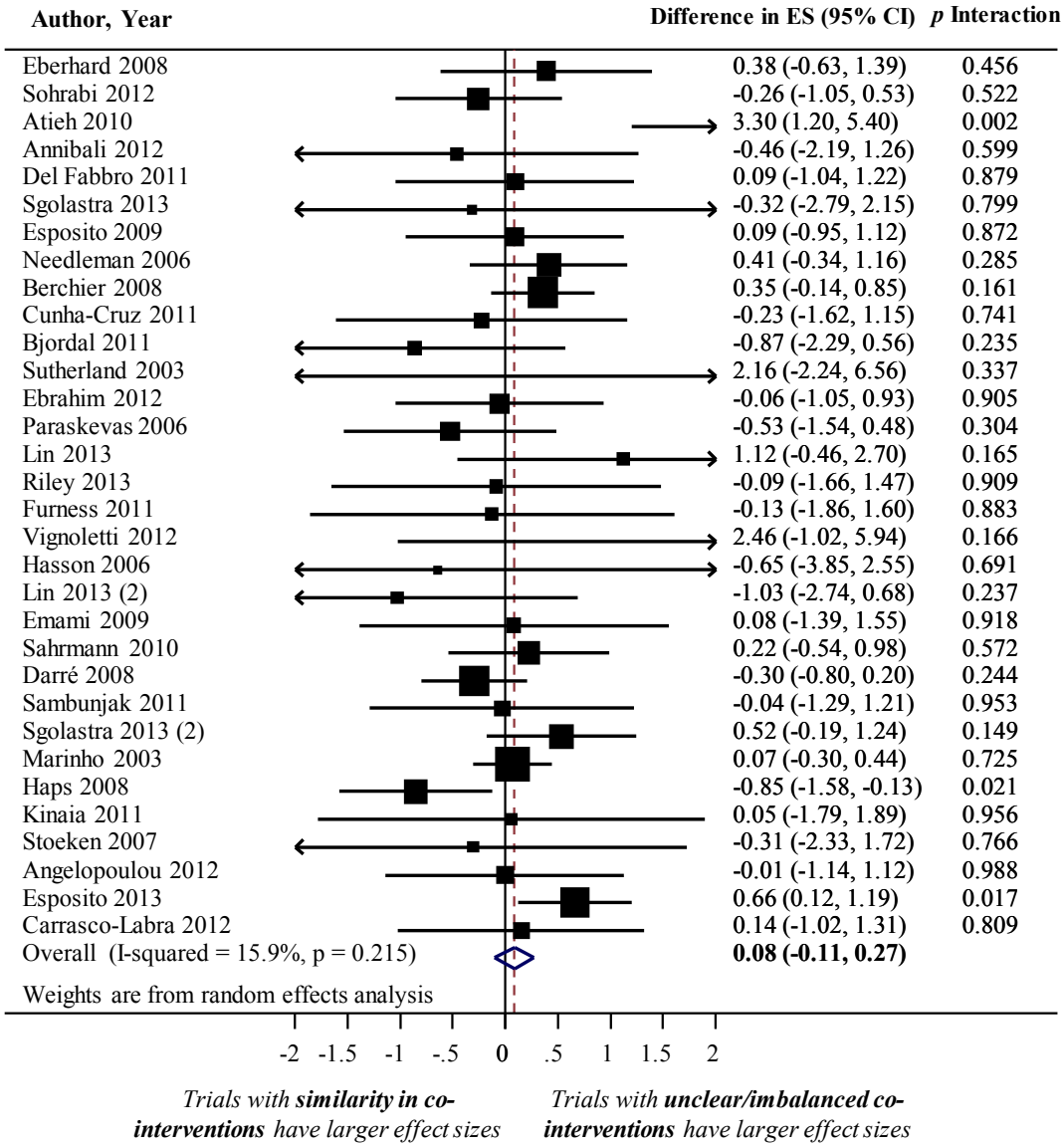

Figure 8.1a. Difference in treatment effect size (ES) estimate between trials with similar and imbalanced cointerventions. A positive value (> 0) across meta-analyses indicates that inadequacy of patient compliance to treatment protocol inflates the treatment effect size estimate.

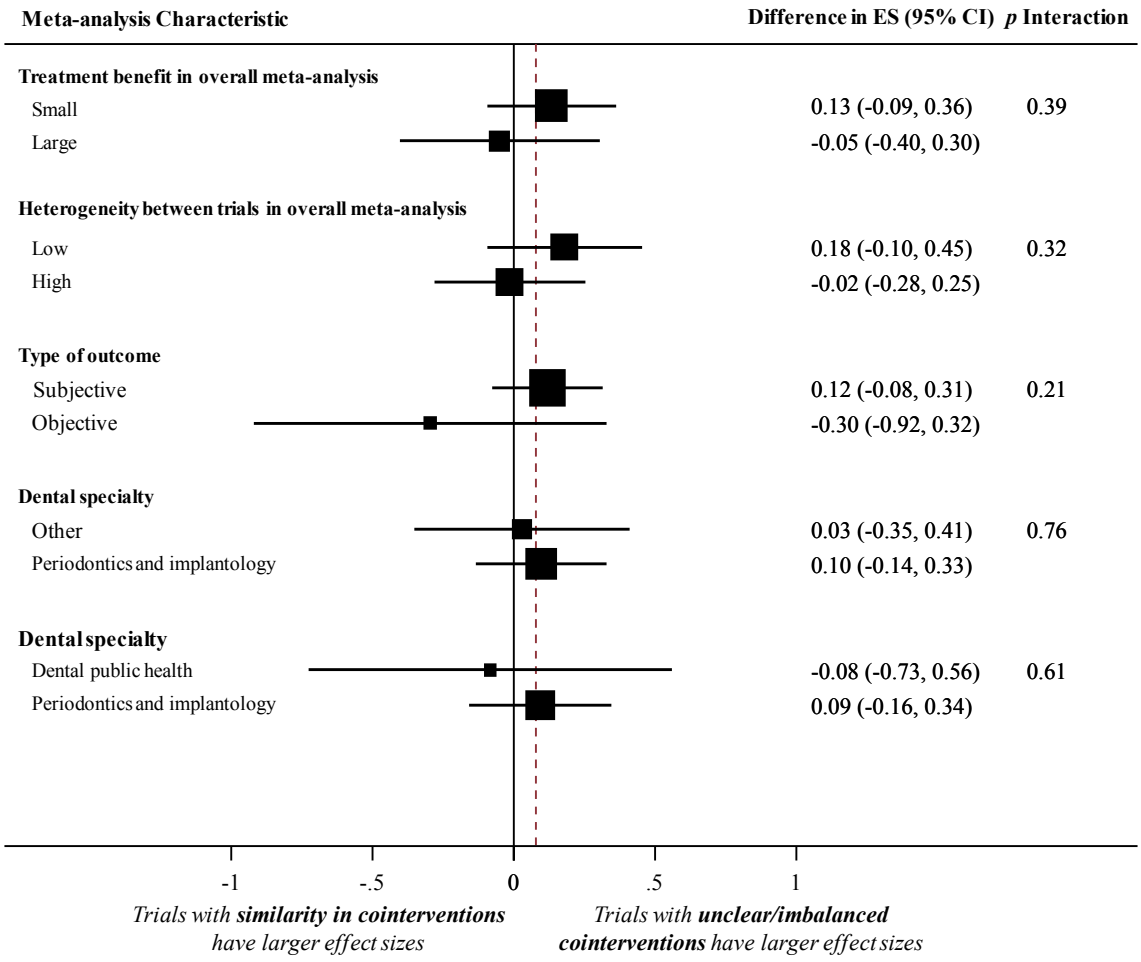

Figure 8.1b. Forest plot of the difference in treatment effect size (ES) estimate between trials with similar and imbalanced cointerventions stratified by meta-analyses characteristics.

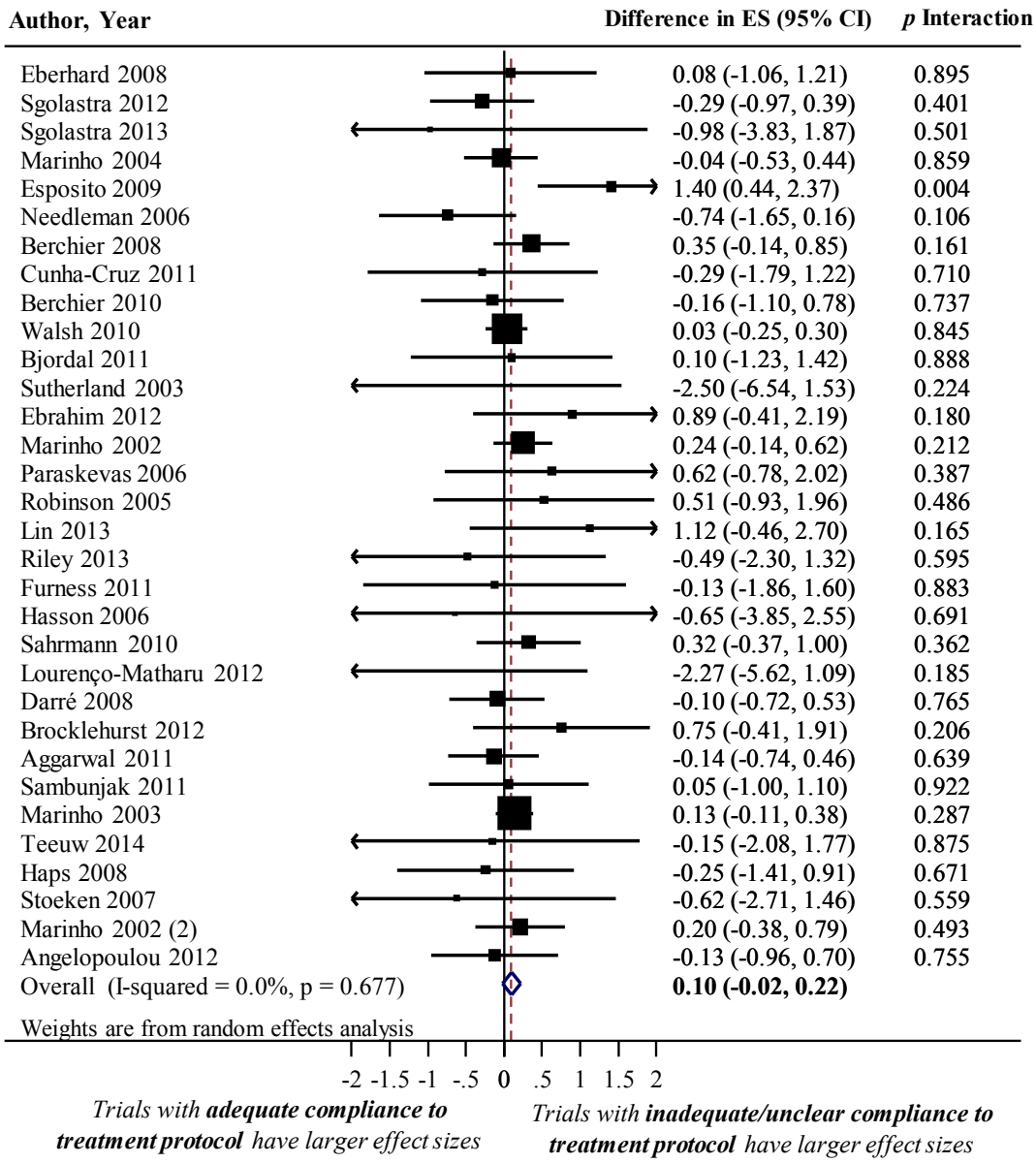

Figure 8.2a. Difference in treatment effect size (ES) estimate between trials with adequate and inadequate/unclear compliance to treatment protocol. A positive value across meta-analyses indicates that inadequacy of compliance to treatment protocol inflates the treatment effect size estimate.

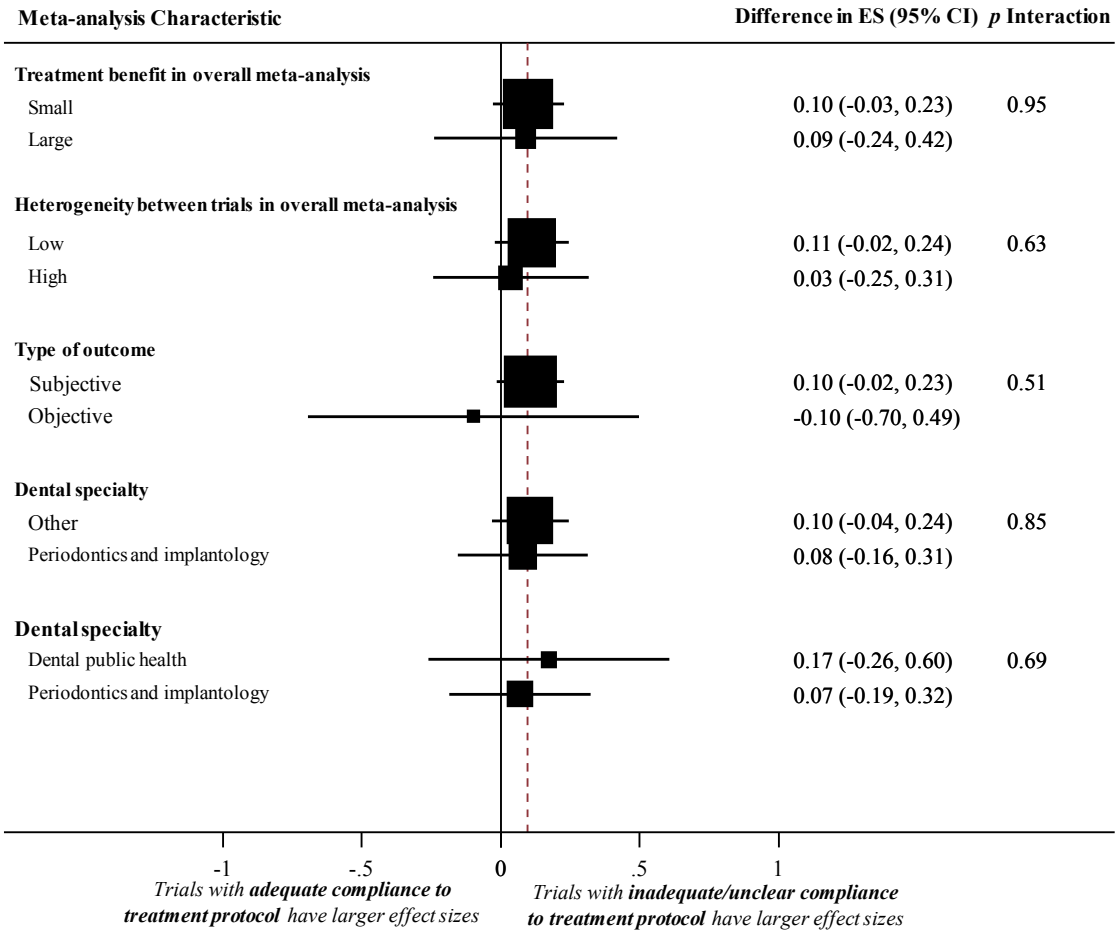

Figure 8.2b. Forest plot of the difference in treatment effect size (ES) estimate between trials with adequate and inadequate/unclear compliance to treatment protocol stratified by meta-analyses characteristics.

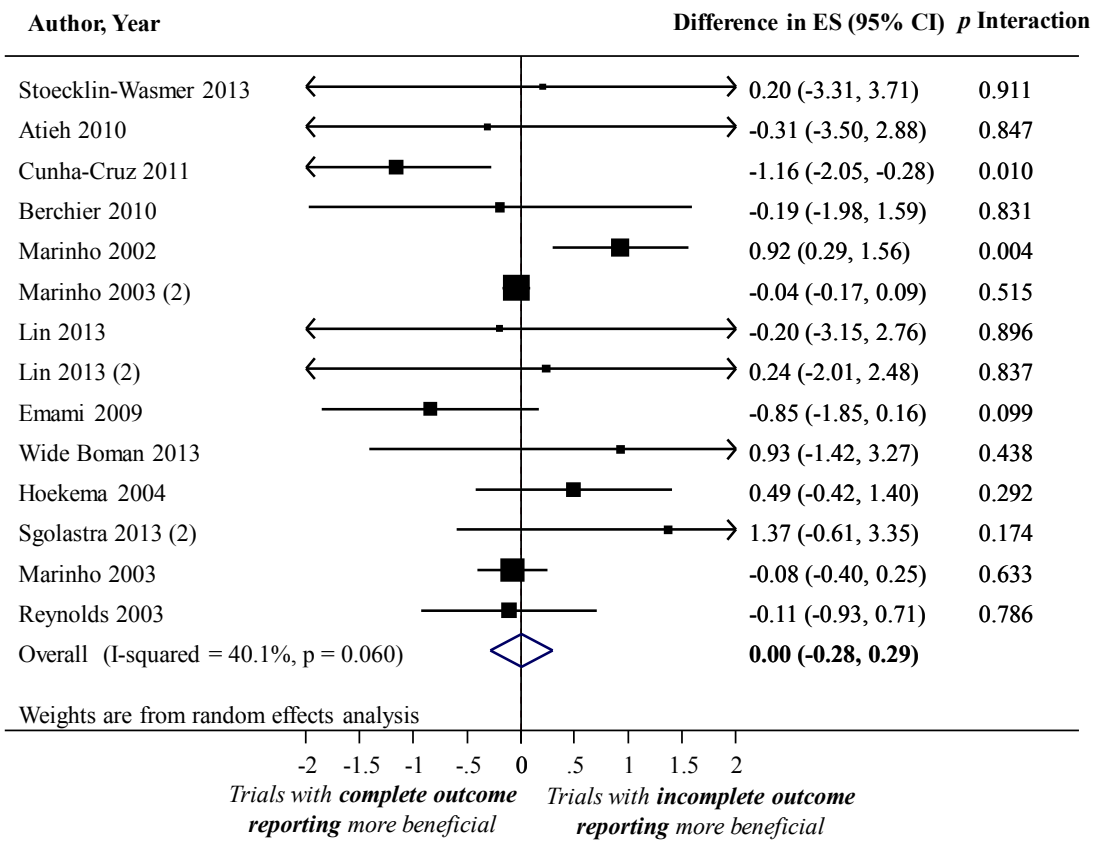

Figure 8.3a. Difference in treatment effect size (ES) estimate between trials with complete and incomplete outcome reporting. A positive value (> 0) across meta-analyses indicates that incomplete outcome reporting inflates the treatment effect size estimate.

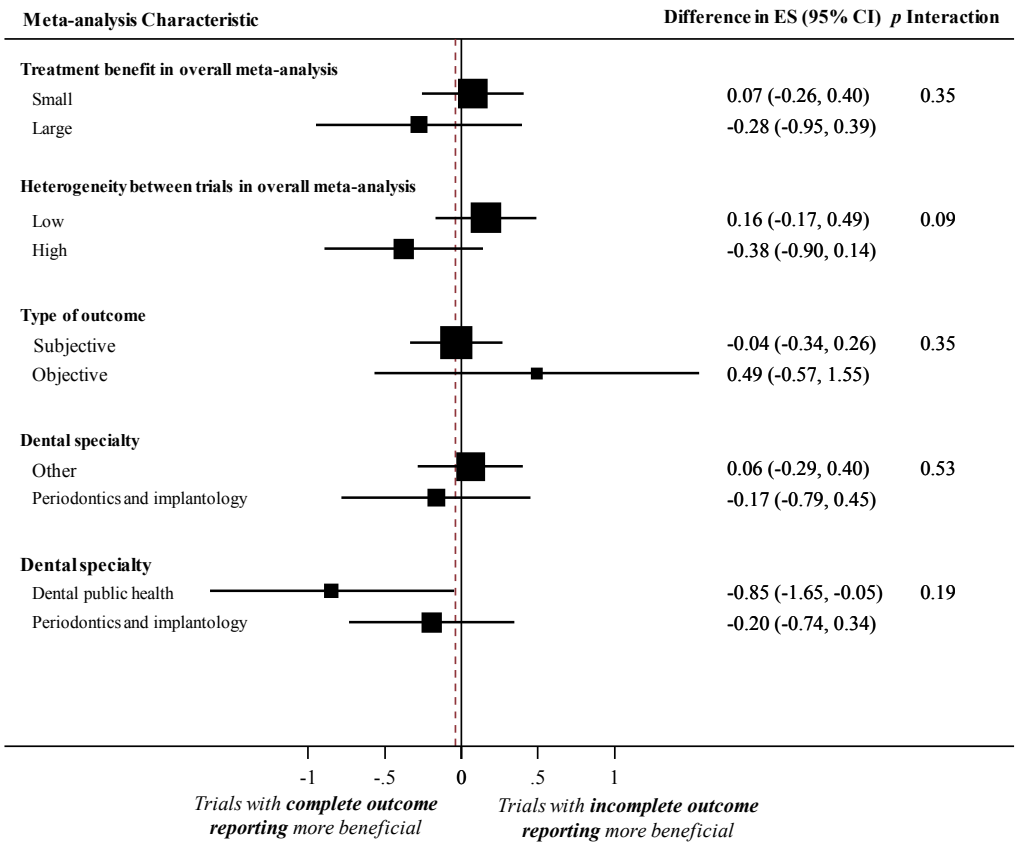

Figure 8.3b. Difference in treatment effect size (ES) estimate between trials with complete and incomplete outcome reporting stratified by meta-analyses characteristics.

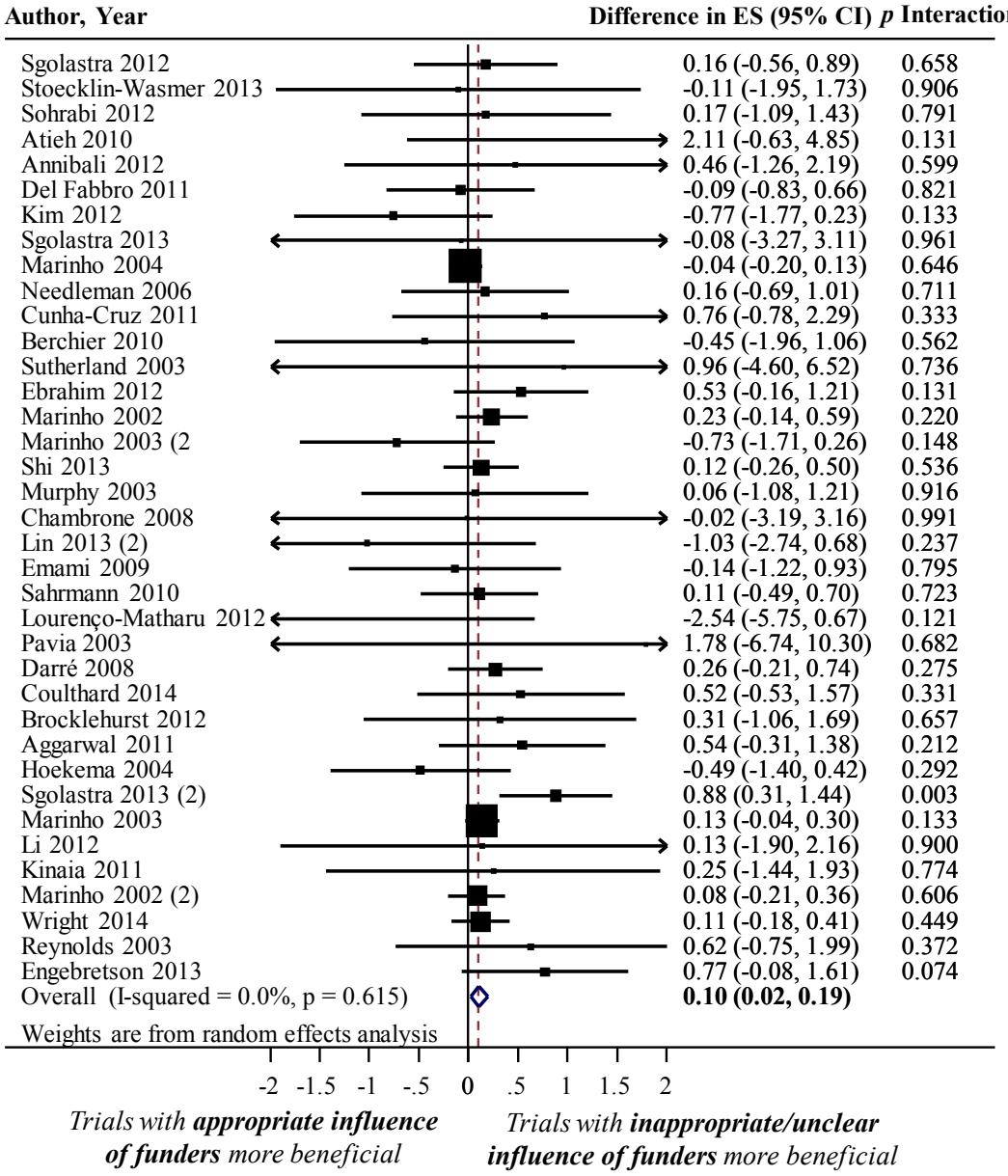

Figure 8.4a. Difference in treatment effect size (ES) estimate between trials with appropriate and inappropriate influence of funders. A positive value (> 0) across meta-analyses indicates that the inappropriate influence of funders inflates the treatment effect size estimate.

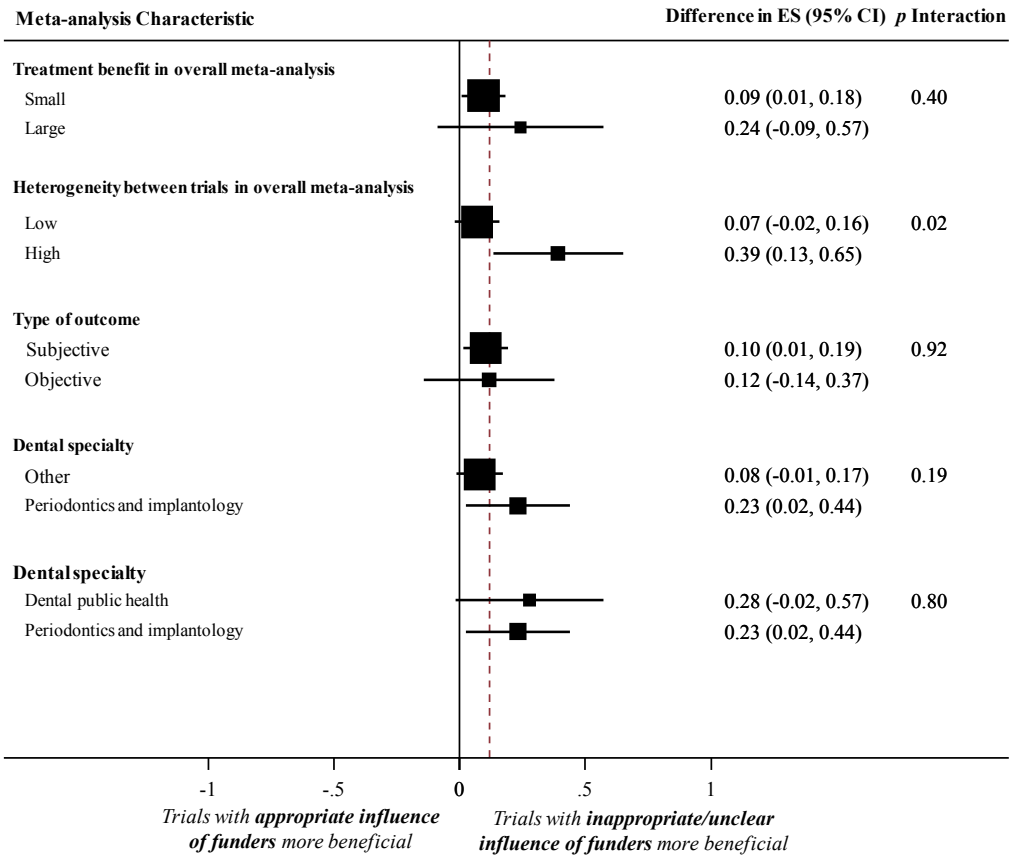

Figure 8.4b. Difference in treatment effect size (ES) estimate between trials with appropriate and inappropriate influence of funders stratified by meta-analyses characteristics.

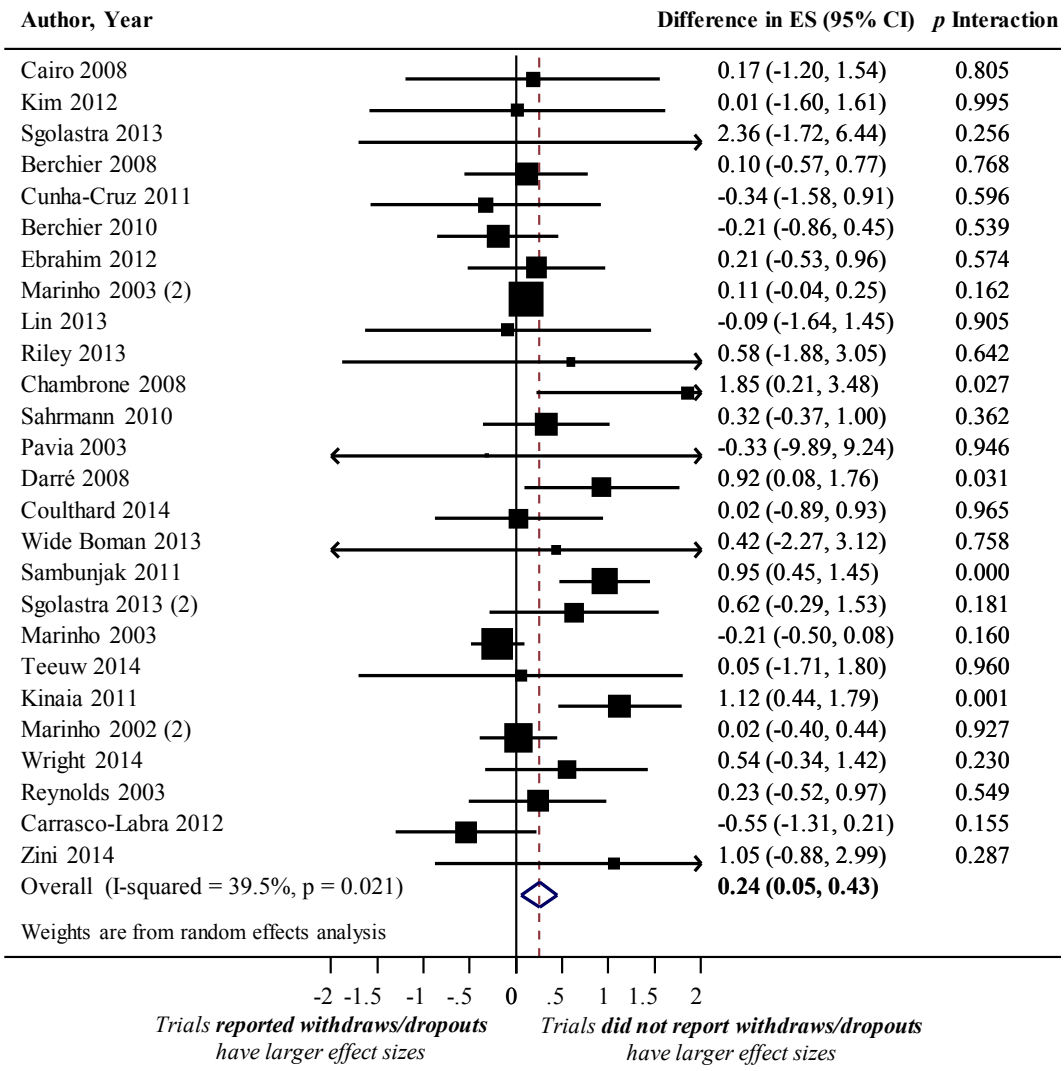

Figure 8.5a. Difference in treatment effect size (ES) estimate between trials reporting dropouts and those not reporting dropouts. A positive value (> 0) across meta-analyses indicates that the failure to report dropouts inflates the treatment effect size estimate.

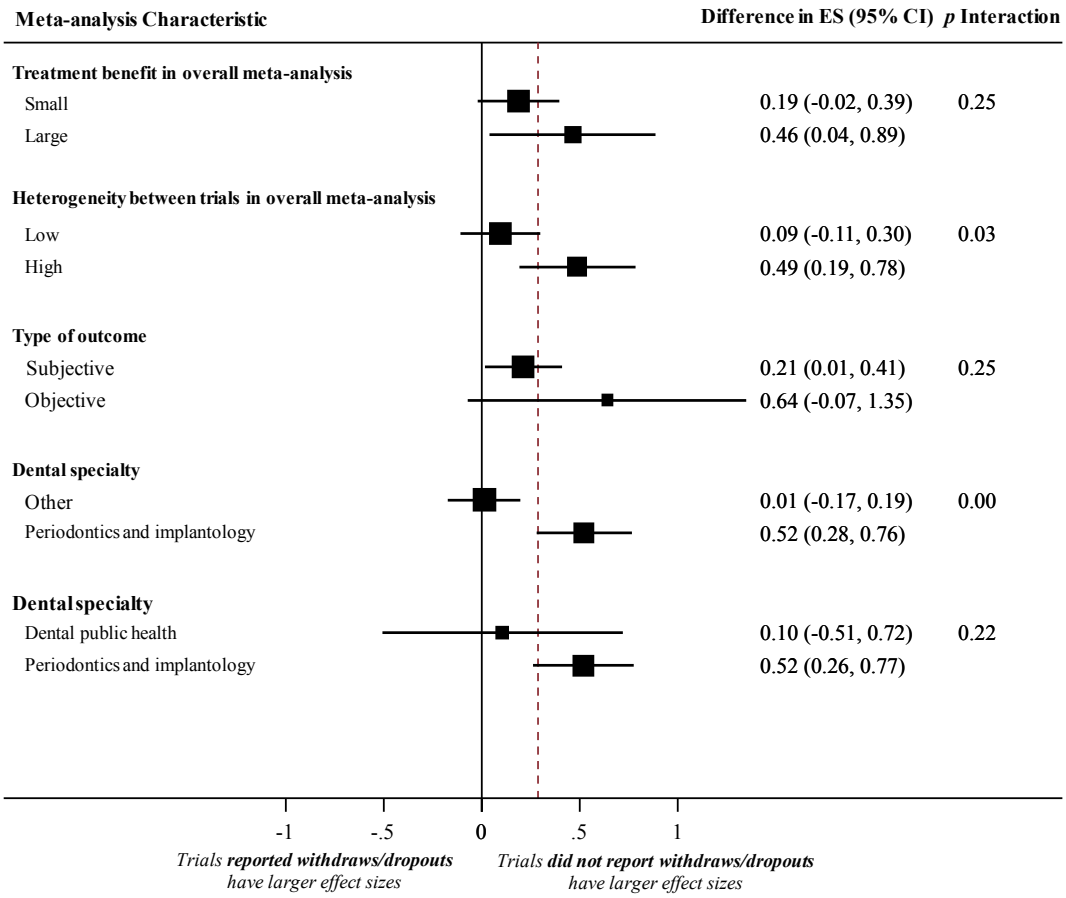

Figure 8.5b. Forest plot of the difference in treatment effect size (ES) estimate between trials reporting dropouts and those not reporting dropouts stratified by meta-analyses characteristics.

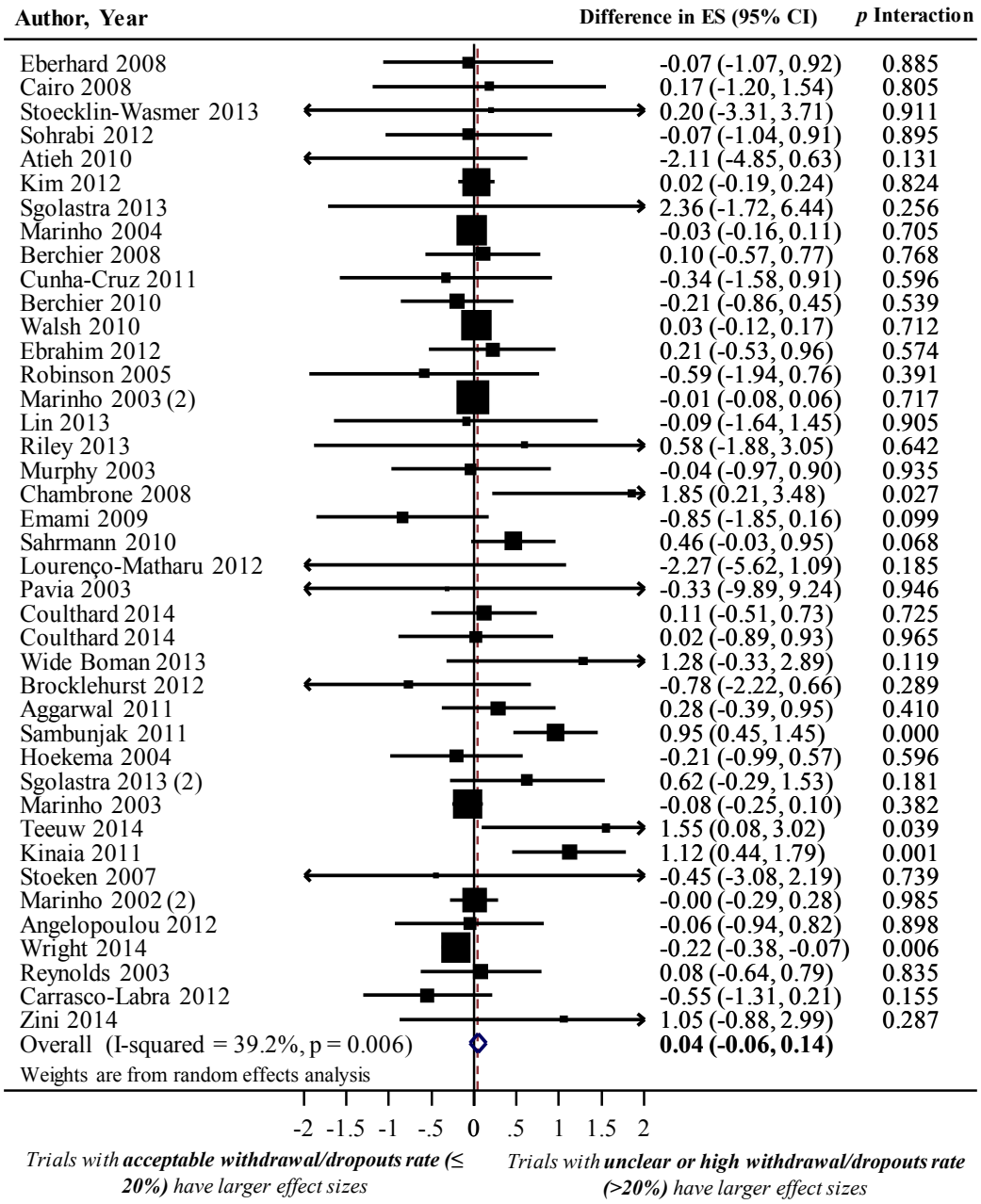

Figure 8.6a. Difference in treatment effect size (ES) estimate between trials with acceptable (≤ 20%) and high (>20%) or unclear withdrawal/dropout rate. A positive value (> 0) across meta-analyses indicates that trials with a high/unclear dropout rate inflates the treatment effect size estimate.

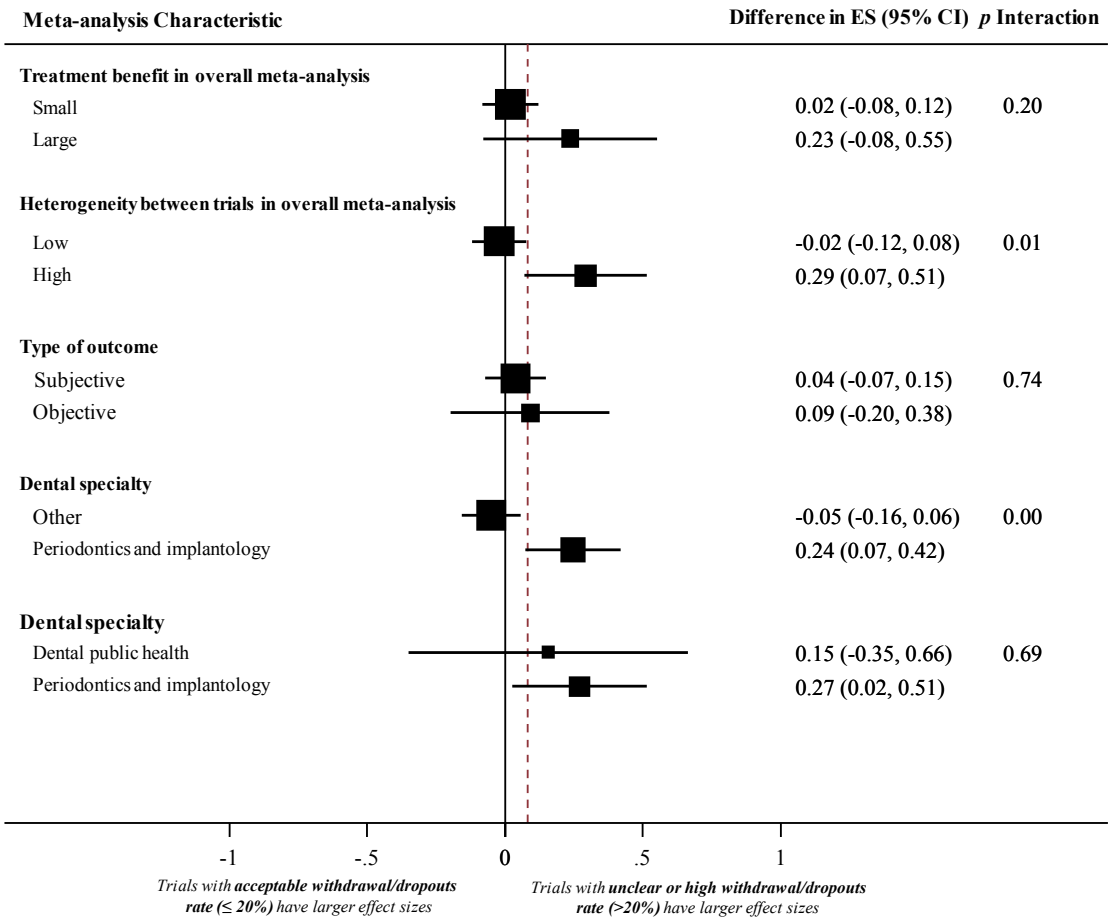

Figure 8.6b. Forest plot of the difference in treatment effect size (ES) estimate between trials with acceptable (≤20%) and high (>20%) or unclear withdrawal/dropout rate stratified by meta-analyses characteristics.

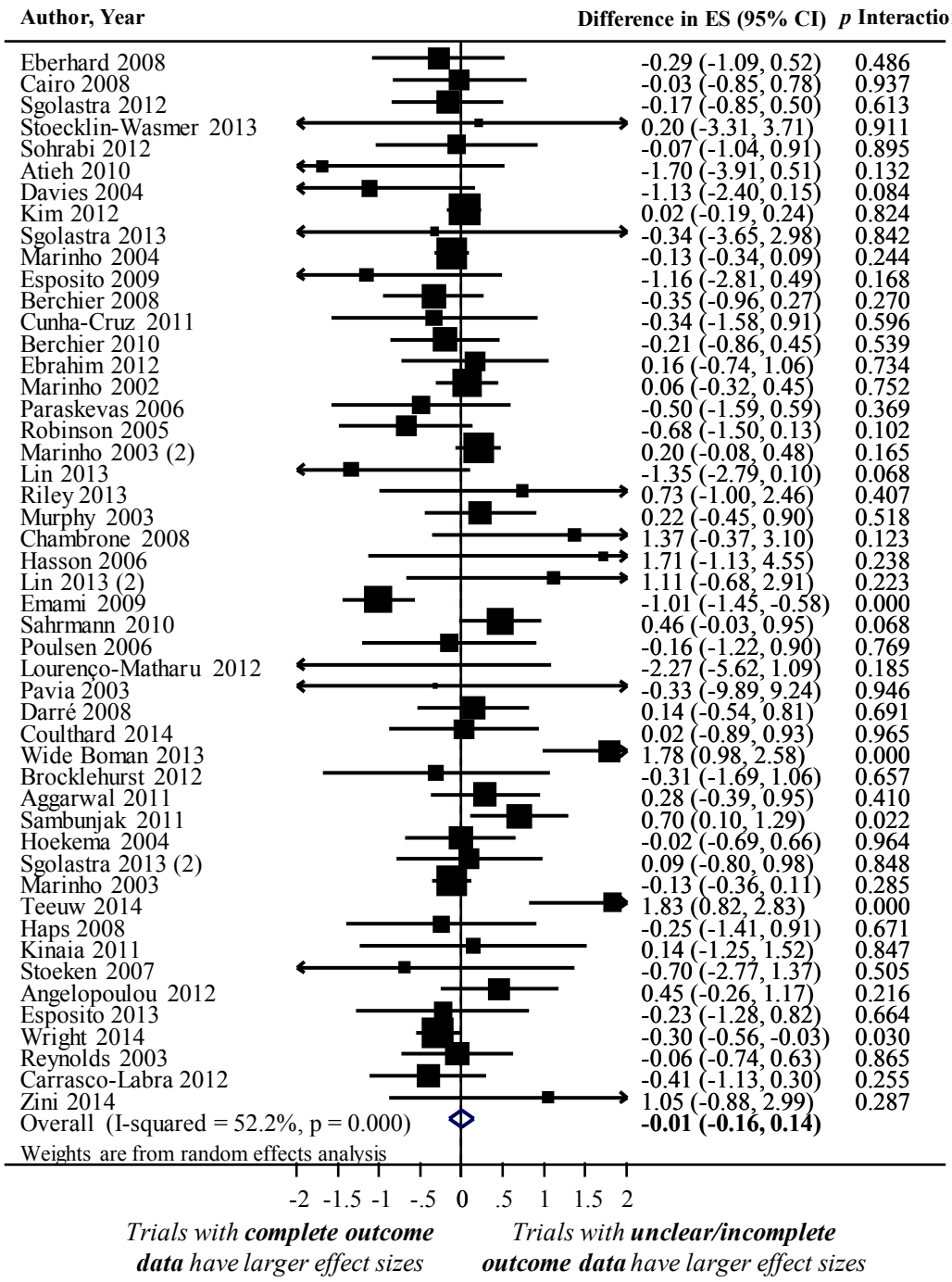

Figure 8.7a. Difference in treatment effect size (ES) estimate between trials with complete and incomplete/unclear outcome data. A positive value (> 0) across meta-analyses indicates that incomplete outcome data inflates the treatment effect size estimate.

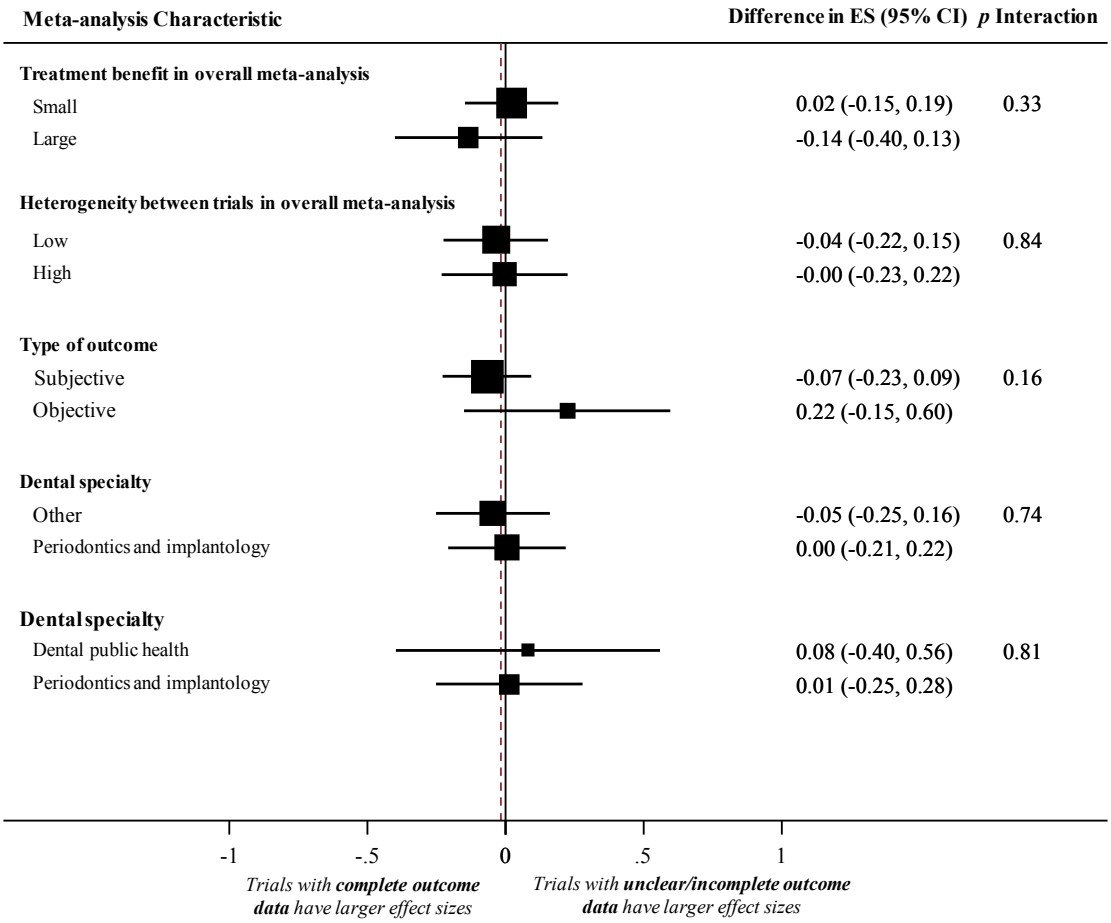

Figure 8.7b. Forest plot of the difference in treatment effect size (ES) estimate between trials with complete and incomplete/unclear outcome data stratified by meta-analyses characteristics.

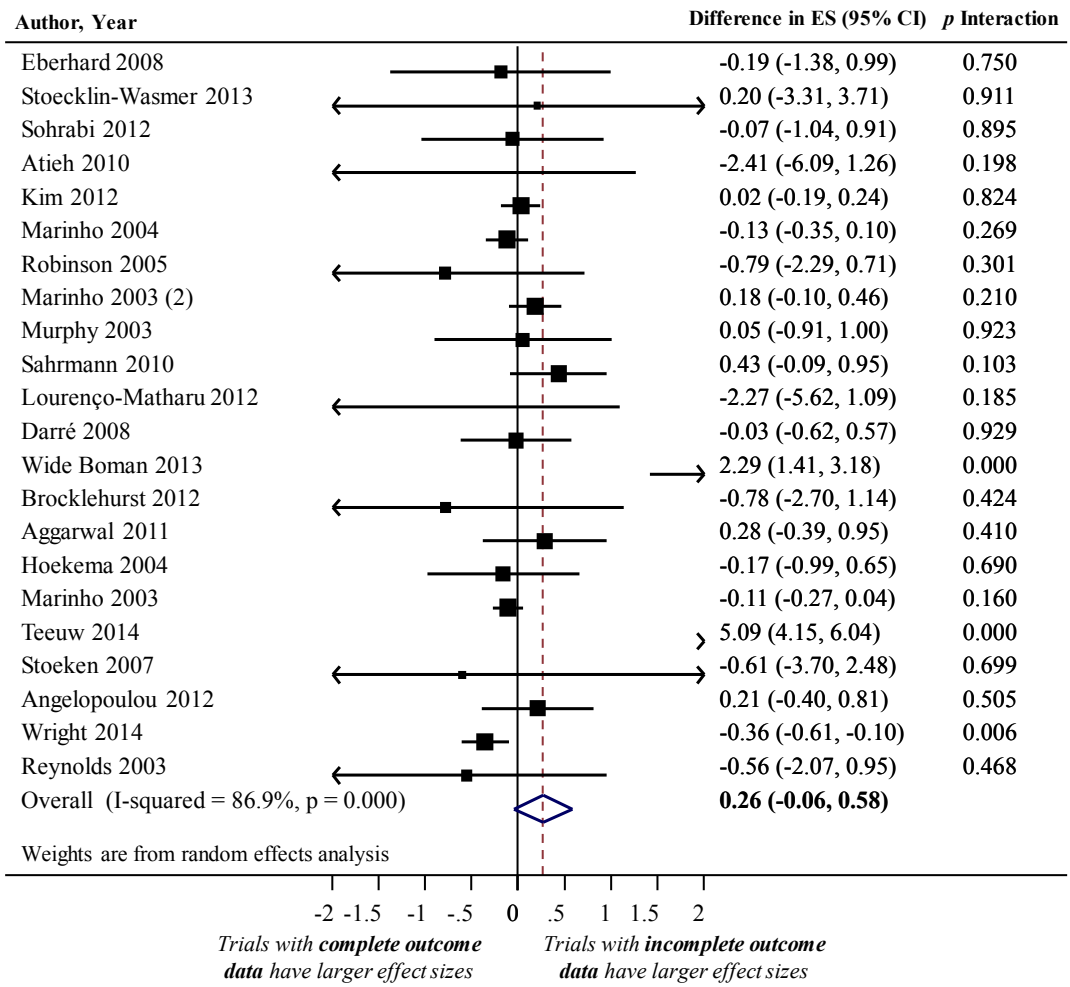

Figure 8.8a. Difference in treatment effect size (ES) estimate between trials with complete and incomplete outcome data. A positive value (> 0) across meta-analyses indicates that incomplete outcome data inflates the treatment effect size estimate.

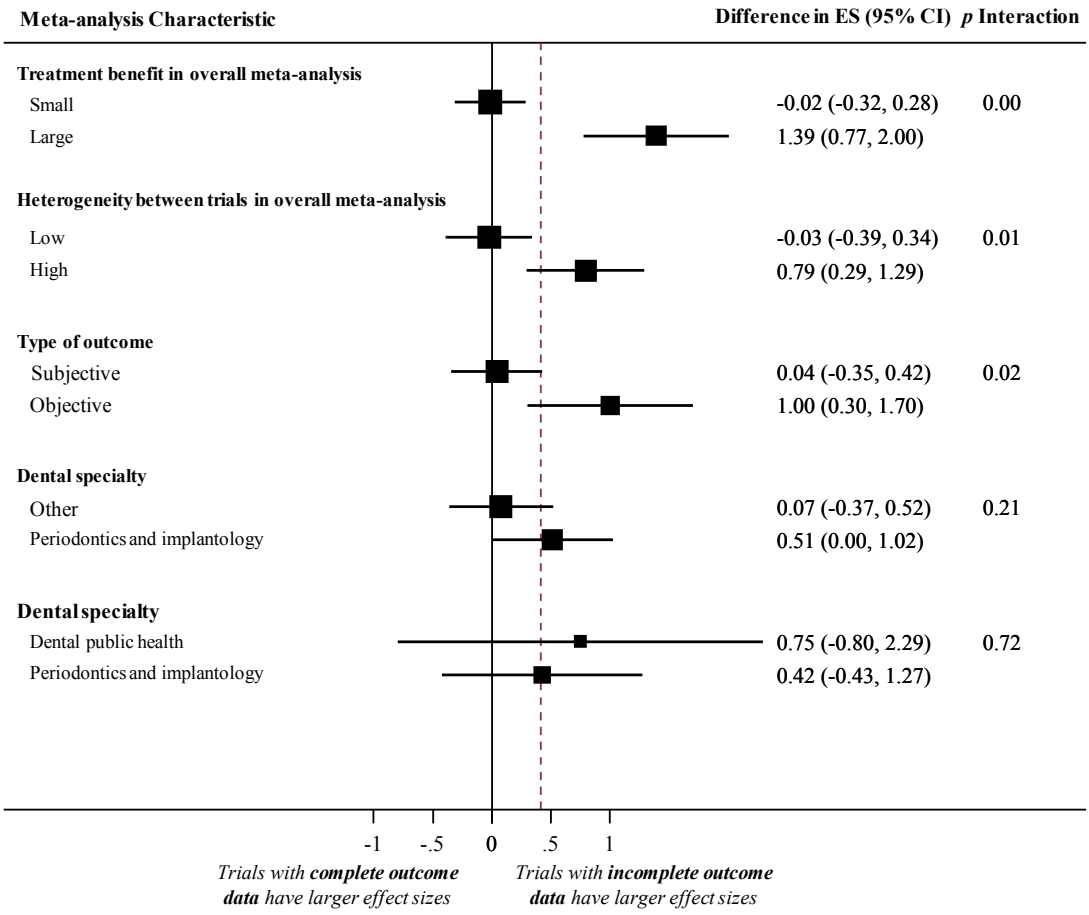

Figure 8.8b. Forest plot of the difference in treatment effect size (ES) estimate between trials with complete and incomplete outcome data stratified by meta-analyses characteristics.

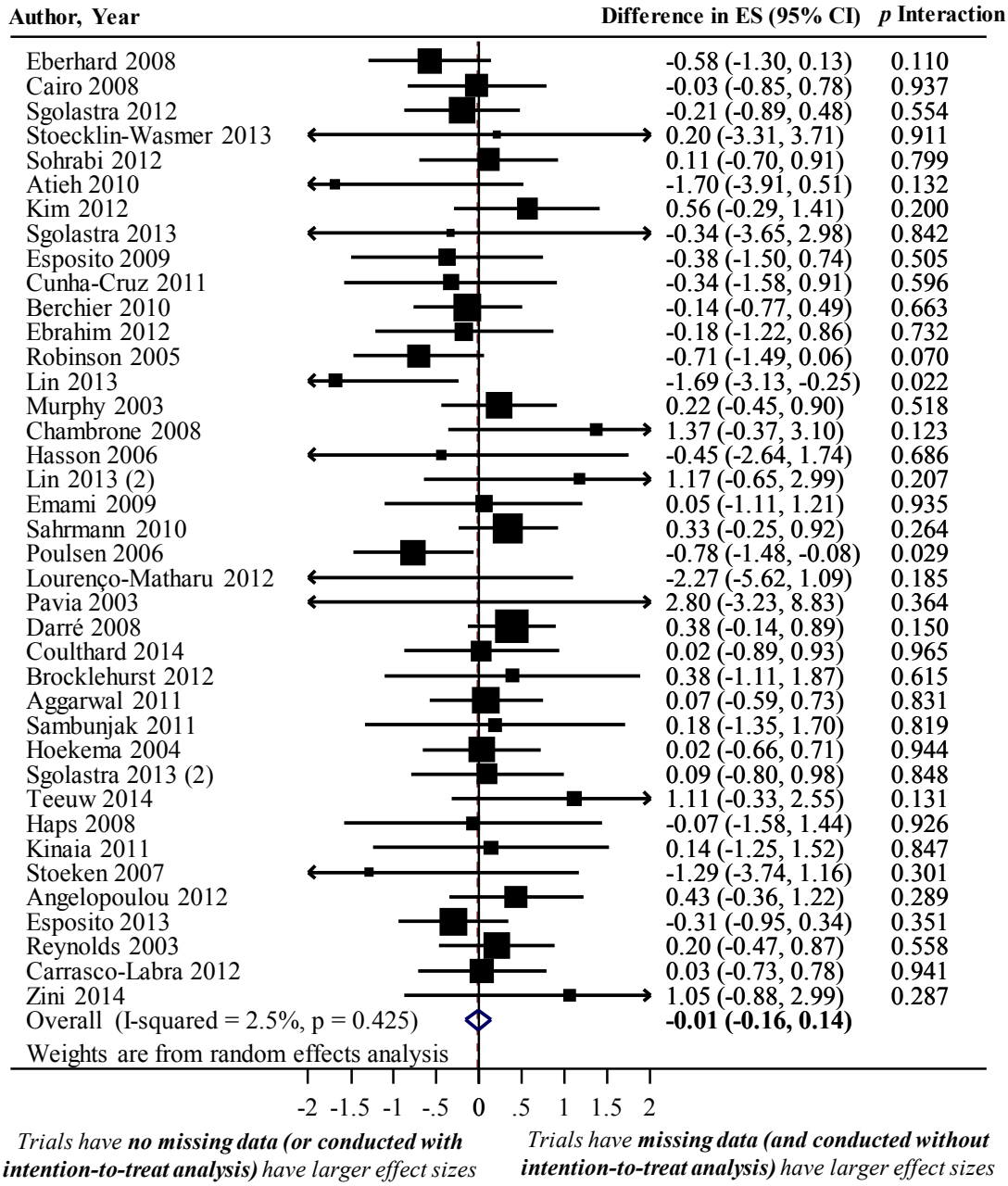

Figure 8.9a. Difference in treatment effect size (ES) estimate between trials with no missing data (or conducted with intention-to-treat analysis) and trials with missing data (or conducted without intention-to-treat analysis). A positive value across meta-analyses indicates that missing data (or lack of intention-to-treat analysis) inflates the treatment effect size estimate.

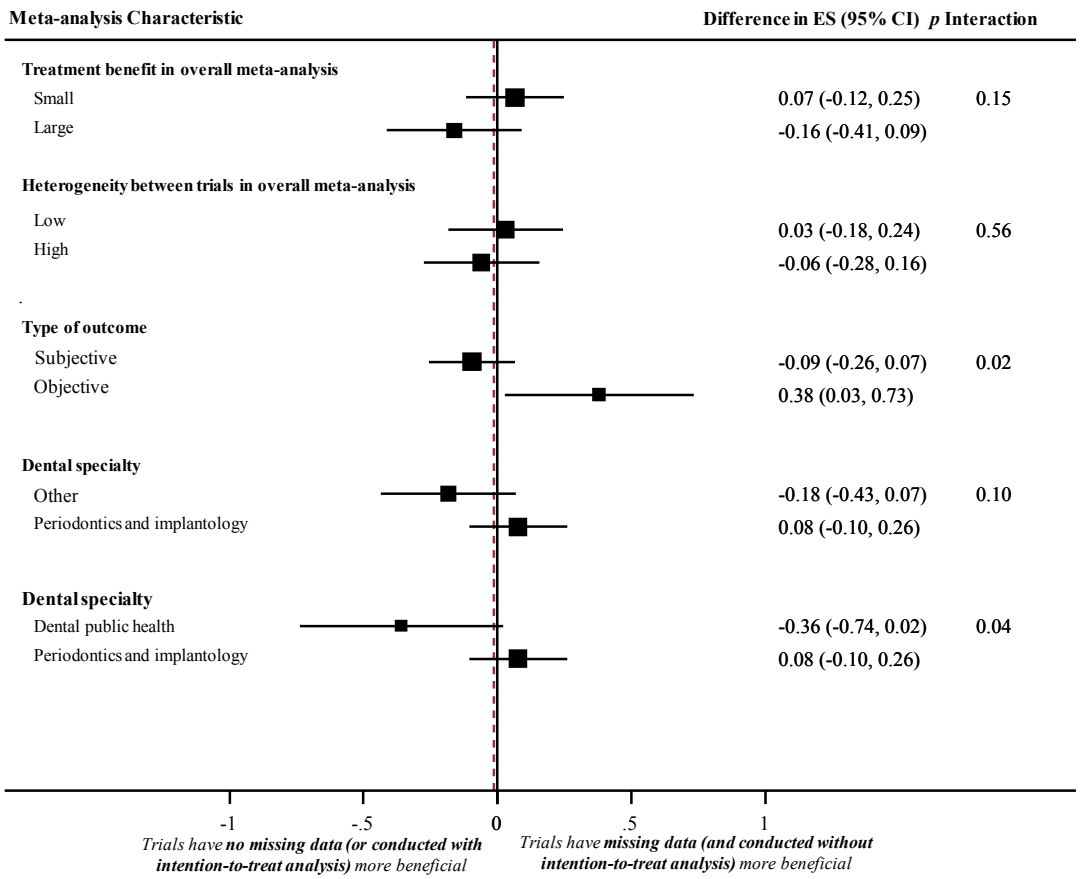

Figure 8.9b. Forest plot of the difference in treatment effect size (ES) estimate between trials with no missing data (or conducted with intention-to-treat analysis) and trials with missing data (or conducted without intention-to-treat analysis) stratified by meta-analyses characteristics.

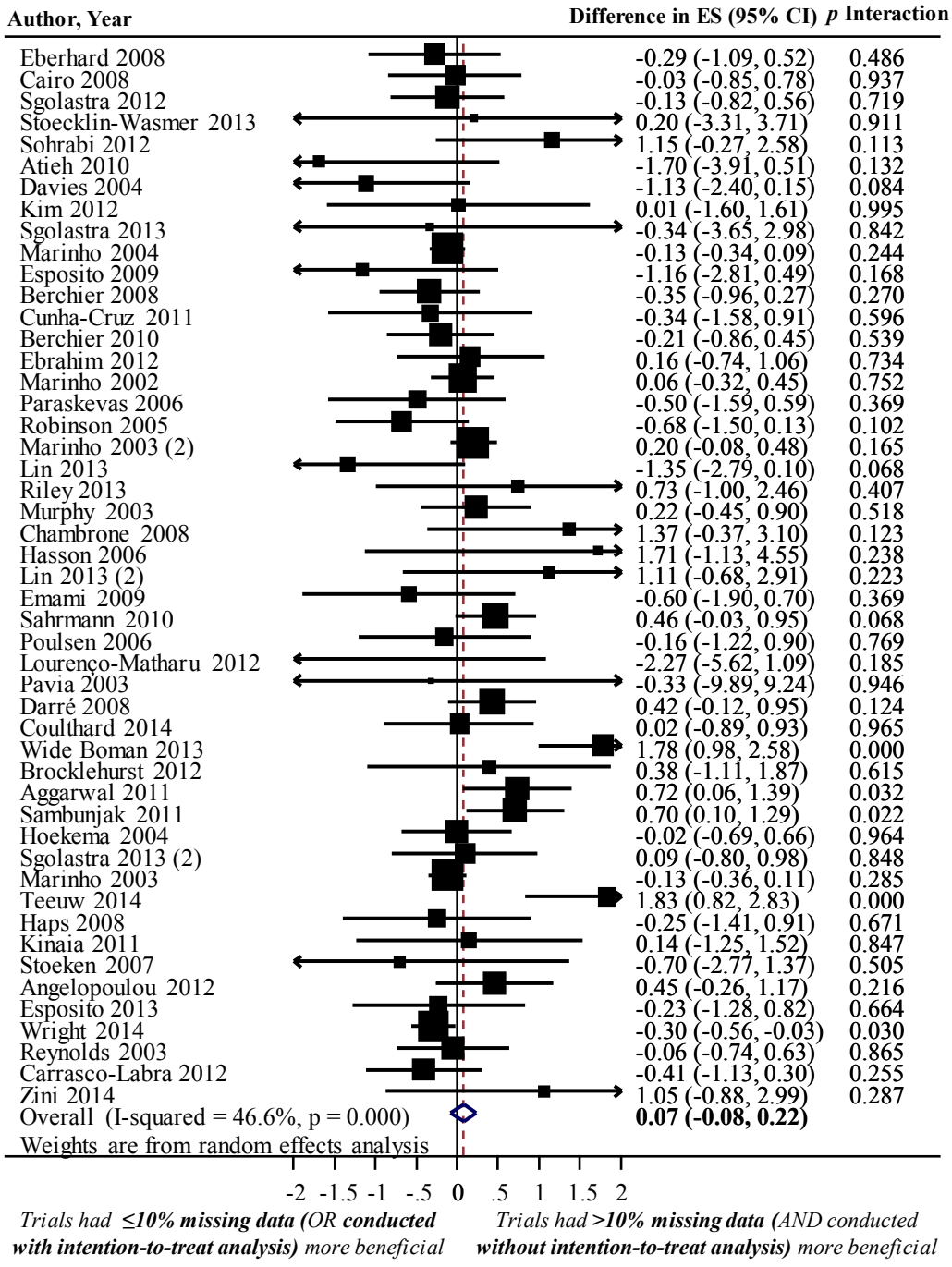

Figure 8.10a. Difference in treatment effect size (ES) estimate between trials with ≤ 10% missing data (or conducted with intention-to-treat analysis) and trials with > 10% missing data (and conducted without intention-to-treat analysis). A positive value across meta-analyses indicates that ≤ 10% missing data inflates the treatment effect size estimate.

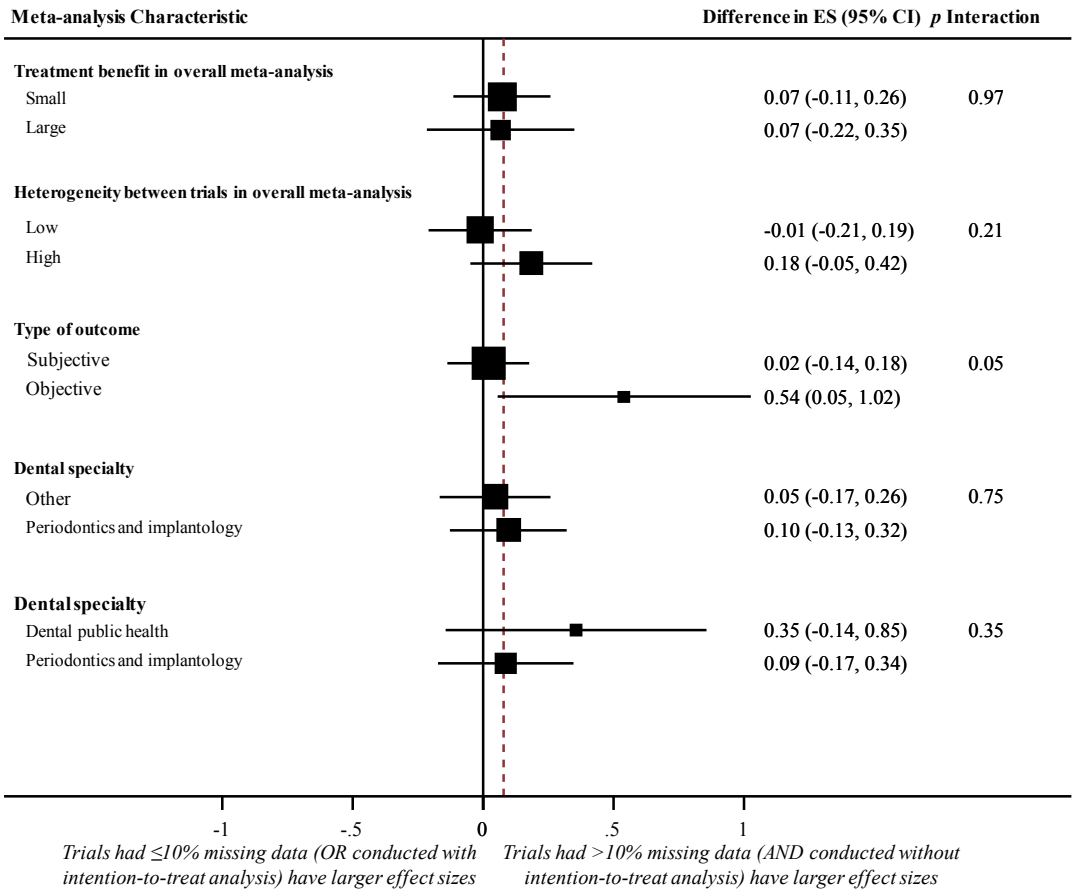

Figure 8.10b. Forest plot of the difference in treatment effect size (ES) estimate between trials with ≤ 10% missing data (or conducted with intention-to-treat analysis) and trials with > 10% missing data (and conducted without intention-to-treat analysis) stratified by meta-analyses characteristics.

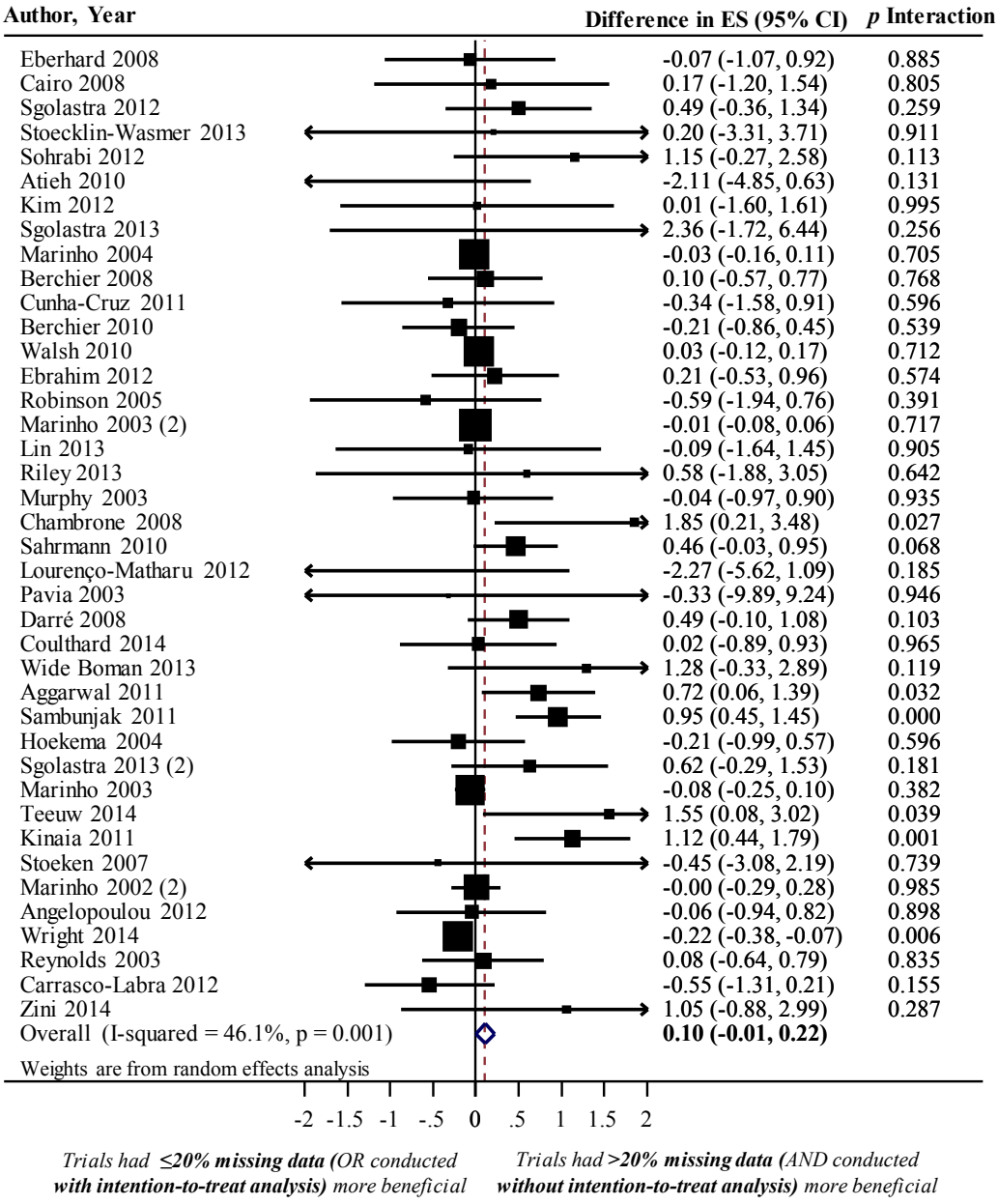

Figure 8.11a. Difference in treatment effect size (ES) estimate between trials with ≤ 20% missing data (or conducted with intention-to-treat analysis) and trials with > 20% missing data (and conducted without intention-to-treat analysis). A positive value across meta-analyses indicates that ≤ 20% missing data (or a lack of intention-to-treat analysis) inflates the treatment effect size estimate.

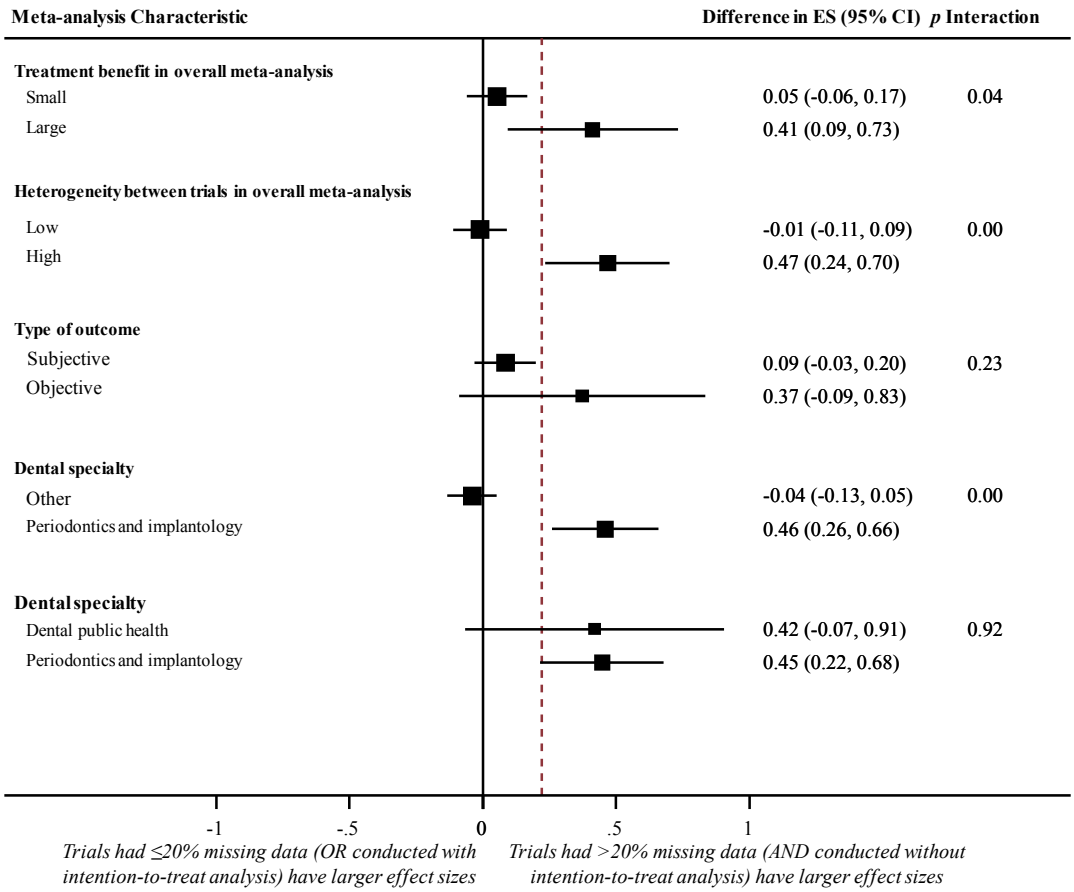

Figure 8.11b. Forest plot of the difference in treatment effect size (ES) estimate between trials with ≤ 20% missing data (or conducted with intention-to-treat analysis) and trials with > 20% missing data (and conducted without intention-to-treat analysis) stratified by meta-analyses characteristics.

References

[1] Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. Int J Surg 2012; 10(1):28-55.

[2] Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. Health Technol Assess 1999; 3(12):i-iv,1-98.

[3] Pandis N, Polychronopoulou A, Madianos P, Makou M, Eliades T. Reporting of research quality characteristics of studies published in 6 major clinical dental specialty journals. J Evid Based Dent Pract 2011; 11:75-83.

[4] Seehra J, Wright NS, Polychronopoulou A, Cobourne MT, Pandis N. Reporting quality of abstracts of randomized controlled trials published in dental specialty journals. J Evid Based Dent Pract 2013; 13:1-8.

[5] Pandis N, Polychronopoulou A, Eliades T. An assessment of quality characteristics of randomised control trials published in dental journals. J Dent 2010; 38(9):713-21.

[6] Dumbrigue HB, Al-Bayat MI, Ng CC, Wakefield CW. Assessment of bias in methodology for randomized controlled trials published on implant dentistry. J Prosthodont 2006; 15:257-63.

[7] Higgins J, Altman D. Chapter 8: Assessing risk of bias in included studies. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions version 50. Chichester, UK: John Wiley & Sons, Ltd.; 2008.

[8] Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. JAMA 2010; 303:1180-7.

[9] Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609-13.

[10] Gluud LL. Bias in clinical intervention research. Am J Epidemiol 2006; 163:493-501.

[11] Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 2002; 287:2973-82.

[12] 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; 336:601-5.

[13] Van Tulder MW, Suttorp M, Morton S, Bouter LM, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine 2009; 34:1685-92.

[14] Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999; 282:1054-60.

[15] Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 2003; 7:1-76.

[16] Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med 2012; 157:429-38.

[17] Hartling L, Hamm MP, Fernandes RM, Dryden DM, Vandermeer B. Quantifying bias in randomized controlled trials in child health: a meta-epidemiological study. PloS One 2014; 9:e88008.

[18] Hempel S, Miles J, Suttorp MJ, Wang Z, Johnsen B, Morton S, et al. Detection of associations between trial quality and effect sizes. Rockville (MD), Agency for Healthcare Research and Quality (US); 2012 Jan. Report No.:12-EHC010-EF.

[19] Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 2003; 326:1167-70.

[20] Needleman I, Worthington H, Moher D, Schulz K, Altman DG. Improving the completeness and transparency of reports of randomized trials in oral health: The CONSORT Statement. Am J Dent 2008; 21:7-12.

[21] Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. Stat Med 2007; 26:2745-58.

[22] Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gotzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int J Epidemiol 2007; 36:847-57.

[23] Hempel S, Suttorp M, Miles J. Empirical evidence of associations between trial quality and effect size. Rockville (MD), Agency for Healthcare Research and Quality (US); 2011 Jun. Report No.:11-EHC045-EF.

[24] Saltaji H, Cummings GG, Armijo-Olivo S, Major MP, Amin M, Major PW, et al. A descriptive analysis of oral health systematic reviews published 1991-2012: cross sectional study. PloS One 2013; 8:e74545.

[25] Koletsi D, Pandis N, Polychronopoulou A, Eliades T. Does published orthodontic research account for clustering effects during statistical data analysis? Eur J Orthod 2012; 34:287-92.

[26] 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; 41:265-70.

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

[28] Armijo-Olivo S, Cummings GG, Fuentes J, Saltaji H, Ha C, Chisholm A, et al. Identifying items to assess methodological quality in physical therapy trials: a factor analysis. Phys Ther 2014; 94:1272-84.

[29] Armijo-Olivo S, Fuentes J, Ospina M, Saltaji H, Hartling L. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: a descriptive analysis. BMC Med Res Methodol 2013; 13:116.

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

[31] Van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the Cochrane Collaboration back review group. Spine 2003; 28:1290-9.

[32] Van Tulder MW, Assendelft WJ, Koes BW, Bouter LM. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for Spinal Disorders. Spine 1997; 22:2323-30.

[33] Moseley AM, Herbert RD, Sherrington C, Maher CG. Evidence for physiotherapy practice: a survey of the Physiotherapy Evidence Database (PEDro). Aust J Physiother 2002; 48:43-9.

[34] Sherrington C, Herbert RD, Maher CG, Moseley AM. PEDro. A database of randomized trials and systematic reviews in physiotherapy. Man Ther 2000; 5:223-6.

[35] Bizzini M, Childs JD, Piva SR, Delitto A. Systematic review of the quality of randomized controlled trials for patellofemoral pain syndrome. J Orthop Sports Phys Ther 2003; 33:4-20.

[36] Armijo-Olivo S, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. Phys Ther 2008; 88:156-75.

[37] Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 1998; 51:1235-41.

[38] Armijo-Olivo S, Fuentes CJ, Rogers T, Hartling L, Saltaji H, Cummings G. How should we evaluate the risk of bias of physical therapy trials? A psychometric and meta-epidemiological approach towards developing guidelines for the design, conduct, and reporting of RCTs in Physical Therapy (PT) area: a study protocol. Syst Rev 2013; 2:88.

[39] Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med 2002; 21:1513-24.

[40] Cohen J. Chapter 1: The concepts of power analysis. In: Cohen J, editor. Statistical power analysis for the behavioral sciences. Second ed. Hillsdale, New Jersey: Academic Press, Inc; 1988. p. 1-17.

[41] DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: An update. Contemp Clin Trials 2007; 28:105-14.

[42] Bero L. Industry sponsorship and research outcome: a Cochrane review. JAMA Intern Med 2013; 173:580-1.

[43] Lundh A, Sismondo S, Lexchin J, Busuioc OA, Bero L. Industry sponsorship and research outcome. Cochrane Database Syst Rev 2012; 12:MR000033.

[44] Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? JAMA 2003; 290:921-8.

[45] Fenwick J, Needleman IG, Moles DR. The effect of bias on the magnitude of clinical outcomes in periodontology: a pilot study. J Clin Periodontol 2008; 35:775-82.

[46] Naci H, Dias S, Ades AE. Industry sponsorship bias in research findings: a network meta-analysis of LDL cholesterol reduction in randomised trials of statins. BMJ 2014; 349:g5741.

[47] Naci H, Dias S, Ades T. No evidence of industry sponsorship bias in statin trials. BMJ 2014; 349:g6579.

[48] Popelut A, Valet F, Fromentin O, Thomas A, Bouchard P. Relationship between sponsorship and failure rate of dental implants: a systematic approach. PloS One 2010; 5:e10274.

[49] Schwendicke F, Tu YK, Blunck U, Paris S, Gostemeyer G. Effect of industry sponsorship on dental restorative trials. J Dent Res 2016; 95:9-16.

[50] Molnar FJ, Hutton B, Fergusson D. Does analysis using "last observation carried forward" introduce bias in dementia research? CMAJ 2008; 179:751-3.

[51] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273:408-12.

[52] Nuesch E, Trelle S, Reichenbach S, Rutjes AW, Burgi E, Scherer M, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. BMJ 2009; 339:b3244.

[53] Altman DG. Missing outcomes in randomized trials: addressing the dilemma. Open Med 2009 May 12; 3(2):e51-3. 2009;3:e51-3.

[54] Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 2004; 291:2457-65.

[55] Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. J Natl Cancer Inst 2005 Jul 20; 97(14):1043-55.

[56] Berkman ND, Santaguida PL, Viswanathan M, Morton SC. The empirical evidence of bias in trials measuring treatment differences. Rockville (MD), Agency for Healthcare Research and Quality (US); 2014 Sep. Report No.:14-EHC050-EF.

[57] Armijo Olivo S, Fuentes CJ, Ospina M, Saltaji H, Hartling L. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: A descriptive analysis. BMC Med Res Methodol 2013; 13:1-19.

[58] De Vet HC, De Bie RA, Van Der Heijden GJMG, Verhagen AP, Sijpkes P, Knipschild PG. Systematic reviews on the basis of methodological criteria. Physiotherapy 1997; 83:284-9.

[59] Van Tulder MW, Assendelft WJJ, Koes BW, Bouter LM, Bombardier C, Nachemson AL, et al. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for spinal disorders. Spine 1997; 22:2323-30.

[60] Moseley AM, Herbert RD, Sherrington C, Maher CG. Evidence for physiotherapy practice: a survey of the Physiotherapy Evidence Database (PEDro). Aust J Physiother 2002; 48:43-9.

[61] Bizzini M, Childs JD, Piva SR. Systematic review of the quality of randomized controlled trials for patellofemoral pain syndrome. J Orthop Sports Phys Ther 2003; 33:4-20.

Chapter 9

Discussion

9.1. Overview of Key Findings

These series of investigations were conducted to quantify and evaluate different forms of bias in randomized controlled trials (RCTs) of oral health interventions. Thereby, key recommendations are provided toward the development of a research framework for appraising, reporting, and conducting RCTs in oral health. The recommendations based on this research have great potential to strengthen dentistry research and, ultimately, clinical practice and oral health outcomes. This research contributes to the body of knowledge in evidence-based practice and has relevance to other health care professions.

9.1.1. Development of a register of oral health systematic reviews

Our register of oral health systematic reviews (SRs) included a total of 1188 oral health (126 Cochrane and 1062 non-Cochrane) SRs published from 1991 through to May 2012, encompassing nine dental specialties. The register was further updated in April 2014 and ultimately included 1408 SRs. We used this register as an umbrella source to select a cohort of meta-analyses and associated RCTs of oral health interventions for a series of methodological studies. While including old reviews in the register could be perceived as irrelevant, it facilitated an examination of the status of dentistry-related SRs since their inception. Overall, we found variation in the methodological characteristics of SRs across the nine dental specialties and according to SR category (Cochrane vs. non-Cochrane). The number of SRs published in the domain of oral health research and within each dental specialty has steadily increased

over the last two decades.

We noted that many methodological characteristics of the published oral health reviews require improvement. For example, only 11 of the 1062 non-Cochrane reviews were updates of previously published reviews. Furthermore, none of the 11 updates identified in our research were considered to be "up-to-date" according to the Cochrane policy, which requires the updating of a review every two years [1]. This is a disappointing fact given that "up-to-date" evidenced-based conclusions are considered essential for decision making [2]. This might be explained by the fact that updates are usually given lower priority by funding agencies, and by editors who tend not to publish updates if the results are the same as those previously published [3]. The results also showed that the research design of the included studies, and the number of included studies, varied across dental specialties and by type of review. In our study, while almost all the Cochrane reviews included RCTs only, a small proportion (17.6%) of the non-Cochrane reviews exclusively included RCTs. Interestingly, a sizable proportion of the Cochrane reviews (17.5%), including nearly a third of the reviews in the field of oral and maxillofacial surgery, included no eligible trials. The low proportion of RCTs highlights the need for more trials to be conducted in the dental specialties, particularly in specialties related to oral and maxilla-facial surgery.

9.1.2. Evaluation of risk of bias assessments of trials in oral health SRs

We found that the investigators of almost one-third of non-Cochrane oral health SRs published from 1991 to 2014 did not assess the risk of bias in primary studies, and that such assessments occurred more often in Cochrane reviews than in non-Cochrane reviews. The fact that investigators of only two-thirds of the oral heath reviews assessed risk of bias is concerning for clinicians who interpret the results, because they will have limited ability to assess how estimates of treatment effect size may have been biased owing to way the study was conducted. Consequently, clinical

decisions may not be made confidently on the basis of the findings in some of the primary studies. The use of risk of bias assessments calculated in this study for oral heath interventions was less than that reported in a 2016 study [4], in which investigators examined a sample of 309 reviews published in the domain of medical research. This suggests that research in the various fields of dentistry is falling behind research in the medical field.

We noted that no tool was identified as being specifically designed to assess risk of bias in oral health trials. However, we found that the Cochrane Collaboration's risk of bias tool [5] was the most commonly used tool in oral health reviews, and that this tool had potential value for oral health SRs. Although the Cochrane tool has played an important role in improving risk of bias assessments in health research, the tool requires further development and improvement. Even the developers of the Cochrane Collaboration's risk of bias tool pointed out that the tool's domains need to be expanded to accommodate its use in different health areas [6, 7]. The developers of the tool [6, 7] and other researchers [8, 9] called for more meta-epidemiologic studies in a wider range of health disciplines to support existing domains and to add new domains.

We found that investigators used nonvalidated risk of bias assessment tools (consisting of items extracted from a variety of tools) in 38% of the included reviews. Modification of risk of bias tools is likely to have affected their validity, and therefore affected the applicability of the results. Without having validation of a newly developed tool or group of quality items, clinicians could question the interpretation of the review findings. This is especially critical when using an overall quality and risk of bias score, which may differ conceptually among tools (for example, placing more or less weight on masking and leaving out concealment); the individual weighting of scored items should also undergo a validation process. Furthermore, we found that the investigators of reviews published in dental journals were less likely to have assessed the risk of bias of individual trials, than the investigators of reviews published in

nondental journals. This raises concerns regarding the quality of evidence of SRs in the former, compared to the latter. The results of our study demonstrated variations across the nine dental specialties; clearly, improvement in the conduct and reporting of SRs in specific dental specialties (such as prosthodontics and restorative dentistry and oral and maxillofacial surgery) is needed.

9.1.3. Assessments of risk of bias of oral health RCTs over time

Overall, our study showed a significant increase in the proportion of trials judged as having adequate quality, or a low risk of bias, over time, in the majority of quality items and risk of bias domains. This encouraging trend is similar to that identified in a recently published report by Reveiz et al. [10]. However, Reveiz et al. used the results of risk of bias assessments reported by the investigators of reviews, rather than by conducting standardized data extraction from each trial. This might be problematic, given the documented low reliability of the Cochrane Collaboration's risk of bias tool [9, 11, 12]. The trend in our study was also comparable to that found in a cohort of child-related trials [13] and medical RCTs [14].

Although an improvement over time was identified in trials of oral health interventions, we found that the results of risk of bias and reported methodological quality were unpropitious, indicating continued substandard quality and high potential for bias in such trials. There was considerable evidence that there is a possibility for further, sizable, improvements in the conduct and reporting of RCTs of oral health interventions. Remarkably, the proportion of trials judged as having a low risk of bias did not exceed 60% in the majority of the risk of bias domains. This is concerning, because of the potential impact of inadequate trial design and conduct on treatment effect size estimates. Clinical decisions made in dental practice may, therefore, not be based on valid findings. For example, allocation concealment and sequence generation were judged to be unclear in 84.8% and 66.7% of the trials, respectively, although this significantly improved over time. It should be noted that one of the clear limitations of

the risk of bias assessment is that, having an "unclear" risk of bias result in a trial, does not necessarily mirror the actual design and conduct of the trial, given that medical and dental journals have a word limit that may restrict the reporting of detailed methodology. This limitation is inherent in all of the quality assessment tools when authors of a clinical trial do not adequately report the characteristics of the methodologies they used [15]. In addition, not all the systematic review authors attempted to contact the authors of the included RCTs to clarify missing information.

Improvements in risk of bias and reported methodological quality of RCTs over time could be attributed to the efforts made by dental journal editors and reviewers to endorse the Consolidated Standards of Reporting Trials (CONSORT) Statement [16, 17] and the mandatory implementation of trial registration policy (as recommended by the ICMJE [18, 19]). Although the CONSORT Statement applies only to reporting quality, it is used, mistakenly, by many dentistry researchers as a methodological quality assessment tool. Endorsement of the CONSORT Statement by dental journal editors and reviewers does not guarantee compliance by trialists [17]. However, implementation of the mandatory trial registration policy [20] is an efficient and effective way to detect potential biases such as selective outcome reporting and publication bias, to promote submissions of dental clinical trials having a low risk of bias, to optimize methodological quality assessment, and to assist systematic reviewers of oral health intervention trials in gauging publication bias [18, 20].

9.1.4. Impact of bias on treatment effect size estimates in RCTs of oral health interventions

Use of the *P*-value to test the significance of the null-hypothesis is widely reported in the medical literature, despite the fact that this procedure does not provide information about the magnitude of the treatment effect and the "precision" of the treatment effect size estimate [21]. In contrast, treatment effect size confidence intervals were encouraged as alternative measures to examine associations within

analyzed data [21, 22] that do not necessarily indicate a clinically relevant effect. The interpretation of study findings on the basis of confidence intervals moves the interpretation from the dichotomy of significance (significant vs. nonsignificant) to an examination of clinical relevance. Therefore, we relied in the current study on differences in treatment effect size estimates and their confidence intervals when interpreting associations between methodological quality criteria and treatment effect size.

Selection bias (sequence generation, allocation concealment, and baseline comparability) and treatment effect size estimate

We show that treatment effect size estimates were 0.13 larger in trials with inadequate sequence generation compared to trials with adequate sequence generation, and 0.15 larger in trials with inadequate allocation concealment compared to trials with adequate allocation concealment. However, baseline comparability was not associated with inflated or underestimated treatment effect size estimates in our study. With respect to the direction and magnitude of the treatment effect, our results agree with studies that consistently showed that inadequate sequence generation could exaggerate treatment effect size estimates by 51% [23], 36% [24], and 11% [25, 26], compared to trials that employed adequate sequence generation. Similarly, inadequate allocation concealment has been associated with an increase in treatment effect size estimates of 52% [23], 34% [27], and 10% [28], compared to trials with adequate allocation concealment. However, this association has not been confirmed in other studies [29, 30]. Previous reports examining the influence of selection bias were restricted to RCTs in specific medical areas such as pediatrics [31], low-back pain [32], osteoarthritis [33], and physical therapy [34].

While the above-mentioned studies [23-26, 29, 30] assessed dichotomous outcomes, two recent studies [31, 34] examined associations between treatment effect

size estimates and inadequate sequence generation and between treatment effect size estimates and inadequate allocation concealment, using continuous outcomes. One of these studies [30] included 287 pediatric trials from 17 meta-analyses, and, based on potential selection bias, found no significant difference in treatment effect size estimates among trials. The second study [34] assessed 275 physical therapy trials included in 22 meta-analyses and found that trials with inadequate allocation concealment displayed a trend toward an exaggeration of treatment effect size compared to trials that employed adequate allocation concealment ($p = 0.06$, moderate evidence against the null hypothesis of no difference between trials with and without adequate concealment of allocation), whereas no differences in treatment effect size estimates were found between trials with adequate or inadequate sequence generation. However, while a meta-epidemiological analysis requires a large number of meta-analyses and trials, the vast majority of the studies mentioned above did not employ large numbers of meta-analyses and trials. The inconsistent findings might be attributed to the use of different statistical approaches; assessing different types of interventions, outcomes and populations [35]; and to the improper inclusion of trials with comparable active interventions (it is difficult to accurately calculate differences in treatment effect size estimates between comparable active interventions).

Performance bias (patient and care-provider blinding, similarity of cointerventions, and compliance to treatment) and treatment effect size estimate

We identified a tendency toward exaggeration of treatment effect size (although not statically significant) based on imbalances in cointerventions (difference in treatment effect size estimate = 0.08, at 95% confidence interval [CI]: -0.11 to 0.27) and inadequacy of compliance to treatment (difference in treatment effect size estimate = 0.10, at 95% CI: -0.02 to 0.22). With respect to blinding, trials with lack of patient blinding had significantly larger treatment effect size estimates (0.12 larger) than trials with adequate patient blinding. However, care-provider blinding was not

related to inflated treatment effect size estimates. Reports examining the impact of lack of patient and care-provider blinding reported inconsistent findings: in [30], treatment effect size estimates were smaller in trials that employed patient blinding than in trials that did not employ patient blinding, whereas [36, 37] cited smaller treatment effect size estimates in trials with the lack of patient blinding than in trials with patient blinding. However, associations between treatment effect size estimates and the presence or lack of blinding were not reported in other studies [32, 33].

The inconsistent findings might be due to the examination of different types of outcomes, interventions, and populations; the implementation of different definitions for quality assessment; and the use of various statistical and modeling approaches [35]. For example, Schulz et al. [38] applied a multiple logistic regression model to analyze data on binary outcomes from 250 trials included in 33 meta-analyses; double blinding was defined on the basis of whether the trial's conduct claimed to be a double-blind. Egger et al. [39] defined "double blinding" based on whether the trial was described as double-blind or included, at least, assessor blinding; the study analyzed data from 304 trials included in 39 meta-analyses with binary outcomes in several medical fields (infection diseases, neurology, among others).

Detection bias (assessor blinding) and treatment effect size estimate

We found that trials with lack of assessor blinding tended to have significantly larger treatment effect size estimates than trials with adequate assessor blinding (difference in treatment effect size estimate = 0.06), although the association was not statistically significant (95% CI -0.06 to 0.18). Published evidence examining the impact of lack of assessor blinding was conducted in particular medical fields such as physical therapy [37], thrombosis and cardiovascular disease [29, 30], pediatrics [31], osteoarthritis [33], and low-back pain [32]. The studies reported inconsistent findings: in two studies [29, 40] the treatment effect size estimates were smaller in trials with assessor blinding, and in two studies [30, 37] the treatment effect size estimates were

smaller in trials with the lack of assessor blinding. However, associations between the presence or lack of blinding and treatment effect size estimates were not reported in some studies [32, 33].

Two recent studies [31, 37] that examined the association between lack of assessor blinding and treatment effect size estimates, using continuous outcomes, also reported inconsistent findings. One study assessed the adequacy of assessor blinding in 287 pediatric trials from 17 meta-analyses [31], and showed no significant difference in treatment effect size estimates between studies, based on potential bias related to lack of blinding. The second study [37] assessed 165 physical therapy trials included in 17 meta-analyses, and found that trials with lack of assessor blinding tended to underestimate treatment effect size compared with trials that employed appropriate blinding (although, the differences were nonstatistically significant).

Performance-detection bias (blinding of both patients and assessors; blinding of patients, assessors, and care-providers) and treatment effect size estimate

Our study is the first meta-epidemiological study conducted in any medical or dental field that examines the impact of blinding of patients and assessors (double blinding), and patients and assessors and care-providers (triple blinding), on treatment effect size estimates in RCTs of oral health interventions with continuous outcomes. We found that while treatment effect size estimations were 0.14 larger in trials with lack of blinding of patients, assessors, and care-providers, blinding of both assessors and patients was associated with the largest overestimation of treatment effect size, 0.19 larger than in trials with lack of blinding of both patients and assessors ; this measured magnitude of bias represents approximately one quarter to one third of the common treatment effect size estimate reported in oral health research [41], for example, in clinical outcomes in periodontology [42]. The fact that treatment effect size estimates in oral health trials may have been biased due to lack of blinding is concerning; clinical decision making related to recommended dental treatments and

modalities may therefore not be based on valid findings.

Because the concept of blinding is implemented at multiple levels of a clinical trial (e.g., patients, assessors, care providers, data analysts, investigators), there is confusion when the implemented level of blinding is described. For example, "double blinding" or "triple blinding" at one level can refer to the blinding of any two or three of the participants at previous levels. Failure to clearly report the levels to which such terms refer, confuses readers of the trial report. A recent study by Kahan et al. [43] indicated that blinding of outcome assessors is uncommonly used, and inadequately reported in a cohort of 258 medical trials.

Attrition bias and treatment effect size estimate

We found significant differences in treatment effect size estimates in oral health trials, based on inadequate reporting of withdrawal/dropout rates, and having > 20% dropout without performing the analysis based on the intention-to-treat approach (compared with trials that had ≤ 20% missing data or were conducted with an intention-to-treat analysis). Furthermore, a tendency toward exaggeration of treatment effect size estimate (at 95% CI: -0.08 to 0.22) was also found, although it was not statistically significant, based on having more than 10% missing data. In contrast, an acceptable dropout rate, the presence of complete outcome data, and an analysis based on the intention-to-treat approach, were not associated with inflated or underestimated treatment effect size estimates. Findings from studies that examined the association between attrition bias and treatment effect size estimate were inconsistent in terms of the direction and magnitude of the associations. This conclusion was based on the definition of attrition, classification, and the type of outcome evaluated [23, 31, 32, 38, 44]. For example, studies that defined attrition bias as dropouts related to intervention, found that trials with a higher dropout level had larger treatment effect size estimates [38]. In a study by Nuesch et al. [45] that assessed 167 trials (included in 14 meta-analyses) investigating patients with osteoarthritis and using pain as an outcome,

concluded that the magnitude and direction of bias associated with patient exclusion was "unpredictable" and that bias frequently inflated treatment effect size estimates. A study by Hartling et al. [31] used the Cochrane Collaboration's risk of bias tool to define attrition bias as "incomplete outcome data" in 287 child-health trials and, based on attrition bias, found no significant difference in treatment effect size estimates among trials.

We found that trials that based the analysis of results on the intention-to-treat approach were not associated with inflated or underestimated treatment effect size estimates. However, in meta-epidemiological analyses of trials with missing data of more than 10% or 20%, the difference between the treatment effect size estimate and the p value increased from 0.07, with a p value of 0.374 (for more than 10% dropouts), to 0.10, with a p value of 0.070 (for more than 20% dropouts). Based on this finding, the acceptable dropout cut-off could be around 20%, with an increase in bias expected in trials with dropout rates above this level. While our study did not show significant differences in treatment effect size estimates based on the acceptability of dropout/withdrawal rates, interestingly, a failure to report dropout/withdrawal rates was associated with a 0.24 increase in treatment effect size estimate; this magnitude represents one third to one half of the treatment effect size estimates observed in many dental interventions [42]. It appears that in the field of dentistry, reporting dropout rates is more important than having a high dropout rate with respect to lowering associated bias. Importantly, a failure to report dropout rates in a trial might be attributed to patients' response to the examined intervention, especially if the frequency of withdrawal and/or the motives for withdrawal vary between the intervention groups [46]. Based on this finding, the reporting of dropout rates could be one of the stratifying factors when synthesizing evidence from meta-analyses, especially in dentistry.

Sponsorship and reporting bias and treatment effect size estimate

The findings of our study support the suggestion that industry sponsorship can bias the end results of clinical trials toward favoring industry-related interventions [47-50]. Bias due to an inappropriate influence of funding was associated with a difference in treatment effect size estimate of 0.10, which represents one-fifth of treatment effect estimates in some dental specialties [42]. While the aforementioned studies described numerous possible scenarios where the level of sponsorship could influence the design, conduct, and reporting of a clinical trial, other studies [31, 51, 52] did not detect a significant influence of funding on the treatment effect size estimate. These contradictory results might be explained by differences in assessing and defining sponsorship bias, types of sponsorship (e.g., pharmacological products and financial supports to trialists), and types of trials evaluated (e.g., placebo-controlled or active control). The results from this study raise the question of whether industry-sponsored oral health trials are sometimes biased for the benefit of the sponsoring company, leading to inappropriate treatment decisions. For example, a recently published report [53] that examined the influence of industry sponsorship in 41 RCTs of dental implants, found that implant failure in sponsored RCTs was much lower than implant failure in nonsponsored RCTs. Conversely, a recent network meta-analysis [54] that assessed the impact of industry sponsorship on 114 dental restorative RCTs, found that material performance rankings did not differ on the basis of sponsorship. That study concluded that the influence of industry sponsorship on RCTs of restorative dentistry was “limited.”

We found that selective outcome reporting was not associated with significant differences in treatment effect size estimates. Although this result agrees with a study by Hartling et al. [17], it disagrees with results of previous studies [58-60] that found larger treatment effect size estimates in trials that employed selective outcome reporting than in trials that did not employ selective outcome reporting. Our finding could be attributed to the fact that the current study relied on the similarity of outcomes presented in methods, compared with outcomes reported in results.

Furthermore, we did not evaluate the published protocols of trials when assessing whether the selective outcome reporting was or was not associated with significant differences in treatment effect size estimates. Future investigations should evaluate the published protocols of the RCTs to determine this association.

9.2. Implications

Several implications for oral health research, policy, practice, and decision-making can be drawn based on the findings of the four-phase study contributing to this dissertation.

9.2.1. Implications for systematic reviewers and meta-analysts

An analysis of the developed register of oral health SRs suggests there is room for improvement in oral health reviews. For example, there is a clear need for more regular updating of SRs to ensure that dental practice decision-making is based on up-to-date information. This includes an examination of where updates are needed and the development of mechanisms to regularly update reviews. Furthermore, the findings highlight the need for more trials to be planned in dental specialties specifically related to oral and maxilla-facial surgery, oral medicine, and oral pathology, and to be conducted and reported with attention to the highest possible standards.

Our evaluations of the risk of bias in reviews of trials of oral health interventions call for authors of SRs to use tested or validated items and assessment tools when assessing the risk of bias of individual studies, and to explicitly report the results for each quality item or risk of bias domain. Also, systematic review authors should state which domain they consider to be the most important for their assessment of quality and interpretation of results, and explain how they condense individual items into a final score. The use of items from different risk of bias assessment tools may be more acceptable than using an overall score for some oral health trials, as long

as the items are linked to important potential biases.

Systematic review authors should consider the potential design characteristics of oral health trials. RCTs of oral health interventions have some unique design characteristics, such as the use of a broad range of concomitant interventions (surgical, nonsurgical, drug, and nondrug), difficulty in applying blinding, and a common use of the split-mouth design. These features can add complexity with respect to reporting and applying strategies that reduce the potential for biases, threatening a study's internal validity and, hence, its external validity.

Exclusion by SR authors of trials of oral health interventions could be considered an acceptable methodological approach when conducting meta-analyses in reviews, based on the inadequacy of the following methodological factors: sequence generation, allocation concealment, patient and assessor blinding, reporting of dropout rate, and influence of funders. Alternatively, systematic review authors should perform sensitivity analyses of included trials based on the above-mentioned factors, and consider the likely bias associated with these factors when interpreting the findings of a quantitative analysis.

9.2.2. Implications for trialists and dental researchers

When conducting trials, dental trialists should report results explicitly and adhere to published guidelines. Systematic research methodology must be taught in dental undergraduate and postgraduate programs to provide dental researchers with adequate knowledge to design, conduct, and report a trial.

9.2.3. Implications for dental research methodologists

Although not without problems, the Cochrane Collaboration's risk of bias tool [5] is the best available approach to assessing risk of bias in oral health SRs.

Improvement of this tool should be ongoing. For example, the "inappropriate influence of funding," in the context of risk of bias assessment, could be added as an individual domain.

9.2.4. Implications for dental journal editors and reviewers

The findings of the current study call for dental journal editors and reviewers to continue to be committed to the international initiatives and statements that have been developed to ensure adequate and appropriate conduct and reporting of RCTs. To minimize the impact of bias on treatment effect size estimates, the risk of inaccurate conclusions being drawn, and, accordingly, inappropriate recommendations being made regarding treatment interventions in dental practice, peer reviewers and editors of dental journals should require authors of SRs and RCTs to adhere to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [55], and should insist that authors show adequate conduct and adequate reporting of risk of bias assessments. Furthermore, they should insist on adequate conduct and reporting of submitted RCTs for publication, because of the expected impact of bias on the treatment effect estimates.

9.2.5. Implications for dental practitioners

To fully appreciate the findings in trial reports, clinicians should have an adequate knowledge of the design, conduction, and reporting of a clinical trial. When interpreting the findings in SR reports, clinicians should be able to correctly identify the type of primary study, to adequately appraise the quality of included studies, and to appreciate the risk of bias assessment. A clinician should be able to determine to what extent the findings of an SR are valid on the basis of whether the investigators have assessed a risk of bias and how the bias was considered when the findings were interpreted. A clinician should be able to estimate the degree to which the conclusions of a meta-analysis are synthesized and interpreted, based on the examples of well-

conducted trials. This knowledge will enable a clinician to deliver the best possible results in his or her own dental practice.

9.2.6. Implications for decision-making and for policy

Oral-health guideline developers and policymakers should continue to follow proper, evidence-based decision-making when formulating guidelines and putting policies about implementation of therapeutic interventions in dental practices. Decision makers, from government and public sectors, should base financial decisions to support oral health services on the findings of SRs and RCTs that have achieved a low risk of bias.

It is clear from the findings in the current study that oral health policy-makers, methodologists, and researchers need to develop initiatives for improving clinical trials. We suggest that the formation of a global oral health initiative that aims to improve the conduct and reporting of oral health trials would spread action within the dental community. One example of a much needed measure to ensure that RCTs meet high standards would be to prioritize methodological criteria in oral health research.

9.3. Strengths and Limitations of the Study

Our four-phase study, conducted in the domain of oral health research, provides a comprehensive and in-depth assessment of the following subjects: (1) oral health SRs, with respect to review characteristics, methodology, and risk of bias assessment of included trials; (2) RCTs of oral health interventions, with respect to reporting, methodological characteristics, risk of bias, and the variation of these factors over time; and (3) associations between treatment effect size estimates and bias in RCTs. The range and size of our sample provided a wide-ranging and thorough evaluation of oral health SRs (over the 23-year period of 1991–2014) and RCTs (over the 58-year-period of 1955–2013). The large number of SRs and RCTs, covering all

designs of trials (including split-mouth design trials) and all dental specialties, increases the statistical power and precision of the analysis and assures the generalizability of the results. Our meta-epidemiological work is the first large study conducted in the domains of dental, oral, and craniofacial research, and is one of a very few meta-epidemiological studies conducted in any medical field to examine the impact of bias on treatment effect size estimates in RCTs. The number of trials assessed in our study is two- to three-times the number of trials used in preceding meta-epidemiological studies conducted in other medical fields. We applied a strict methodology to data collection and data analysis, based on our previous meta-epidemiological work, to improve our methodological approach and address potentially limiting factors associated with this type of methodological research. For example, we performed a standardized data extraction (in duplicate by two assessors), rather than relying on the risk of bias assessment reported in the SRs. Thus, our work provides evidence that is novel and of high priority and interest to researchers and methodologists, particularly in the field of oral health research.

Our study has several potential limitations. First, our empirical evidence examined only published reports, not actual conducted randomized trials; this could lead to bias based on reported methodological characteristics. Second, we did not contact authors of the SRs or the RCTs for missing data, given that a large proportion of the SRs and RCTs were published before the year 2005 when corresponding author information was not provided or not up-to-date. Third, we performed data extraction and analyses based on the information given by the authors of the SRs and RCTs in the published reports. This approach, although widely used, limits the identification of actual bias when trial authors do not adequately report study elements. Additionally, we used our judgment to assign each RCT to a primary dental specialty (e.g., dental public health), whereas the RCT could be classified under more than one specialty (e.g., both pediatric dentistry and dental public health).

Certain levels of heterogeneity are expected in our meta-epidemiological

study, given that these types of studies are built on numerous entities of analyses (meta-analysis, trials, and participants) that have a distinct potential for heterogeneity [35]. By applying a cautious methodology to data collection and analysis, and by assembling a large number of meta-analyses and trials in our study, we increased the power of the study and minimized heterogeneity. Also, we restricted our data analysis to trials in which the direction of expected treatment effect estimate was visible, including trials involving a control or placebo intervention; this procedure reduced heterogeneity and confounding factors in the analyses, allowing for the detection of significant effects of methodological characteristics. Additionally, the inclusion of trials with comparable active interventions (where identification of the direction of treatment effect is difficult) leads to inaccuracy in the calculated difference in treatment effect size estimates. Finally, our study did not assess the likely effects of interactions with other design biases. Such an assessment would necessitate a multivariate analysis with a larger number of meta-analyses and trials [36].

9.4. Future Research

The developed register of oral health SRs should be further updated; an evaluation of additional methodological aspects related to the conduct and reporting of oral health SRs should be explored. For example, there is a need to assess publication bias, and the factors associated with its conduct. In the context of methodological aspects related to SRs of oral health interventions, an area of particular weakness is the implementation of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach [56] in dentistry to evaluate the strength of a body of evidence.

Future meta-epidemiological studies should be highly powered and use consistent methodological approaches for trial assessment to accommodate potential heterogeneity sources associated with these types of studies [35]. Such studies should

also use valid and reliable quality assessment tools when assessing the risk of bias of individual trials, and explicitly report findings for each risk of bias domain. Future meta-epidemiological investigators could examine associations between methodologic and nonmethodologic characteristics and the magnitude of treatment effect size estimates in dichotomous outcomes in RCTs of oral health interventions. They could determine if other nonmethodological trial characteristics, such as the number of centers (i.e., multicenters vs. single-center), the type of funding, the sample size, and the design of a trial (i.e., split-mouth vs. parallel) are associated with different treatment effect size estimates in oral health RCTs. Another potential track for future investigations is to examine associations between methodologic and nonmethodologic characteristics and the magnitude of the treatment effect estimates on specific outcomes in RCTs of oral health interventions. There is also a clear need for investigators of future meta-epidemiological studies to assemble a greater number of meta-analyses and trials by synthesizing results from different disciplines and datasets. Moreover, investigators should use other meta-epidemiological analyses to identify an exact cut-off ceiling for acceptable dropout levels in RCTs of oral health interventions.

Finally, future research could explore the perceptions of investigators conducting RCTs and SRs regarding systemic biases in the field of oral health research, given the undesirable influence of bias on treatment effect size estimates.

References

[1] Shea B, Moher D, Graham I, Pham B, Tugwell P. A comparison of the quality of Cochrane reviews and systematic reviews published in paper-based journals. Eval Health Prof 2002; 25:116-29.

[2] Moher D, Tsertsvadze A. Systematic reviews: when is an update an update? Lancet 2006; 367:881-3.

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

[4] Seehra J, Pandis N, Koletsi D, Fleming PS. Use of quality assessment tools in systematic reviews was varied and inconsistent. J Clin Epidemiol 2016; 69:179-84 e5.

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

[6] Hrobjartsson A, Boutron I, Turner L, Altman DG, Moher D. Assessing risk of bias in randomised clinical trials included in Cochrane Reviews: the why is easy, the how is a challenge. Cochrane Database Syst Rev 2013; 4:ED000058.

[7] Savovic J, Weeks L, Sterne JA, Turner L, Altman DG, Moher D, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. Syst Rev 2014; 3:37.

[8] Hartling L, Hamm MP, Milne A, Vandermeer B, Santaguida PL, Ansari M, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. J Clin Epidemiol 2013; 66:973-81.

[9] Armijo-Olivo S, Ospina M, da Costa BR, Egger M, Saltaji H, Fuentes J, et al. Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. PloS One 2014; 9:e96920.

[10] Reveiz L, Chapman E, Asial S, Munoz S, Bonfill X, Alonso-Coello P. Risk of bias of randomized trials over time. J Clin Epidemiol 2015; 68:1036-45.

[11] Hartling L, Hamm MP, Milne A, Vandermeer B, Santaguida PL, Ansari M, et al. Testing the Risk of Bias Tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. J Clin Epidemiol 2012; 66:973-81.

[12] Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials? Evaluation of risk of bias assessments in systematic reviews. BMJ 2013; 346:f1798.

[13] Thomson D, Hartling L, Cohen E, Vandermeer B, Tjosvold L, Klassen TP. Controlled trials in children: quantity, methodological quality and descriptive characteristics of pediatric controlled trials published 1948–2006. PloS One 2010; 5.

[14] Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ 2010; 340:c723.

[15] Hamm MP, Hartling L, Milne A, Tjosvold L, Vandermeer B, Thomson D, et al. A descriptive analysis of a representative sample of pediatric randomized controlled trials published in 2007. BMC Pediatr 2010; 10:96.

[16] Pandis N, Turpin DL. Enhancing CONSORT compliance for improved reporting of randomized controlled trials. Am J Orthod Dentofacial Orthop 2014; 145:1.

[17] Turpin DL. Challenges faced by journal editors. Am J Orthod Dentofacial Orthop 2015; 147:285-9.

[18] Hooft L, Korevaar DA, Molenaar N, Bossuyt PM, Scholten RJ. Endorsement of ICMJE's Clinical Trial Registration Policy: a survey among journal editors. Neth J Med 2014; 72:349-55.

[19] Smail-Faugeron V, Fron-Chabouis H, Durieux P. Clinical trial registration in oral health journals. J Dent Res 2015; 94:8S-13S.

[20] De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. Lancet 2004; 364:911-2.

[21] Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev Camb Philos Soc 2007; 82:591-605.

[22] Dunst CJ, Hamby DW. Guide for calculating and interpreting effect sizes and confidence intervals in intellectual and developmental disability research studies. J Intellect Dev Disabil 2012; 37:89-99.

[23] Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2001; 135:982-9.

[24] Linde K, Scholz M, Ramirez G, Clausius N, Melchart D, Jonas WB. Impact of study quality on outcome in placebo-controlled trials of homeopathy. J Clin Epidemiol 1999; 52:631-6.

[25] Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. Health Technol Assess 1999; 3:i-iv, 1-98.

[26] Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998; 352:609-13.

[27] Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med 2002; 21:1513-24.

[28] Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gotzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int J Epidemiol 2007; 36:847-57.

[29] Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999; 282:1054-60.

[30] Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 2002; 287:2973-82.

[31] Hartling L, Hamm MP, Fernandes RM, Dryden DM, Vandermeer B. Quantifying bias in randomized controlled trials in child health: a meta-epidemiological study. PloS One 2014; 9:e88008.

[32] van Tulder MW, Suttorp M, Morton S, Bouter LM, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine 2009; 34:1685-92.

[33] Nuesch E, Reichenbach S, Trelle S, Rutjes AWS, Liewald K, Sterchi R, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: A meta-epidemiologic study. Arthritis Rheum 2009; 61:1633-41.

[34] Armijo-Olivo S, Saltaji H, da Costa BR, Fuentes J, Ha C, Cummings GG. What is the influence of randomisation sequence generation and allocation concealment on treatment effects of physical therapy trials? A meta-epidemiological study. BMJ Open 2015; 5:e008562.

[35] Berkman ND, Santaguida PL, Viswanathan M, Morton SC. The empirical evidence of bias in trials measuring treatment differences. Rockville (MD), Agency for Healthcare Research and Quality (US); 2014 Sep. Report No.:14-EHC050-EF.

[36] Hempel S, Miles J, Suttorp MJ, Wang Z, Johnsen B, Morton S, et al. Detection of associations between trial quality and effect sizes. Rockville (MD), Agency for Healthcare Research and Quality (US); 2012 Jan. Report No.:12-EHC010-EF.

[37] Armijo-Olivo S, Fuentes J, da Costa BR, Saltaji H, Ha C, Cummings GG. Blinding in physical therapy trials and its association with treatment effects: A meta-epidemiological study. Am J Phys Med Rehabil 2016 May 4. [Epub ahead of print].

[38] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273:408-12.

[39] Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 2003; 7:1-76.

[40] Hrobjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ 2012; 344:e1119.

[41] Pandis N. The effect size. Am J Orthod Dentofacial Orthop 2012; 142:739-40.

[42] Fenwick J, Needleman IG, Moles DR. The effect of bias on the magnitude of clinical outcomes in periodontology: a pilot study. J Clin Periodontol 2008; 35:775-82.

[43] Kahan BC, Rehal S, Cro S. Blinded outcome assessment was infrequently used and poorly reported in open trials. PloS One 2015; 10:e0131926.

[44] Mhaskar R, Djulbegovic B, Magazin A, Soares HP, Kumar A. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. J Clin Epidemiol 2012; 65:602-9.

[45] Nuesch E, Trelle S, Reichenbach S, Rutjes AW, Burgi E, Scherer M, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. BMJ 2009; 339:b3244.

[46] Altman DG. Missing outcomes in randomized trials: addressing the dilemma. Open Med 2009 May 12; 3(2):e51-3.

[47] Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 2003; 326:1167-70.

[48] Bero L. Industry sponsorship and research outcome: a Cochrane review. JAMA Intern Med 2013 Apr 8; 173(7):580-1.

[49] Lundh A, Sismondo S, Lexchin J, Busuioc OA, Bero L. Industry sponsorship and research outcome. Cochrane Database Syst Rev 2012; 12:MR000033.

[50] Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? JAMA 2003; 290:921-8.

[51] Naci H, Dias S, Ades AE. Industry sponsorship bias in research findings: a network meta-analysis of LDL cholesterol reduction in randomised trials of statins. BMJ 2014; 349:g5741.

[52] Naci H, Dias S, Ades T. No evidence of industry sponsorship bias in statin trials. BMJ 2014; 349:g6579.

[53] Popelut A, Valet F, Fromentin O, Thomas A, Bouchard P. Relationship between sponsorship and failure rate of dental implants: a systematic approach. PloS One 2010; 5:e10274.

[54] Schwendicke F, Tu YK, Blunck U, Paris S, Gostemeyer G. Effect of industry sponsorship on dental restorative trials. J Dent Res 2016; 95:9-16.

[55] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009; 62:1006-12.

[56] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008; 336:924-6.

Chapter 10

Conclusions

The following conclusions can be drawn, based on the findings of the four-phase study contributing to this dissertation:

- Epidemiological and descriptive characteristics of the 1188 oral health systematic reviews, encompassing the nine dental specialties included in the register of reviews, and varied across the nine dental specialties and by review category (Cochrane vs. non-Cochrane).

- A study of systematic reviews of oral health trials showed that risk of bias was not assessed in a considerable portion of reviews published between 1991 and 2014. Reviews published in dental journals were less likely to assess risk of bias of individual trials than reviews published in nondental journals. While the Cochrane risk of bias tool was the most commonly used, no tool has been specifically designed for assessing the methodologic quality of oral health trials.

- An examination of change over time in the state of 540 oral health randomized trials showed that methodological quality and reporting quality were, in general, substandard, indicating a high potential for bias in oral health trials. The proportion of trials judged as having a low risk of bias did not exceed 60% in the majority of risk of bias domains. However, a significant increase over time in the proportion of trials judged as having a low risk of bias was identified in the majority of the quality items and risk of bias domains.

- Using our register of oral health reviews, we were able to quantify biases associated with the methodology employed in 540 oral health randomized trials included in a cohort of 64 meta-analyses. Associations were apparent between

inflated treatment effect size estimates and inadequacy in the conduct and reporting of the trials. Significant differences in treatment effect size estimates were identified in oral health trials based on inadequacy of sequence generation, inadequacy of allocation concealment, lack of patient and assessor blinding, inadequate reporting of the withdrawal/dropout rate, and inappropriate influence of funders. Trials with the inadequate conduct of the aforementioned quality items had significantly larger treatment effect size estimates than trials that employed adequate conduct in the same items. Although not statistically significant, a tendency toward exaggeration of treatment effect size was also found based on imbalances in cointerventions, inadequacy of compliance to treatment, incomplete outcome data and having dropout without performing intention-to-treat approach. On the contrary, baseline imbalance, caregiver blinding, an acceptable dropout rate (≤ 20%), selective outcome reporting, and analysis based on the intention-to-treat approach, were not associated with inflated or underestimated treatment effect size.

Based on this evidence, systematic reviewers may consider excluding trials (conducted in the domains of dental, oral, and craniofacial research) with inadequacy in the aforementioned quality criteria from meta-analyses, or, alternatively, should perform sensitivity analyses based on adequacy of these criteria. Furthermore, dental journal editors and reviewers should insist on adequate conduct and reporting of trials reports submitted for publication because of the expected impact of bias on treatment effect size estimates.

References

Agha R, Cooper D, Muir G. The reporting quality of randomised controlled trials in surgery: a systematic review. Int J Surg 2007; 5:413-22.

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.

Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? JAMA 2003; 290:921-8.

Altman DG, Dore CJ. Baseline comparisons in clinical trials. Lancet 1990; 335:1476.

Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001; 134:663-94.

Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001;134(8):663-694.

Altman DG. Missing outcomes in randomized trials: addressing the dilemma. Open Med 2009 May 12; 3(2):e51-3. 2009;3:e51-3.

American Dental Association - Database of Systematic Reviews. Available at: http://ebd.ada.org/SystematicReviews.aspx. Accessed May 2012.

American Dental Association (ADA). Specialty Definitions 2015. Available at: http://www.ada.org/495.aspx. Accessed July 2015.

Angelopoulou MV, Vlachou V, Halazonetis DJ. Pharmacological management of pain during orthodontic treatment: a meta-analysis. Orthod Craniofac Res 2012; 15:71-83.

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; 39:1097-113.

Armijo-Olivo S, Cummings GG, Fuentes J, Saltaji H, Ha C, Chisholm A, et al. Identifying items to assess methodological quality in physical therapy trials: a factor analysis. Phys Ther 2014; 94:1272-84.

Armijo-Olivo S, da Costa BR, Cummings GG, et al. PEDro or Cochrane to assess the quality of clinical trials? A meta-epidemiological study. PloS One 2015;10(7):e0132634.

Armijo-Olivo S, Fuentes CJ, Ospina M, Saltaji H, Hartling L. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: A descriptive analysis. BMC Med Res Methodol 2013; 13:1-19.

Armijo-Olivo S, Fuentes CJ, Rogers T, Hartling L, Saltaji H, Cummings G. How should we evaluate the risk of bias of physical therapy trials? A psychometric and meta-epidemiological approach towards developing guidelines for the design, conduct, and reporting of RCTs in physical therapy (PT) area: a study protocol. Syst Rev 2013; 2:88.

Armijo-Olivo S, Fuentes J, da Costa BR, Saltaji H, Ha C, Cummings GG. Blinding in physical therapy trials and its association with treatment effects: A meta-epidemiological study. Am J Phys Med Rehabil 2016 May 4.

Armijo-Olivo S, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. Phys Ther 2008; 88:156-75.

Armijo-Olivo S, Ospina M, da Costa BR, Egger M, Saltaji H, Fuentes CJ, et al. Poor reliability between cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. PloS One 2014; 9(5): e96920.

Armijo-Olivo S, Saltaji H, da Costa BR, Fuentes J, Ha C, Cummings GG. What is the influence of randomisation sequence generation and allocation concealment on treatment effects of physical therapy trials? A meta-epidemiological study. BMJ Open 2015; 5:e008562.

Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool—methodological research. J Eval Clin Pract 2012;18(1):12-18.

Atieh MA, Ibrahim HM, Atieh AH. Platform switching for marginal bone preservation around dental implants: a systematic review and meta-analysis. J Periodontol 2010; 81:1350-66.

Aziz T, Compton S, Nassar U, Matthews D, Ansari K, Flores-Mir C. Methodological quality and descriptive characteristics of prosthodontic-related systematic reviews. J Oral Rehabil. J Oral Rehabil. 2013; 40(4):263-78.

Bader J, Ismail A. Survey of systematic reviews in dentistry. J Am Dent Assoc. 2004; 135:464-73.

Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 2002; 287:2973-82.

Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. JAMA 2010; 303:1180-7.

Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996; 276:637-9.

Berchier CE, Slot DE, Haps S, Van der Weijden GA. The efficacy of dental floss in addition to a toothbrush on plaque and parameters of gingival inflammation: a systematic review. Int J Dent Hyg 2008; 6:265-79.

Berchier CE, Slot DE, Van der Weijden GA. The efficacy of 0.12% chlorhexidine mouthrinse compared with 0.2% on plaque accumulation and periodontal parameters: a systematic review. J Clin Periodontol 2010; 37:829-39.

Berger VW, Exner DV. Detecting selection bias in randomized clinical trials. Contr Clin Trials 1999; 20:319-27.

Berkman ND, Santaguida PL, Viswanathan M, Morton SC. The empirical evidence of bias in trials measuring treatment differences. Rockville (MD), Agency for Healthcare Research and Quality (US); 2014 Sep. Report No.:14-EHC050-EF.

Bero L. Industry sponsorship and research outcome: a Cochrane review. JAMA Intern Med 2013; 173:580-1.

Bialy L, Vandermeer B, Lacaze-Masmonteil T, Dryden DM, Hartling L. A meta-epidemiological study to examine the association between bias and treatment effects in neonatal trials. Evid Based Child Health 2014; 9:1052-9.

Bizzini M, Childs JD, Piva SR, Delitto A. Systematic review of the quality of randomized controlled trials for patellofemoral pain syndrome. J Orthop Sports Phys Ther 2003; 33:4-20.

Bizzini M, Childs JD, Piva SR. Systematic review of the quality of randomized controlled trials for patellofemoral pain syndrome. J Orthop Sports Phys Ther 2003; 33:4-20.

Bjordal JM, Bensadoun RJ, Tuner J, Frigo L, Gjerde K, Lopes-Martins RA. A systematic review with meta-analysis of the effect of low-level laser therapy (LLLT) in cancer therapy-induced oral mucositis. Support Care Cancer 2011; 19:1069-77.

Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med 2008; 148:295-309.

Boutron I, Ravaud P. Classification systems to improve assessment of risk of bias. J Clin Epidemiol 2012;65(3):236-238.

Bow S, Klassen J, Chisholm A, Tjosvold L, Thomson D, Klassen TP, et al. A descriptive analysis of child-relevant systematic reviews in the Cochrane Database of Systematic Reviews. BMC Pediatrics 2010; 10:34.

Brito JP, Tsapas A, Griebeler ML, Wang Z, Prutsky GJ, Domecq JP, et al. Systematic reviews supporting practice guideline recommendations lack protection against bias. J Clin Epidemiol 2013; 66:633-8.

Brocklehurst P, Tickle M, Glenny AM, Lewis MA, Pemberton MN, Taylor J, et al. Systemic interventions for recurrent aphthous stomatitis (mouth ulcers). Cochrane Database Syst Rev 2012; 9:CD005411.

Cairo F, Pagliaro U, Nieri M. Treatment of gingival recession with coronally advanced flap procedures: a systematic review. J Clin Periodontol 2008; 35:136-62.

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:e441-57.

Chalmers TC, Smith H Jr, Blackburn B, et al. A method for assessing the quality of a randomized control trial. Control Clin Trials 1981;2(1):31-49.

Chambrone L, Chambrone D, Pustiglioni FE, Chambrone LA, Lima LA. Can subepithelial connective tissue grafts be considered the gold standard procedure in the treatment of Miller Class I and II recession-type defects? J Dent 2008; 36:659-71.

Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 2004; 291:2457-65.

Cohen J. Chapter 1: The Concepts of Power Analysis. In: Cohen J, editor. Statistical Power Analysis for the Behavioral Sciences. Second ed. Hillsdale, New Jersey: Academic Press, Inc. 1988. p. 1-17.

Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000; 342:1887-92.

Cook DA, West CP. Conducting systematic reviews in medical education: A stepwise approach. Med Educ 2012; 46:943-52.

Coulthard P, Bailey E, Esposito M, Furness S, Renton TF, Worthington HV. Surgical techniques for the removal of mandibular wisdom teeth. Cochrane Database Syst Rev 2014; 7:CD004345.

Cunha-Cruz J, Stout JR, Heaton LJ, Wataha JC. Dentin hypersensitivity and oxalates: a systematic review. J Dent Res 2011; 90:304-10.

Darre L, Vergnes JN, Gourdy P, Sixou M. Efficacy of periodontal treatment on glycaemic control in diabetic patients: A meta-analysis of interventional studies. Diabetes Metab 2008; 34:497-506.

Davies RM, Ellwood RP, Davies GM. The effectiveness of a toothpaste containing triclosan and polyvinyl-methyl ether maleic acid copolymer in improving plaque control and gingival health: a systematic review. J Clin Periodontol 2004; 31:1029-33.

De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. Lancet 2004; 364:911-2.

De Angelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Is this clinical trial fully registered? —A statement from the International Committee of Medical Journal Editors. N Engl J Med 2005; 352:2436-8.

de Boer MR, Waterlander WE, Kuijper LD, Steenhuis IH, Twisk JW. Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. Int J Behav Nutr Phys Act 2015; 12:4.

De Vet HC, De Bie RA, Van Der Heijden GJMG, Verhagen AP, Sijpkes P, Knipschild PG. Systematic reviews on the basis of methodological criteria. Physiotherapy 1997; 83:284-9.

Deacon SA, Glenny AM, 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.

Dechartres A, Charles P, Hopewell S, Ravaud P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. J Clin Epidemiol 2011; 64:136-44.

Del Fabbro M, Bortolin M, Taschieri S, Weinstein R. Is platelet concentrate advantageous for the surgical treatment of periodontal diseases? A systematic review and meta-analysis. J Periodontol 2011; 82:1100-11.

Delgado-Rodriguez M, Llorca J. Bias. J Epidemiol Community Health 2004; 58:635-41.

DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: An update. Contemp Clin Trials 2007; 28:105-14.

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

Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, et al. Need for expertise based randomised controlled trials. BMJ 2005; 330:88.

Devereaux PJ, Choi PTL, El-Dika S, Bhandari M, Montori VM, Schünemann HJ, et al. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. J Clin Epidemiol 2004; 57:1232-6.

Doig GS, Simpson F. Randomization and allocation concealment: a practical guide for researchers. J Crit Care 2005; 20:187-91.

Dumbrigue HB, Al-Bayat MI, Ng CC, Wakefield CW. Assessment of bias in methodology for randomized controlled trials published on implant dentistry. J Prosthodont 2006; 15:257-63.

Dunst CJ, Hamby DW. Guide for calculating and interpreting effect sizes and confidence intervals in intellectual and developmental disability research studies. J Intellect Dev Disabil 2012; 37:89-99.

Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PloS One 2008;3: e3081.

Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias – an updated review. PloS One 2013; 8:e66844.

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

Ebrahim S, Montoya L, Busse JW, Carrasco-Labra A, Guyatt GH. The effectiveness of splint therapy in patients with temporomandibular disorders: a systematic review and meta-analysis. J Am Dent Assoc 2012; 143:847-57.

Edwards P, Clarke M, DiGuiseppi C, Pratap S, Roberts I, Wentz R. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. Stat Med 2002; 21:1635-40.

Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 2003; 7:1-76.

Elkins MR. Assessing baseline comparability in randomised trials. J Physiother 2015; 61:228-30.

Emami E, Heydecke G, Rompre 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; 20:533-44.

Engebretson S, Kocher T. Evidence that periodontal treatment improves diabetes outcomes: a systematic review and meta-analysis. J Periodontol 2013; 84:S153-69.

Epstein JB, Hong C, Logan RM, Barasch A, Gordon SM, Oberle-Edwards L, et al. A systematic review of orofacial pain in patients receiving cancer therapy. Support Care Cancer 2010; 18:1023-31.

Esposito M, Grusovin MG, Maghaireh H, Worthington HV. Interventions for replacing missing teeth: different times for loading dental implants. Cochrane Database Syst Rev 2013; 3:CD003878.

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.

Esposito M, Murray-Curtis L, Grusovin MG, Coulthard P, Worthington HV. Interventions for replacing missing teeth: different types of dental implants. Cochrane Database Syst Rev 2007: CD003815.

Faggion CM, Huda F, Wasiak J. Use of methodological tools for assessing the quality of studies in periodontology and implant dentistry: a systematic review. J Clin Periodontol 2014;41(6):625-631.

Faggion CM, Jr., Wu YC, Scheidgen M, Tu YK. Effect of risk of bias on the effect size of meta-analytic estimates in randomized controlled trials in periodontology and implant dentistry. PloS One 2015; 10:e0139030.

Fenwick J, Needleman IG, Moles DR. The effect of bias on the magnitude of clinical outcomes in periodontology: A pilot study. J Clin Periodontol 2008; 35:775-82.

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; 41:265-70.

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. Eur J Orthod 2013; 35(2):244-8.

Furlan AD, Pennick V, Bombardier C, Van Tulder M. 2009 Updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine 2009; 34:1929-41.

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.

Giannobile WV. Improving clinical trials in dentistry. J Dent Res 2015; 94:6S-7S.

Glenny AM, Esposito M, Coulthard P, Worthington HV. The assessment of systematic reviews in dentistry. Eur J Oral Sci 2003; 111:85-92.

Gluud LL. Bias in clinical intervention research. Am J Epidemiol 2006; 163:493-501.

Greenland S. Quality scores are useless and potentially misleading: reply to “re: a critical look at some popular analytic methods.” Am J Epidemiol 1994;140(3):300-301.

Guyatt G, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B, et al. Grading strength of recommendations and quality of evidence in clinical guidelines: report from the American College of Chest Physicians task force. Chest 2006; 129:174-81.

Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011; 64:383-94.

Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011; 64:395-400.

Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008; 336:924-6.

Haahr MT, Hrobjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. Clin Trials 2006; 3:360-5.

Hadorn DC, Baker D, Hodges JS, Hicks N. Rating the quality of evidence for clinical practice guidelines. J Clin Epidemiol 1996;49(7):749-754.

Hamm MP, Hartling L, Milne A, Tjosvold L, Vandermeer B, Thomson D, et al. A descriptive analysis of a representative sample of pediatric randomized controlled trials published in 2007. BMC Pediatr 2010; 10:96.

Haps S, Slot DE, Berchier CE, Van der Weijden GA. The effect of cetylpyridinium chloride-containing mouth rinses as adjuncts to toothbrushing on plaque and parameters of gingival inflammation: a systematic review. Int J Dent Hyg 2008; 6:290-303.

Harrison JE. Clinical trials in orthodontics II: assessment of the quality of reporting of clinical trials published in three orthodontic journals between 1989 and 1998. J Orthod 2003; 30:309-15.

Hartling L, Bond K, Vandermeer B, Seida J, Dryden D, Rowe B. Applying the Risk of Bias Tool in a systematic review of combination longacting betaagonists and inhaled corticosteroids for persistent asthma. PLoS Med 2011; 6.

Hartling L, Hamm MP, Fernandes RM, Dryden DM, Vandermeer B. Quantifying bias in randomized controlled trials in child health: a meta-epidemiological study. PloS One 2014; 9:e88008.

Hartling L, Hamm MP, Milne A, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. J Clin Epidemiol 2012;66(9):973-981.

Hartling L, Hamm MP, Milne A, Vandermeer B, Santaguida PL, Ansari M, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. J Clin Epidemiol 2013; 66:973-81.

Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. BMJ 2009;339:b4102.

Hasson H, Ismail AI, Neiva G. Home-based chemically-induced whitening of teeth in adults. Cochrane Database Syst Rev 2006: CD006202.

Hempel S, Miles J, Suttorp MJ, Wang Z, Johnsen B, Morton S, et al. Detection of associations between trial quality and effect sizes. Rockville (MD), Agency for Healthcare Research and Quality (US); 2012 Jan. Report No.:12-EHC010-EF.

Hempel S, Suttorp M, Miles J. Empirical evidence of associations between trial quality and effect size. Rockville (MD), Agency for Healthcare Research and Quality (US); 2011 Jun. Report No.:11-EHC045-EF.

Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. J Clin Epidemiol 2006; 59(12):1249-56.

Higgins J, Altman D. Chapter 8: Assessing risk of bias in included studies. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions version 50. Chichester, UK: John Wiley & Sons, Ltd.; 2008.

Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011; 343:d5928.

Higgins JP, Green S, eds. Cochrane handbook for systematic reviews of interventions version 5.0.0. Cochrane Collaboration. 2008.

Hoekema A, Stegenga B, De Bont LG. Efficacy and co-morbidity of oral appliances in the treatment of obstructive sleep apnea-hypopnea: a systematic review. Crit Rev Oral Biol Med 2004; 15:137-55.

Hooft L, Korevaar DA, Molenaar N, Bossuyt PM, Scholten RJ. Endorsement of ICMJE's Clinical Trial Registration Policy: a survey among journal editors. Neth J Med 2014; 72:349-55.

Hopewell S, Clarke M, Lusher A, Lefebvre C, Westby M. A comparison of handsearching versus MEDLINE searching to identify reports of randomized controlled trials. Stat Med 2002; 21:1625-34.

Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ 2010; 340:c723.

Hrobjartsson A, Boutron I, Turner L, Altman DG, Moher D. Assessing risk of bias in randomised clinical trials included in Cochrane Reviews: the why is easy, the how is a challenge. Cochrane Database Syst Rev 2013; 4:ED000058.

Hrobjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ 2012; 344:e1119.

Iqbal A, Glenny AM. General dental practitioners' knowledge of and attitudes towards evidence based practice. Br Dent J 2002; 193:587-91.

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

Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001; 323:42-6.

Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999; 282:1054-60.

Kahan BC, Rehal S, Cro S. Blinded outcome assessment was infrequently used and poorly reported in open trials. PloS One 2015; 10:e0131926.

Kahan BC, Rehal S, Cro S. Risk of selection bias in randomised trials. Trials 2015; 16:405.

Karanicolas PJ, Farrokhyar F, Bhandari M. Practical tips for surgical research: blinding: who, what, when, why, how? Can J Surg 2010; 53:345-8.

Kim AJ, Lo AJ, Pullin DA, Thornton-Johnson DS, Karimbux NY. Scaling and root planing treatment for periodontitis to reduce preterm birth and low birth weight: a systematic review and meta-analysis of randomized controlled trials. J Periodontol 2012; 83:1508-19.

Kinaia BM, Steiger J, Neely AL, Shah M, Bhola M. Treatment of Class II molar furcation involvement: meta-analyses of reentry results. J Periodontol 2011; 82:413-28.

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(12):1181-1187.

Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2001; 135:982-9.

Knottnerus JA, Tugwell P. Selection-related bias, an ongoing concern in doing and publishing research. J Clin Epidemiol 2014; 67:1057-8.

Koletsi D, Karagianni A, Pandis N, Makou M, Polychronopoulou A, Eliades T. Are studies reporting significant results more likely to be published? Am J Orthod Dentofacial Orthop 2009; 136:632 e1-5.

Koletsi D, Pandis N, Fleming PS. Sample size in orthodontic randomized controlled trials: are numbers justified? Eur J Orthod 2013;1:1.

Koletsi D, Pandis N, Polychronopoulou A, Eliades T. Does published orthodontic research account for clustering effects during statistical data analysis? Eur J Orthod 2012; 34:287-92.

Koletsi D, Pandis N, Polychronopoulou A, et al. Mislabeling controlled clinical trials (CCTs) as ‘randomized clinical trials (RCTs)’ in dental specialty journals. J Evid Based Dent Pract 2012;12:124–30.

Koletsi D, Pandis N, Polychronopoulou A, et al. What’s in a title? An assessment of whether randomized controlled trial in a title means that it is one. Am J Orthod Dentofacial Orthop 2012;141:679–85.

Koletsi D, Spineli LM, Lempesi E, Pandis N. Risk of bias and magnitude of effect in orthodontic randomized controlled trials: a meta-epidemiological review. Eur J Orthod 2016; 38(3):308-12.

Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 2007: MR000012.

Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. J Natl Cancer Inst 2005 Jul 20; 97(14):1043-55.

Lempesi E, Koletsi D, Fleming PS, Pandis N. The reporting quality of randomized controlled trials in orthodontics. J Evid Based Dent Pract 2014; 14(2):46-52.

Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 2003; 326:1167-70.

Li W, Xiao L, Hu J. The use of enamel matrix derivative alone versus in combination with bone grafts to treat patients with periodontal intrabony defects: a meta-analysis. J Am Dent Assoc 2012; 143:e46-56.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009; 339:b2700.

Lin GH, Chan HL, Bashutski JD, Oh TJ, Wang HL. The effect of flapless surgery on implant survival and marginal bone level: a systematic review and meta-analysis. J Periodontol 2014; 85:e91-103.

Lin PY, Cheng YW, Chu CY, Chien KL, Lin CP, Tu YK. In-office treatment for dentin hypersensitivity: a systematic review and network meta-analysis. J Clin Periodontol 2013; 40:53-64.

Linde K, Scholz M, Ramirez G, Clausius N, Melchart D, Jonas WB. Impact of study quality on outcome in placebo-controlled trials of homeopathy. J Clin Epidemiol 1999; 52:631-6.

Lourenco-Matharu L, Ashley PF, Furness S. Sedation of children undergoing dental treatment. Cochrane Database Syst Rev 2012; 3:CD003877.

Lundh A, Sismondo S, Lexchin J, Busuioc OA, Bero L. Industry sponsorship and research outcome. Cochrane Database Syst Rev 2012; 12:MR000033.

Major MP, Major PW, Flores-Mir C. An evaluation of search and selection methods used in dental systematic reviews published in English. J Am Dent Assoc 2006; 137:1252-7.

Major MP, Major PW, Flores-Mir C. Benchmarking of reported search and selection methods of systematic reviews by dental speciality. Evid Based Dent 2007;8:66–70.

Marinho VC, Higgins JP, Logan S, Sheiham A. Fluoride gels for preventing dental caries in children and adolescents. Cochrane Database Syst Rev 2002: CD002280.

Marinho VC, Higgins JP, Logan S, Sheiham A. Fluoride mouthrinses for preventing dental caries in children and adolescents. Cochrane Database Syst Rev 2003: CD002284.

Marinho VC, Higgins JP, Logan S, Sheiham A. Fluoride varnishes for preventing dental caries in children and adolescents. Cochrane Database Syst Rev 2002: CD002279.

Marinho VC, Higgins JP, Sheiham A, Logan S. Combinations of topical fluoride (toothpastes, mouthrinses, gels, varnishes) versus single topical fluoride for preventing dental caries in children and adolescents. Cochrane Database Syst Rev 2004: CD002781.

Marinho VC, Higgins JP, Sheiham A, Logan S. Fluoride toothpastes for preventing dental caries in children and adolescents. Cochrane Database Syst Rev 2003: CD002278.

Markiewicz MR, Brady MF, Ding EL, Dodson TB. Corticosteroids reduce postoperative morbidity after third molar surgery: a systematic review and meta-analysis. J Oral Maxillofac Surg 2008; 66:1881-94.

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.

McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. BMJ 2002; 324:1448-51.

Mhaskar R, Djulbegovic B, Magazin A, Soares HP, Kumar A. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. J Clin Epidemiol 2012; 65:602-9.

Millett D. Bias in systematic reviews? J Orthod 2011; 38:158-60.

Moher D, Cook DJ, Jadad AR, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. Health Technol Assess 1999;3:i–iv, 1–98.

Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10:28–55.

Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. Int J Surg 2011.

Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010; 63:e1-37.

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

Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010; 8:336-41.

Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009; 62:1006-12.

Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609-13.

Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 2001; 285:1987-91.

Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. J Am Podiatr Med Assoc 2001; 91:437-42.

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

Moher D, Tsertsvadze A. Systematic reviews: when is an update an update? Lancet 2006; 367:881-3.

Molnar FJ, Hutton B, Fergusson D. Does analysis using "last observation carried forward" introduce bias in dementia research? CMAJ 2008; 179:751-3.

Montenegro R, Needleman I, Moles D, Tonetti M. Quality of RCTs in periodontology—a systematic review. J Dent Res 2002; 81:866-70.

Moseley AM, Herbert RD, Maher CG, Sherrington C, Elkins MR. Reported quality of randomized controlled trials of physiotherapy interventions has improved over time. J Clin Epidemiol 2011; 64:594-601.

Moseley AM, Herbert RD, Sherrington C, Maher CG. Evidence for physiotherapy practice: a survey of the Physiotherapy Evidence Database (PEDro). Aust J Physiother 2002; 48:43-9.

Moustafa K. The disaster of the impact factor. Sci Eng Ethics 2015;21(1):139-142.

Murphy KG, Gunsolley JC. Guided tissue regeneration for the treatment of periodontal intrabony and furcation defects. A systematic review. Ann Periodontol 2003; 8:266-302.

Naci H, Dias S, Ades T. No evidence of industry sponsorship bias in statin trials. BMJ 2014; 349:g6579.

Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev Camb Philos Soc 2007; 82:591-605.

Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. BMJ 1997; 315:617-9.

Needleman I, Moher D, Altman DG, Schulz KF, Moles DR, Worthington H. Improving the clarity and transparency of reporting health research: a shared obligation and responsibility. J Dent Res 2008; 87:894-5.

Needleman I, Moles DR, Worthington H. Evidence-based periodontology, systematic reviews and research quality. Periodontol 2000. 2005; 37:12-28.

Needleman I, Worthington H, Moher D, Schulz K, Altman DG. Improving the completeness and transparency of reports of randomized trials in oral health: the CONSORT statement. Am J Dent 2008;21(1):7-12.

Needleman IG, Worthington HV, Giedrys-Leeper E, Tucker RJ. Guided tissue regeneration for periodontal infra-bony defects. Cochrane Database Syst Rev 2006: CD001724.

Nieri M, Clauser C, Franceschi D, Pagliaro U, Saletta D, Pini-Prato G. Randomized clinical trials in implant therapy: relationships among methodological, statistical, clinical, paratextual features and number of citations. Clin Oral Implants Res 2007; 18:419-31.

Nuesch E, Reichenbach S, Trelle S, Rutjes AW, Liewald K, Sterchi R, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. Arthritis Rheum 2009; 61:1633-41.

Nuesch E, Trelle S, Reichenbach S, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. BMJ 2009;339:b3244.

Organizations that have endorsed or that are using GRADE. http://www.gradeworkinggroup.org/society/index.htm. Accessed Feb 23, 2015.

Pandis N, Fleming PS, Worthington H, Salanti G. The quality of the evidence according to GRADE is predominantly low or very low in oral health systematic reviews. PLoS One 2015;10(7):e0131644.

Pandis N, Polychronopoulou A, Eliades T. An assessment of quality characteristics of randomised control trials published in dental journals. J Dent 2010; 38 (9):713-21.

Pandis N, Polychronopoulou A, Eliades T. Randomization in clinical trials in orthodontics: its significance in research design and methods to achieve it. Eur J Orthod 2011; 33:684-90.

Pandis N, Polychronopoulou A, Madianos P, Makou M, Eliades T. Reporting of research quality characteristics of studies published in 6 major clinical dental specialty journals. J Evid Based Dent Pract 2011; 11:75-83.

Pandis N, Turpin DL. Enhancing CONSORT compliance for improved reporting of randomized controlled trials. Am J Orthod Dentofacial Orthop 2014; 145(1):1.

Pandis N. Blinding or masking. Am J Orthod Dentofacial Orthop 2012; 141:389-90.

Pandis N. Randomization. Part 1: sequence generation. Am J Orthod Dentofacial Orthop 2011; 140:747-8.

Pandis N. Randomization. Part 3: allocation concealment and randomization implementation. Am J Orthod Dentofacial Orthop 2012; 141:126-8.

Pandis N. The effect size. Am J Orthod Dentofacial Orthop 2012; 142:739-40.

Panic N, Leoncini E, de Belvis G, Ricciardi W, Boccia S. Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. PloS One 2013;8(12):e83138.

Papageorgiou SN, Papadopoulos MA, Athanasiou AE. Evaluation of methodology and quality characteristics of systematic reviews in orthodontics. Orthodontics & Craniofacial Research 2011; 14:116-37.

Paraskevas S, van der Weijden GA. A review of the effects of stannous fluoride on gingivitis. J Clin Periodontol 2006; 33:1-13.

Pavia M, Nobile CG, Angelillo IF. Meta-analysis of local tetracycline in treating chronic periodontitis. J Periodontol 2003; 74:916-32.

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

Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gotzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int J Epidemiol 2007; 36:847-57.

Piromchai P, Thanaviratananich S, Laopaiboon M. Systemic antibiotics for chronic rhinosinusitis without nasal polyps in adults. Cochrane Database Syst Rev 2011: CD008233.

Polychronopoulou A, Pandis N, Eliades T. Appropriateness of reporting statistical results in orthodontics: the dominance of P values over confidence intervals. Eur J Orthod 2011; 33:22-5.

Popelut A, Valet F, Fromentin O, Thomas A, Bouchard P. Relationship between sponsorship and failure rate of dental implants: a systematic approach. PloS One 2010; 5:e10274.

Poulsen S, Errboe M, Lescay Mevil Y, Glenny AM. Potassium containing toothpastes for dentine hypersensitivity. Cochrane Database Syst Rev 2006: CD001476.

Reveiz L, Chapman E, Asial S, Munoz S, Bonfill X, Alonso-Coello P. Risk of bias of randomized trials over time. J Clin Epidemiol 2015; 68:1036-45.

Reynolds MA, Aichelmann-Reidy ME, Branch-Mays GL, Gunsolley JC. The efficacy of bone replacement grafts in the treatment of periodontal osseous defects. A systematic review. Ann Periodontol 2003; 8:227-65.

Richards D. Critically appraising systematic reviews. Evid Based Dent 2010; 11:27-9.

Riley P, Lamont T. Triclosan/copolymer containing toothpastes for oral health. Cochrane Database Syst Rev 2013; 12:CD010514.

Robinson PG, Deacon SA, Deery C, Heanue M, Walmsley AD, Worthington HV, et al. Manual versus powered toothbrushing for oral health. Cochrane Database Syst Rev 2005: CD002281.

Sahrmann P, Puhan MA, Attin T, Schmidlin PR. Systematic review on the effect of rinsing with povidone-iodine during nonsurgical periodontal therapy. J Periodontal Res 2010; 45:153-64.

Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, Flores-Mir C. Methodological characteristics and treatment effect sizes in oral health randomised controlled trials: Is there a relationship? Protocol for a meta-epidemiological study. BMJ Open 2014; 4:e004527.

Saltaji H, Cummings GG, Armijo-Olivo S, Major MP, Amin M, Major PW, et al. A descriptive analysis of oral health systematic reviews published 1991-2012: cross sectional study. PloS One 2013; 8:e74545.

Saltaji H, Flores-Mir C, Major PW. Letter to the editor, "Clinical trial registration in oral health journals." J Dent Res 2015; 94:103S.

Saltaji H. Bias is the key challenge in orthodontic research. Am J Orthod Dentofacial Orthop 2015; 148:8.

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.

Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med 2012; 157:429-38.

Savovic J, Weeks L, Sterne JA, Turner L, Altman DG, Moher D, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. Syst Rev 2014; 3:37.

Schulz KF, Altman DG, Moher D. Allocation concealment in clinical trials. JAMA 2002; 288:2406-7; author reply 8-9.

Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273:408-12.

Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. Lancet 2002; 359:614-8.

Schulz KF, Grimes DA. Blinding in randomised trials: Hiding who got what. Lancet 2002; 359:696-700.

Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. Lancet 2002; 359:515-9.

Schulz KF. Randomized controlled trials. Clin Obstet Gynecol 1998; 41:245-56.

Schwendicke F, Tu YK, Blunck U, Paris S, Gostemeyer G. Effect of industry sponsorship on dental restorative trials. J Dent Res 2016; 95:9-16.

Seehra J, Pandis N, Koletsi D, Fleming PS. Use of quality assessment tools in systematic reviews was varied and inconsistent. J Clin Epidemiol 2016; 69:179-84 e5.

Seehra J, Wright NS, Polychronopoulou A, Cobourne MT, Pandis N. Reporting quality of abstracts of randomized controlled trials published in dental specialty journals. J Evid Based Dent Pract 2013; 13:1-8.

Sgolastra F, Petrucci A, Gatto R, Monaco A. Effectiveness of systemic amoxicillin/metronidazole as an adjunctive therapy to full-mouth scaling and root planing in the treatment of aggressive periodontitis: a systematic review and meta-analysis. J Periodontol 2012; 83:731-43.

Sgolastra F, Petrucci A, Severino M, Gatto R, Monaco A. Lasers for the treatment of dentin hypersensitivity: a meta-analysis. J Dent Res 2013; 92:492-9.

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; 40:514-26.

Shea B, Moher D, Graham I, Pham B, Tugwell P. A comparison of the quality of Cochrane reviews and systematic reviews published in paper-based journals. Eval Health Prof. 2002; 25:116-29.

Sherrington C, Herbert RD, Maher CG, Moseley AM. PEDro. A database of randomized trials and systematic reviews in physiotherapy. Man Ther 2000; 5:223-6.

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.

Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. Stat Med 2007; 26:2745-58.

Smail-Faugeron V, Fron-Chabouis H, Durieux P. Clinical trial registration in oral health journals. J Dent Res 2015; 94:8S-13S.

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 Contr Clin Trials. J Periodontol 2012; 83:453-64.

Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. BMJ 2001; 323:101-5.

Sterne JAC, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med 2002; 21:1513-24.

Stoecklin-Wasmer C, Rutjes AW, da Costa BR, Salvi GE, Juni P, Sculean A. Absorbable collagen membranes for periodontal regeneration: a systematic review. J Dent Res 2013; 92:773-81.

Stoeken JE, Paraskevas S, van der Weijden GA. The long-term effect of a mouthrinse containing essential oils on dental plaque and gingivitis: a systematic review. J Periodontol 2007; 78:1218-28.

Suarez-Almazor ME, Belseck E, Homik J, Dorgan M, Ramos-Remus C. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. Control Clin Trials 2000; 21:476-87.

Sutherland S, Matthews DC. Emergency management of acute apical periodontitis in the permanent dentition: a systematic review of the literature. J Can Dent Assoc 2003; 69:160.

Swiglo BA, Murad MH, Schunemann HJ, Kunz R, Vigersky RA, Guyatt GH, et al. 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; 93:666-73.

Teeuw WJ, Slot DE, Susanto H, Gerdes VE, Abbas F, D'Aiuto F, et al. Treatment of periodontitis improves the atherosclerotic profile: a systematic review and meta-analysis. J Clin Periodontol 2014; 41:70-9.

Thomson D, Hartling L, Cohen E, Vandermeer B, Tjosvold L, Klassen TP. Controlled trials in children: quantity, methodological quality and descriptive characteristics of pediatric controlled trials published 1948–2006. PloS One 2010; 5(9).

Trees AH, Howe TE, Grant M, Gray HG. WITHDRAWN: Exercise for treating anterior cruciate ligament injuries in combination with collateral ligament and meniscal damage of the knee in adults. Cochrane Database Syst Rev 2011: CD005961.

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.

Turner L, Galipeau J, Garritty C, Manheimer E, Wieland LS, Yazdi F, et al. An Evaluation of Epidemiological and Reporting Characteristics of Complementary and Alternative Medicine (CAM) Systematic Reviews (SRs). PloS One 2013; 8:e53536.

Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Syst Rev 2012; 1:60.

Turpin DL. Challenges faced by journal editors. Am J Orthod Dentofacial Orthop 2015; 147:285-9.

Turpin DL. The need to register clinical trials in orthodontics. Am J Orthod Dentofacial Orthop 2006; 130:429-30.

Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials? Evaluation of risk of bias assessments in systematic reviews. BMJ 2013; 346:f1798.

Van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. Spine 2003; 28:1290-9.

Van Tulder MW, Assendelft WJJ, Koes BW, Bouter LM, Bombardier C, Nachemson AL, et al. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for spinal disorders. Spine 1997; 22:2323-30.

Van Tulder MW, Suttorp M, Morton S, Bouter LM, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine 2009; 34:1685-92.

Vere J, Joshi R. Quality assessment of randomised controlled trials of dental implant surgery and prosthodontics published from 2004 to 2008: a systematic review. Clin Oral Implants Res 2011; 22:1338-45.

Verhagen AP, de Vet HC, de Bie RA, Boers M, van den Brandt PA. The art of quality assessment of RCTs included in systematic reviews. J Clin Epidemiol 2001; 54:651-4.

Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 1998; 51:1235-41.

Vignoletti F, Matesanz P, Rodrigo D, Figuero E, Martin C, Sanz M. Surgical protocols for ridge preservation after tooth extraction. A systematic review. Clin Oral Implants Res 2012; 23 Suppl 5:22-38.

Walsh T, Worthington HV, Glenny AM, 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.

Wide Boman U, Carlsson V, Westin M, Hakeberg M. Psychological treatment of dental anxiety among adults: a systematic review. Eur J Oral Sci 2013; 121:225-34.

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; 336:601-5.

Wright JT, Hanson N, Ristic H, Whall CW, Estrich CG, Zentz RR. Fluoride toothpaste efficacy and safety in children younger than 6 years: a systematic review. J Am Dent Assoc 2014; 145:182-9.

Yusof ZY, Han LJ, San PP, Ramli AS. Evidence-based practice among a group of Malaysian dental practitioners. J Dent Educ 2008; 72:1333-42.

Zini A, Krivoroutski Y, Vered Y. Primary prevention of dental erosion by calcium and fluoride: a systematic review. Int J Dent Hyg 2014; 12:17-24.

2014 Journal Citation Reports Science Edition. Thomson Reuters, 2015.

Appendices

Chapter 1: **Introduction**

Appendix 1.A. Letter to the editor, "Clinical Trial Registration in Oral Health Journals"
Appendix 1.B. Bias is the key challenge in orthodontic research

Chapter 2: **A descriptive analysis of oral health systematic reviews published 1991-2012: A cross sectional study**

Appendix 2.A. Search strategies and results from different electronic databases

Appendix 2.B. Continent and country of corresponding author of oral health systematic reviews
Appendix 2.C. Authors and affiliation of oral health systematic reviews

Appendix 2.D. Focus and interventions of oral health systematic reviews

Appendix 2.E. Study designs of studies included in oral health systematic reviews

Appendix 2.F. Number of included studies of oral health systematic reviews

Appendix 2.G. Number of studies that contributed data to the largest meta-analysis in oral health systematic reviews

Chapter 3: **Evaluation of risk of bias assessment of trials in systematic reviews of oral health interventions, 1991-2014: A methodology study**

Appendix 3.A. Search strategy used in the study

Chapter 5: **Randomized clinical trials in dentistry: Risk of bias and reported methodological quality over time, 1955-2013**

Appendix 5.A. Tools and items to assess quality of randomized trials

Appendix 5.B. Guidelines for evaluating quality assessment of trials based on most common tools used in health research
Appendix 5.C. The Cochrane Collaboration's tool for assessing risk of bias

Appendix 6.D. Guidelines for evaluating the risk of bias of trials

Appendix 1.A

Letter to the Editor, "Clinical Trial Registration in Oral Health Journals" *

Humam Saltaji, Carlos Flores-Mir, Paul W. Major

A copy of this peer-reviewed letter has been published in the J Dent Res (JDR); Saltaji et al., J Dent Res. 2015 Mar;94(3 Suppl):103S. doi: 10.1177/0022034514563954.

The article by Smaïl-Faugeron et al. sheds light on the need for specific mechanisms to inform oral health researchers about the importance of clinical trial registration [1]. We wholeheartedly agree with their plea and hope that this important report spurs action within the dental community. Moreover, we have identified 2 areas that could have strengthened the submission.

Trial registration has been proposed as a potential solution to prevent biased reporting in clinical trials [2]. It aimed at reducing 2 main sources of bias: publication bias and selective reporting [3], the latter having been included in one of the 6 domains of the Cochrane Collaboration Risk of Bias tool [4, 5]. The article in question, though, examined the risk of bias based on only 2 domains of the Risk of Bias tool (sequence generation and allocation concealment) that are not directly related to trial registration, while not considering the main domain of the tool (selective reporting) that assesses reporting bias. It is not clear why the authors chose to do so.

Furthermore, the authors classified journals into 3 categories, based on a screening performed in December 2013 using the journals' Web sites, while including trials published in 2013. This procedure might have created a sampling bias affecting the findings, because some trials might have been submitted in 2011/2012 when the rules were not in place, and published in 2013 when the rules were in effect. Last, the

authors could have confirmed with the journals directly about the specific date after which regulation for clinical trial registration was required.

Appendix 1.B

Bias is the key challenge in orthodontic research*

Humam Saltaji

A copy of this peer-reviewed letter has been published in the American Journal of Orthodontics and Dentofacial Orthopedics (AJODO); Saltaji. Am J Orthod Dentofacial Orthop. 2015 Jul;148(1):8. doi: 10.1016/j.ajodo.2015.04.014.

The timely and thoughtful editorial by Dr Turpin [6] sheds light on challenges encountered by orthodontic journal editors and reviewers when appraising the design, conduct, and reporting of research to facilitate the assessment of the study's quality. Progress has certainly been made with the *AJO-DO*'s adoption of the Consolidated Standards of Reporting Trials (CONSORT) statement to promote transparency and reporting quality of randomized controlled trials [7, 8]. This is critical because high-quality randomized controlled trials contribute greatly to the strength of a body of evidence assessed in systematic reviews, informing decisions by practitioners and practice guidelines.

As Dr Turpin noted, adoption of the CONSORT does not guarantee compliance; I argue further that relying on this approach may enable a false sense of security to reviewers and readers, that good reporting is sufficient for meeting methodologic quality standards, including a low risk of bias. Transparent reporting would, for example, fail to safeguard against recognizing important risks of bias such as selective outcome reporting by trial authors (ie, preferential reporting of beneficial results) who have failed to document their primary outcomes and analyses in an a priori fashion. Another bias that is fundamental to determining the strength of a body of evidence is publication bias, whereby studies having large beneficial effect sizes are published sooner and in journals with higher impacts. These biases have been shown

to overestimate the magnitude of the treatment effects of the clinical trials and can skew the overall conclusions in meta-analyses [9]. The most effective and efficient way to detect either of them is to review the trial registries.

To optimize methodologic quality assessment, promote submissions of clinical trials having a low risk of bias, and assist systematic review authors in the dental field in assessing publication bias, I urge the *AJO-DO* to consider the recommendations of the International Committee of Medical Journal Editors by implementing mandatory trial registration. This action was taken 10 years ago by 11 leading medical journals, lately by over 300 medical journals [10] and recently by a leading dental journal [11]. It was recently reported that only 23% of the randomized controlled trials published in 15 dental journals were registered [1].

Although Dr Turpin's stated concern of the challenge of conflicts of interest is appreciated, given the nature of orthodontic interventions, I suggest that bias should be considered the main challenge encountered and in need of more reflection.

Appendix 2.A. Search strategies and results from different electronic databases

Database	Search Strategy	Results
PubMed	((systematic review* OR meta-analys*)) AND (dent* OR tooth OR teeth OR orthodon* OR oral surg* OR endodon* OR periodon* OR prosthodon* OR pedodon* OR pediatric* AND dentistry OR paediatric* AND dentistry OR dent* AND public health OR oral pathology)	1505
Embase	(systematic review* or meta-analys*).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui] AND (dent* OR tooth OR teeth OR orthodon* OR oral surg* OR endodon* OR periodon* OR prosthodon* OR pedodon* OR pediatric* dentistry OR paediatric* dentistry OR dent* public health OR oral pathology).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui]	2196
MEDLINE (Ovid)	(systematic review* or meta-analys*).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui] AND (dent* OR tooth OR teeth OR orthodon* OR oral surg* OR endodon* OR periodon* OR prosthodon* OR pedodon* OR pediatric* dentistry OR paediatric* dentistry OR dent* public health OR oral pathology).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui]	1709
ISI Web of Science	Topic=(dent* OR tooth OR teeth OR orthodon* OR oral surg* OR endodon* OR periodon* OR prosthodon* OR pedodon* OR pediatric* AND dentistry OR paediatric* AND dentistry OR dent* AND public health OR oral pathology) AND Topic=(systematic review* OR meta-analys*)	1872
EMB Reviews-Cochrane Database of Systematic Reviews	(systematic review* or meta-analys*).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui] AND (dent* OR tooth OR teeth OR orthodon* OR oral surg* OR endodon* OR periodon* OR prosthodon* OR pedodon* OR pediatric* dentistry OR paediatric* dentistry OR dent* public health OR oral pathology).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui]	559
HealthSTAR	(systematic review* or meta-analys*).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui] AND (dent* OR tooth OR teeth OR orthodon* OR oral surg* OR endodon* OR periodon* OR prosthodon* OR pedodon* OR pediatric* dentistry OR paediatric* dentistry OR dent* public health OR oral pathology).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui]	1828
Total electronic databases searches		9669
Duplicates		2854
Final		6815

Appendix 2.B. Continent and country of corresponding author of oral health systematic reviews

	Overall (No. Overall=1188: NCRs=1062 & CRs=126)	**Oral Medicine & Oral Pathology** (No. Overall=162: NCRs=140 & CRs=22)	**Dental Public Health** (No. Overall=184: NCRs=163 & CRs=21)	**Prosthodontics & Restorative Dentistry** (No. Overall=198: NCRs=179 & CRs=19)	**Pediatric Dentistry** (No. Overall=50: NCRs=42 & CRs=8)	**Endodontics** (No. Overall=54: NCRs=47 & CRs=7)	**Periodontics** (No. Overall=212: NCRs=203 & CRs=9)	**Orthodontics &Dentofacial Orthopedics** (No. Overall=138: NCRs=123 & CRs=15)	**Oral and Maxillofacial Surgery** (No. Overall=159: NCRs=134 & CRs=25)	**Oral and Maxillo-facial Radiology** (No. Overall=31: NCRs=31 & CRs=0)
Continent of Corresponding Author, n (% total)										
Overall (Cochrane & Non-Cochrane Reviews)										
Europe	645 (54.3)	87 (53.7)	88 (47.8)	102 (51.5)	22 (44)	17 (31.5)	126(59.4)	81 (58.7)	106 (66.7)	16 (51.6)
North America	303 (25.5)	46 (28.4)	60 (32.6)	46 (23.2)	4 (8)	16 (29.6)	51 (24.1)	37 (26.8)	33 (20.8)	10 (32.3)
Asia	99 (8.3)	19 (11.7)	7 (3.8)	15 (7.6)	5 (10)	16 (29.6)	12 (5.7)	10 (7.5)	14 (8.8)	1 (3.2)
South America	61 (5.1)	4 (2.5)	6 (3.3)	11 (5.6)	8 (16)	2 (3.7)	17 (8.0)	7 (5.1)	4 (2.5)	2 (6.5)
Australia	47 (4.0)	2 (1.2)	8 (4.3)	17 (8.6)	9 (18)	3 (5.6)	6 (2.8)	0 (0.0)	1 (0.6)	1 (3.2)
Africa	33 (2.8)	4 (2.5)	15 (8.2)	7 (3.5)	2 (4)	0 (0.0)	(0.0)	3 (2.2)	1 (0.6)	1 (3.2)
Cochrane Reviews										
Europe	99 (78.6)	14 (63.6)	20 (95.2)	10 (52.6)	6 (75.0)	5 (71.4)	8 (88.9)	12 (80.0)	24 (96)	0 (0.0)
North America	2 (1.6)	0 (0.0)	0 (0.0)	2 (10.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Asia	13 (10.3)	6 (27.3)	1 (4.8)	2 (10.5)	1 (12.5)	1 (14.3)	0 (0.0)	1 (6.7)	1 (4)	0 (0.0)
South America	10 (7.9)	1 (4.5)	0 (0.0)	5 (26.3)	0 (0.0)	1 (14.3)	1 (11.1)	2 (13.3)	0 (0.0)	0 (0.0)
Australia	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Africa	2 (1.6)	1 (4.5)	0 (0.0)	0 (0.0)	1 (12.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Non-Cochrane Reviews										
Europe	546 (51.4)	73 (52.1)	68 (41.7)	92 (51.4)	16 (38.1)	12 (25.5)	118 (58.1)	69 (56.1)	82 (61.2)	16 (51.6)
North America	301 (28.3)	46 (32.9)	60 (36.8)	44 (24.6)	4 (9.5)	16 (34)	51 (25.1)	37 (30.1)	33 (24.6)	10 (32.3)
Asia	86 (8.1)	13 (9.3)	6 (3.7)	13 (7.3)	4 (9.5)	15 (31.9)	12 (5.9)	9 (7.3)	13 (9.7)	1 (3.2)
South America	51 (4.8)	3 (2.1)	6 (3.75)	6 (3.4)	8 (19)	1 (2.1)	16 (7.9)	5 (4.1)	4 (3)	2 (6.5)
Australia	47 (4.4)	2 (1.4)	8 (4.9)	17 (9.5)	9 (21.4)	3 (6.4)	6 (3)	0 (0.0)	1 (0.7)	1 (3.2)
Africa	31 (2.9)	3 (2.1)	15 (9.2)	7 (3.9)	1 (2.4)	0 (0.0)	0 (0.0)	3 (2.4)	1 (0.7)	1 (3.2)
Country of Corresponding Author, n (% total)										
Overall (Cochrane & Non-Cochrane Reviews)										
No. of countries	47	25	22	29	17	18	29	22	23	14
USA	218 (18.4)	31 (19.1)	46 (25)	33 (16.7)	4 (8.0)	15 (27.8)	46 (21.7)	12 (8.7)	27 (17)	4 (12.9)
UK	196 (16.5)	33 (20.4)	37 (20.1)	15 (7.6)	14 (28)	8 (14.8)	24 (11.3)	27 (19.6)	36 (22.6)	2 (6.5)
Canada	85 (7.2)	15 (9.3)	14 (7.6)	13 (6.6)	0 (0.0)	1 (1.9)	5 (2.4)	25 (18.1)	6 (3.8)	6 (19.4)
The Netherlands	82 (6.9)	8 (4.9)	18 (9.8)	12 (6.1)	1 (2.0)	0 (0.0)	14 (6.6)	11 (8.0)	14 (8.8)	4 (12.9)
Switzerland	67 (5.6)	4 (2.5)	4 (2.2)	24 (12.1)	1 (2.0)	1 (1.9)	16 (7.5)	5 (3.6)	12 (7.5)	0 (0.0)
Italy	65 (5.5)	17 (10.5)	3 (1.6)	5 (2.5)	0 (0.0)	4 (7.4)	14 (6.6)	9 (6.5)	12 (7.5)	1 (3.2)

Brazil	57 (4.8)	3 (1.9)	6 (3.3)	10 (5.1)	7 (14.0)	2 (3.7)	16 (7.5)	7 (5.1)	4 (2.5)	2 (6.5)
Germany	46 (3.9)	2 (1.2)	1 (0.5)	18 (9.1)	1 (2.0)	0 (0.0)	13 (6.1)	2 (1.4)	8 (5.0)	1 (3.2)
Sweden	40 (3.4)	6 (3.7)	8 (4.3)	2 (1.0)	1 (2.0)	1 (1.9)	11 (5.2)	7 (5.1)	2 (1.3)	2 (6.5)
China	40 (3.4)	8 (4.9)	5 (2.7)	2 (1.0)	2 (4.0)	3 (5.6)	2 (0.9)	8 (5.8)	10 (6.3)	0 (0.0)
Greece	28 (2.4)	1 (0.6)	0 (0.0)	5 (2.5)	1 (2.0)	0 (0.0)	10 (4.7)	10 (7.2)	1 (0.6)	0 (0.0)
Australia	28 (2.4)	0 (0.0)	4 (2.2)	8 (4.0)	9 (18.0)	3 (5.6)	3 (1.4)	0 (0.0)	1 (0.6)	0 (0.0)
Spain	25 (2.1)	9 (5.6)	1 (0.5)	1 (0.5)	0 (0.0)	0 (0.0)	4 (1.9)	2 (1.4)	8 (5.0)	0 (0.0)
South Africa	25 (2.1)	1 (0.6)	15 (8.2)	7 (3.5)	1 (2.0)	0 (0.0)	0 (0.0)	1 (0.7)	0 (0.0)	0 (0.0)
Denmark	19 (1.6)	2 (1.2)	4 (2.2)	0 (0.0)	1 (2.0)	1 (1.9)	4 (1.9)	0 (0.0)	5 (3.1)	2 (6.5)
Belgium	16 (1.3)	0 (0.0)	1 (0.5)	4 (2.0)	0 (0.0)	0 (0.0)	5 (2.4)	2 (1.4)	2 (1.3)	2 (6.5)
New Zealand	16 (1.3)	1 (0.6)	2 (1.1)	9 (4.5)	0 (0.0)	0 (0.0)	3 (1.4)	0 (0.0)	0 (0.0)	1 (3.2)
Japan	16 (1.3)	2 (1.2)	0 (0.0)	10 (5.1)	1 (2.0)	1 (1.9)	1 (0.5)	0 (0.0)	1 (0.6)	0 (0.0)
Other	119 (10)	19 (11.7)	15 (8.2)	20 (10.1)	6 (12.0)	14 (25.9)	21 (9.9)	10 (7.2)	10 (6.3)	4 (12.9)
Cochrane Reviews										
No. of countries	20	8	5	9	3	4	3	3	5	0
UK	82 (65.1)	12 (54.5)	16 (76.2)	7 (36.8)	6 (75)	2 (28.6)	7 (77.8)	12 (80.0)	20 (80)	0 (0.0)
Brazil	9 (7.1)	0 (0.0)	0 (0.0)	5 (26.3)	0 (0.0)	1 (14.3)	1 (11.1)	2 (13.3)	0 (0.0)	0 (0.0)
Bahrain	6 (4.8)	3 (13.6)	0 (0.0)	1 (5.3)	0 (0.0)	1 (14.3)	0 (0.0)	0 (0.0)	1 (4.0)	0 (0.0)
China	5 (4.0)	2 (9.1)	1 (4.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (6.7)	1 (4.0)	0 (0.0)
Germany	4 (3.2)	1 (4.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (11.1)	0 (0.0)	2 (8.0)	0 (0.0)
Italy	4 (3.2)	1 (4.5)	0 (0.0)	0 (0.0)	0 (0.0)	3 (42.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
France	2 (1.6)	0 (0.0)	1 (4.8)	1 (5.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Finland	2 (1.6)	0 (0.0)	2 (9.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Other	12 (9.5)	3 (13.6)	1 (4.8)	5 (26.3)	2 (25)	0 (0.0)	0 (0.0)	0 (0.0)	1 (4.0)	0 (0.0)
Non-Cochrane Reviews										
No. of countries	47	22	22	26	17	18	29	22	23	14
USA	217 (20.4)	31 (22.5)	46 (28.2)	32 (17.9)	4 (9.5)	15 (31.9)	46 (22.7)	12 (9.8)	27 (20.1)	4 (12.9)
UK	114 (10.7)	21 (15)	21 (12.9)	8 (4.5)	8 (19)	6 (12.8)	17 (8.4)	15 (12.2)	16 (11.9)	2 (6.5)
Canada	84 (7.9)	15 (10.7)	14 (8.6)	12 (6.7)	0 (0.0)	1 (2.1)	5 (2.5)	25 (20.3)	6 (4.5)	6 (19.4)
The Netherlands	81 (7.6)	8 (5.7)	18 (11.0)	12 (6.7)	1 (2.4)	0 (0.0)	14 (6.9)	11 (8.9)	13 (9.7)	4 (12.9)
Switzerland	67 (6.3)	4 (2.9)	4 (2.5)	24 (13.4)	1 (2.4)	1 (2.1)	16 (7.9)	5 (4.1)	12 (9.0)	0 (0.0)
Italy	61 (5.7)	16 (11.4)	3 (1.8)	5 (2.8)	0 (0.0)	1 (2.1)	14 (6.9)	9 (7.3)	12 (9.0)	1 (3.2)
Brazil	48 (4.5)	3 (2.1)	6 (3.7)	5 (2.8)	7 (16.7)	1 (2.1)	15 (7.4)	5 (4.1)	4 (3.0)	2 (6.5)
Germany	42 (4.0)	1 (0.7)	1 (0.6)	18 (10.1)	1 (2.4)	0 (0.0)	12 (5.9)	2 (1.6)	6 (4.5)	1 (3.2)
Sweden	40 (3.8)	6 (4.3)	8 (4.9)	2 (1.1)	1 (2.4)	1 (2.1)	11 (5.4)	7 (5.7)	2 (1.5)	2 (6.5)

Greece	28 (2.6)	1 (0.7)	0 (0.0)	5 (2.8)	1 (2.4)	0 (0.0)	10 (4.9)	10 (8.1)	1 (0.7)	0 (0.0)
Australia	28 (2.6)	0 (0.0)	4 (2.5)	8 (4.5)	9 (21.4)	3 (6.4)	3 (1.5)	0 (0.0)	1 (0.7)	0 (0.0)
Other	252 (23.7)	34 (24.3)	38 (23.3)	48 (26.8)	9 (21.4)	18 (38.3)	40 (19.7)	22 (17.9)	34 (25.4)	9 (29)

Appendix 2.C. Authors and affiliation of oral health systematic reviews

	Overall (No. Overall=1188: NCRs=1062 & CRs=126)	**Oral Medicine & Oral Pathology** (No. Overall=162: NCRs=140 & CRs=22)	**Dental Public Health** (No. Overall=184: NCRs=163 & CRs=21)	**Prosthodontics & Restorative Dentistry** (No. Overall=198: NCRs=179 & CRs=19)	**Pediatric Dentistry** (No. Overall=50: NCRs=42 & CRs=8)	**Endodontics** (No. Overall=54: NCRs=47 & CRs=7)	**Periodontics** (No. Overall=212: NCRs=203 & CRs=9)	**Orthodontics &Dentofacial Orthopedics** (No. Overall=138: NCRs=123 & CRs=15)	**Oral and Maxillofacial Surgery** (No. Overall=159: NCRs=134 & CRs=25)	**Oral and Maxillo-facial Radiology** (No. Overall=31: NCRs=31 & CRs=0)
Number of Authors										
Overall (Cochrane & Non-Cochrane Reviews)										
Number of authors, median (IQR)										
	4 (2, 5)	4 (3, 6)	3 (2, 4.75)	3 (2, 5)	3 (2, 4)	4 (3, 5)	3.5 (2.25, 5)	3 (3, 5)	4 (3, 5)	3.5 (2.25, 4.75)
Number of authors, n (% total)										
1	78 (6.6)	13 (8.0)	18 (9.8)	10 (5.1)	6 (12)	0 (0.0)	13 (6.1)	6 (4.3)	7 (4.4)	5 (16.1)
2-3	505 (42.5)	55 (34)	83 (45.1)	90 (45.5)	20 (40)	23 (42.6)	93 (43.9)	66 (47.8)	64 (40.3)	11 (35.5)
4-6	520 (43.8)	72 (44.4)	65 (35.3)	88 (44.4)	19 (38)	29 (53.7)	98 (46.2)	58 (42.0)	77 (48.4)	14 (45.2)
≥ 7	85 (7.2)	22 (13.6)	18 (9.8)	10 (5.1)	5 (10)	2 (3.7)	8 (3.8)	8 (5.8)	11 (6.9)	1 (3.2)
Cochrane Reviews										
Number of authors, median (IQR)										
	5 (4, 6)	4.5 (3.75,6)	4 (4, 6.5)	4 (3, 6)	4 (3,4.75)	5 (4, 5)	5 (4, 5)	5 (3, 6)	5 (4, 6.75)	N/A
Number of authors, n (% total)										
1	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
2-3	26 (20.6)	5 (22.7)	1 (4.8)	6 (31.5)	3 (37.5)	0 (0.0)	1 (11.1)	5 (33.3)	5 (20.0)	0 (0.0)
4-6	81 (64.3)	15 (68.2)	15 (71.4)	11 (57.9)	5 (62.5)	7 (100)	8 (88.9)	7 (46.7)	13 (52.0)	0 (0.0)
≥ 7	19 (15.1)	2 (9.1)	5 (23.8)	2 (10.5)	0 (0.0)	0 (0.0)	0 (0.0)	3 (20.0)	7 (28.0)	0 (0.0)
Non-Cochrane Reviews										
Number of authors, median (IQR)										
	3 (2, 5)	4 (3, 5.75)	3 (2, 4)	3 (2, 4)	2.5 (2, 4.25)	4 (3, 5)	3 (2, 5)	3 (2, 5)	4 (2, 5)	3 (2, 4)
Number of authors, n (% total)										
1	78 (7.3)	13 (9.3)	18 (11.0)	10 (5.6)	6 (14.3)	0 (0.0)	13 (6.4)	6 (4.9)	7 (5.2)	5 (16.1)
2-3	479 (45.1)	50 (35.7)	82 (50.3)	84 (46.9)	17 (40.5)	23 (48.9)	92 (45.3)	61 (49.6)	59 (44.0)	11 (35.5)
4-6	439 (41.3)	57 (40.7)	50 (30.7)	77 (43.0)	14 (33.3)	22 (46.8)	90 (44.3)	51 (41.5)	64 (47.8)	14 (45.2)
≥ 7	66 (6.2)	20 (14.3)	13 (80.0)	8 (4.5)	5 (11.9)	2 (4.3)	8 (3.9)	5 (4.1)	4 (3.0)	1 (3.2)
Number of Schools/Affiliations										
Overall (Cochrane & Non-Cochrane Reviews)										
Number of schools, median (IQR)										
	2 (1, 3)	2 (1, 3)	2 (1, 3)	2 (1, 3)	1.5 (1, 2)	2 (1, 3)	2 (1, 3)	2 (1, 2)	2 (1, 3)	2 (1, 2)

Number of schools, n (% total)										
1	454 (38.2)	44 (27.2)	66 (35.9)	81 (40.9)	25 (50)	25 (46.3)	74 (34.9)	60 (43.5)	67 (42.1)	12 (38.7)
2-3	573 (48.2)	82 (50.6)	82 (44.6)	94 (47.5)	19 (38)	19 (35.2)	118(55.7)	65 (47.1)	77 (48.4)	17 (54.8)
≥ 4	161 (13.6)	36 (22.2)	36 (19.6)	23 (11.6)	6 (12)	10 (18.5)	20 (9.4)	13 (9.4)	15 (9.4)	2 (6.5)
Cochrane Reviews										
Number of schools, median (IQR)										
	3 (2, 4)	3 (2, 4)	4 (3, 4.5)	3 (2, 4)	3 (1.5, 4)	4 (2, 4)	2 (1.5, 3.5)	3 (2, 5)	2 (1, 4)	N/A
Number of schools, n (% total)										
1	19 (15.1)	2 (9.1)	0 (0.0)	2 (10.5)	2 (25.0)	1 (14.3)	2 (22.2)	1 (6.7)	9 (36.0)	0 (0.0)
2-3	57 (45.2)	13 (59.1)	7 (33.3)	11 (57.9)	3 (37.5)	2 (28.6)	5 (55.6)	8 (53.3)	8 (32.0)	0 (0.0)
≥ 4	50 (39.7)	7 (31.8)	14 (66.7)	6 (31.6)	3 (37.5)	4 (57.1)	2 (22.2)	6 (40.0)	8 (32.0)	0 (0.0)
Non-Cochrane Reviews										
Number of schools, median (IQR)										
	2 (1, 3)	2 (1, 3)	2 (1, 3)	2 (1, 2)	1 (1, 2)	1 (1, 3)	2 (1, 3)	2 (1, 2)	2 (1, 2)	2 (1, 2)
Number of schools, n (% total)										
1	435 (41.0)	42 (30.0)	66 (40.5)	79 (44.1)	23 (54.8)	24 (51.1)	72 (35.5)	59 (48)	58 (43.3)	12 (38.7)
2-3	516 (48.6)	69 (49.3)	75 (46.0)	83 (46.4)	16 (38.1)	17 (36.2)	113(55.7)	57 (46.3)	69 (51.5)	17 (54.8)
≥ 4	111 (10.5)	29 (20.7)	22 (13.5)	17 (19.5)	3 (7.1)	6 (12.8)	18 (8.9)	7 (5.7)	7 (5.2)	2 (6.5)

NCRs, non-Cochrane Reviews; CRs, Cochrane Reviews; IQR, interquartile range, N/A, not applicable.

Appendix 2.D. Focus and interventions of oral health systematic reviews

	Overall (No. Overall=1188: NCRs=1062 & CRs=126)	**Oral Medicine & Oral Pathology** (No. Overall=162: NCRs=140 & CRs=22)	**Dental Public Health** (No. Overall=184: NCRs=163 & CRs=21)	**Prosthodontics & Restorative Dentistry** (No. Overall=198: NCRs=179 & CRs=19)	**Pediatric Dentistry** (No. Overall=50: NCRs=42 & CRs=8)	**Endodontics** (No. Overall=54: NCRs=47 & CRs=7)	**Periodontics** (No. Overall=212: NCRs=203 & CRs=9)	**Orthodontics &Dentofacial Orthopedics** (No. Overall=138: NCRs=123 & CRs=15)	**Oral and Maxillofacial Surgery** (No. Overall=159: NCRs=134 & CRs=25)	**Oral and Maxillo-facial Radiology** (No. Overall=31: NCRs=31 & CRs=0)
Type of Review, N (% Total)										
Overall (Cochrane & Non-Cochrane Reviews)										
Therapeutic	894 (75.3)	80 (49.4)	133 (72.3)	183 (92.4)	38 (76)	48 (88.9)	151 (71.2)	113 (81.9)	146 (91.8)	2 (6.5)
Non-therapeutic	294 (24.7)	82 (50.6)	51 (27.7)	15 (7.6)	12 (24)	6 (11.1)	61 (28.8)	25 (18.1)	13 (8.2)	29 (93.5)
Cochrane Reviews										
Therapeutic	126 (100)	22 (100)	21 (100)	19 (100)	8 (100)	7 (100)	9 (100)	15 (100)	25 (100)	0 (0.0)
Non-therapeutic	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Non-Cochrane Reviews										
Therapeutic	768 (72.3)	58 (41.4)	112 (68.7)	164 (91.6)	30 (71.4)	41 (87.2)	142 (70.0)	98 (79.7)	121 (90.3)	2 (6.5)
Non-therapeutic	294 (27.7)	82 (58.6)	51 (31.3)	15 (8.4)	12 (28.6)	6 (12.8)	61 (30.0)	25 (20.3)	13 (9.7)	29 (93.5)
Focus of Non-therapeutic SRs, N (% Total)										
Overall/Non-Cochrane Reviews										
Total Number	N=294	N=82	N=51	N=15	N=12	N=6	N=61	N=25	N=13	N=29
Diagnosis/ Prognosis	112 (38.1)	24 (29.3)	16 (31.4)	3 (20.0)	4 (33.3)	1 (16.7)	22 (36.1)	9 (36.0)	8 (61.5)	25 (86.2)
Epidemiology	150 (51)	56 (68.3)	26 (51.0)	10 (66.7)	6 (50.0)	3 (50.0)	36 (59.0)	8 (32.0)	1 (7.7)	4 (13.8)
Psychological/ Educational/Policy/Quality of studies	32 (10.9)	2 (2.4)	9 (17.6)	2 (13.3)	2 (16.7)	2 (33.3)	3 (4.9)	8 (32.0)	4 (30.8)	0 (0.0)
Type of Intervention in Therapeutic SRs, Category I, N (% Total)										
Overall (Cochrane & Non-Cochrane Reviews)										
Total Number	N=894	N=80	N=133	N=183	N=38	N=48	N=151	N=113	N=146	N=2
Drug	219 (24.5)	45 (56.2)	74 (55.6)	10 (5.5)	10 (26.3)	13 (27.1)	28 (18.5)	6 (5.3)	33 (22.6)	0 (0.0)
Non-drug	577 (64.5)	19 (23.8)	51 (38.3)	169 (92.3)	19 (50.0)	18 (37.5)	90 (59.6)	105 (92.9)	104 (71.2)	2 (100)
Both	98 (11.0)	16 (20.0)	8 (6.0)	4 (2.2)	9 (23.7)	17 (35.4)	33 (21.9)	2 (1.8)	9 (6.2)	0 (0.0)
Type of Intervention in Therapeutic SRs, Category II, N (% Total)										
Overall (Cochrane & Non-Cochrane Reviews)										
Total Number	N=894	N=80	N=133	N=183	N=38	N=48	N=151	N=113	N=146	N=2

Surgical	151 (16.9)	0 (0.0)	0 (0.0)	8 (4.4)	0 (0.0)	3 (6.2)	41 (27.2)	8 (7.1)	91 (62.3)	0 (0.0)
Non-surgical	651 (72.8)	67 (83.8)	132 (99.2)	160 (87.4)	36 (94.7)	39 (81.2)	82 (54.3)	94 (83.2)	39 (26.7)	2 (100)
Both	92 (10.3)	13 (16.2)	1 (0.8)	15 (8.2)	2 (5.3)	6 (12.5)	28 (18.5)	11 (9.7)	16 (11.0)	0 (0.0)
Type of Intervention in Therapeutic SRs, Category III, N (% Total)										
Overall (Cochrane & Non-Cochrane Reviews)										
Total Number	N=894	N=80	N=133	N=183	N=38	N=48	N=151	N=113	N=146	N=2
Surgical	145 (16.2)	1 (1.2)	0 (0.0)	9 (4.9)	0 (0.0)	3 (6.2)	39 (25.8)	7 (6.2)	86 (58.9)	0 (0.0)
Device	163 (18.2)	0 (0.0)	0 (0.0)	83 (45.4)	3 (7.9)	1 (2.1)	3 (2.0)	70 (61.9)	3 (2.1)	0 (0.0)
Drug	194 (21.7)	46 (57.5)	65 (48.9)	6 (3.3)	7 (18.4)	5(10.4)	28 (18.5)	4 (3.5)	33 (22.6)	0 (0.0)
Dental Material	96 (10.7)	1 (1.2)	25 (18.8)	47 (25.7)	7 (18.4)	8 (16.7)	0 (0.0)	7 (6.2)	1 (0.7)	0 (0.0)
Psychological/ Educational/Policy	31 (3.5)	3 (3.8)	17 (12.8)	1 (0.5)	4 (10.5)	1 (2.1)	2 (1.3)	2 (1.8)	1 (0.7)	0 (0.0)
Other	105 (11.7)	11 (13.8)	18 (13.5)	16 (8.8)	7 (18.4)	14 (29.2)	30 (19.9)	5 (4.4)	2 (1.4)	2 (100)
Multiple/ Combined	160 (17.9)	18 (22.5)	8 (6.0)	21 (11.5)	10 (26.3)	16 (33.3)	49 (32.5)	18 (15.9)	20 (13.7)	0 (0.0)
NCRs, non-Cochrane Reviews; CRs, Cochrane Reviews; N/A, not applicable.										

Appendix 2.E. Study designs of studies included in oral health systematic reviews

	Overall (No. Overall=1188: NCRs=1062 & CRs=126)	**Oral Medicine & Oral Pathology** (No. Overall=162: NCRs=140 & CRs=22)	**Dental Public Health** (No. Overall=184: NCRs=163 & CRs=21)	**Prosthodontics & Restorative Dentistry** (No. Overall=198: NCRs=179 & CRs=19)	**Pediatric Dentistry** (No. Overall=50: NCRs=42 & CRs=8)	**Endodontics** (No. Overall=54: NCRs=47 & CRs=7)	**Periodontics** (No. Overall=212: NCRs=203 & CRs=9)	**Orthodontics &Dentofacial Orthopedics** (No. Overall=138: NCRs=123 & CRs=15)	**Oral and Maxillofacial Surgery** (No. Overall=159: NCRs=134 & CRs=25)	**Oral and Maxillo-facial Radiology** (No. Overall=31: NCRs=31 & CRs=0)
Study Designs of SRs with Eligible Studies										
Overall (Cochrane & Non-Cochrane Reviews)										
Total Number	N=1163	N=161	N=183	N=193	N=48	N=50	N=212	N=134	N=151	N=31
RCTs only	283 (24.3)	39 (24.2)	61 (33.3)	33 (17.1)	10 (20.8)	16 (32.0)	62 (29.2)	21 (15.7)	40 (26.5)	1 (3.2)
CCTs only	10 (0.9)	0 (0.0)	0 (0.0)	3 (1.6)	0 (0.0)	0 (0.0)	1 (0.5)	2 (1.5)	4 (2.6)	0 (0.0)
RCTs and CCTs	71 (6.1)	5 (3.1)	18 (9.8)	11 (5.7)	4 (8.3)	0 (0.0)	9 (4.2)	15 (11.2)	9 (6.0)	0 (0.0)
RCTs and other designs	326 (28.0)	31 (19.3)	46 (25.1)	76 (39.4)	13 (27.1)	20 (40.0)	59 (27.8)	30 (22.4)	46 (30.5)	5 (16.1)
Non-RCTs	424 (36.5)	84 (52.2)	52 (28.4)	54 (28.0)	21 (43.8)	12 (24.0)	72 (34)	61 (45.5)	43 (28.5)	25 (80.6)
Unclear/Not reported	49 (4.2)	2 (1.2)	6 (3.3)	16 (8.3)	0 (0.0)	2 (4.0)	9 (4.2)	5 (3.7)	9 (6.0)	0 (0.0)
Cochrane Reviews										
Total Number	N=104	N=21	N=20	N=14	N=6	N=5	N=9	N=11	N=18	N=0
RCTs only	97 (93.3)	19 (90.5)	19 (95)	13 (92.9)	6 (100)	5 (100)	9 (100)	8 (72.7)	18 (100)	0 (0.0)
CCTs only	1 (1.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (9.1)	0 (0.0)	0 (0.0)
RCTs and CCTs	4 (3.8)	1 (4.8)	0 (0.0)	1 (7.1)	0 (0.0)	0 (0.0)	0 (0.0)	2 (18.2)	0 (0.0)	0 (0.0)
RCTs and other designs	1 (1.0)	0 (0.0)	1 (5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Non-RCTs	1 (1.0)	1 (4.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Unclear/Not reported	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Non-Cochrane Reviews										
Total Number	N=1059	N=140	N=163	N=179	N=42	N=45	N=203	N=123	N=133	N=31
RCTs only	186 (17.6)	20 (14.3)	42 (25.8)	20 (11.2)	4 (9.5)	11 (24.4)	53 (26.1)	13 (10.6)	22 (16.5)	1 (3.2)
CCTs only	9 (0.8)	0 (0.0)	0 (0.0)	3 (1.7)	0 (0.0)	0 (0.0)	1 (0.5)	1 (0.8)	4 (3.0)	0 (0.0)
RCTs and CCTs	67 (6.3)	4 (2.9)	18 (11.0)	10 (5.6)	4 (9.5)	0 (0.0)	9 (4.4)	13 (10.6)	9 (6.8)	0 (0.0)
RCTs and other designs	325 (30.7)	31 (22.1)	45 (27.6)	76 (42.5)	13 (31.0)	20 (44.4)	59 (29.1)	30 (24.4)	46 (34.6)	5 (16.1)
Non-RCTs	423 (39.9)	83 (59.3)	52 (31.9)	54 (30.2)	21 (50.0)	12 (26.7)	72 (35.5)	61 (49.6)	43 (32.3)	25 (80.6)
Unclear/Not reported	49 (4.6)	2 (1.4)	6 (3.7)	16 (8.9)	0 (0.0)	2 (4.4)	9 (4.4)	5 (4.1)	9 (6.8)	0 (0.0)
NCRs, non-Cochrane reviews; CRs, Cochrane reviews; RCTs, randomized controlled trials; CCTs, controlled clinical trials.										

Appendix 2.F. Number of included studies of oral health systematic reviews

	Overall (No. Overall=1188 : NCRs=1062 & CRs=126)	**Oral Medicine & Oral Pathology** (No. Overall=162: NCRs=140 & CRs=22)	**Dental Public Health** (No. Overall=184: NCRs=163 & CRs=21)	**Prosthodontics & Restorative Dentistry** (No. Overall=198: NCRs=179 & CRs=19)	**Pediatric Dentistry** (No. Overall=50: NCRs=42 & CRs=8)	**Endodontics** (No. Overall=54: NCRs=47 & CRs=7)	**Periodontics** (No. Overall=212 : NCRs=203 & CRs=9)	**Orthodontics &Dentofacial Orthopedics** (No. Overall=138: NCRs=123 & CRs=15)	**Oral and Maxillofacial Surgery** (No. Overall=159: NCRs=134 & CRs=25)	**Oral and Maxillo-facial Radiology** (No. Overall=31: NCRs=31 & CRs=0)
Number of Included Studies										
Overall (Cochrane & Non-Cochrane Reviews)										
Number of included studies, median (IQR)										
	14 (7, 28)	16.5 (9, 31)	16 (8, 33)	13 (8, 27)	12 (5.5, 25.5)	11 (4, 30)	15 (9, 28)	11 (5.75, 18.25)	13 (7, 25)	15.5 (7,28)
Number of included studies, n (total)										
0	25 (2.1)	1 (0.6)	1 (0.5)	5 (2.5)	2 (4.0)	4 (7.4)	0 (0.0)	4 (2.9)	8 (5.0)	0 (0.0)
1-5	166 (14.0)	21 (13.0)	16 (8.7)	25 (12.6)	10 (20.0)	12 (22.2)	25 (11.8)	29 (21.0)	24 (15.1)	4 (12.9)
6-15	433 (36.4)	49 (30.2)	70 (38.0)	69 (34.8)	20 (40.0)	15 (27.8)	81 (38.2)	61 (44.2)	57 (35.8)	11 (35.5)
16-30	261 (22.0)	42 (25.9)	42 (22.8)	44 (22.2)	8 (16.0)	9 (16.7)	52 (24.5)	22 (15.9)	34 (21.4)	8 (25.8)
>30	251 (21.1)	39 (24.1)	50 (27.2)	40 (20.2)	9 (18.0)	13 (24.1)	45 (21.2)	18 (13.0)	30 (18.9)	7 (22.6)
Unclear/Not reported	52 (4.4)	10 (6.2)	5 (2.7)	15 (7.6)	1 (2.0)	1 (1.9)	9 (4.2)	4 (2.9)	6 (3.8)	1 (3.2)
Number of included RCTs, median (IQR)										
	1 (0, 7)	0 (0, 6)	4 (0, 11.5)	1 (0, 6)	0 (0, 5)	3 (0, 6)	3 (0, 10)	0 (0, 4)	1 (0, 7)	0 (0, 0)
Number of included RCTs, n (% total)										
0	461 (38.3)	85 (52.5)	53 (28.8)	74 (34.9)	23 (46.0)	16 (29.6)	23 (46.0)	68 (49.3)	55 (34.6)	25 (80.6)
1-2	116 (9.8)	14 (8.6)	14 (7.6)	9 (4.2)	3 (6.0)	7 (13.0)	3 (6.0)	17 (12.3)	25 (15.7)	2 (6.5)
3-4	72 (6.1)	8 (4.9)	10 (5.4)	19 (9.0)	5 (10.0)	9 (16.7)	5 (10.0)	6 (4.3)	8 (5.0)	0 (0.0)
5-10	183 (15.4)	16 (9.9)	34 (18.5)	39 (18.4)	7 (14.0)	13 (24.1)	7 (14.0)	17 (12.3)	22 (13.8)	0 (0.0)
11-20	96 (8.1)	15 (9.3)	26 (14.1)	13 (6.1)	4 (8.0)	3 (5.6)	4 (8.0)	10 (7.2)	14 (8.8)	1 (3.2)
>20	75 (75)	12 (7.4)	16 (8.7)	27 (12.7)	3 (6.0)	0 (0.0)	3 (6.0)	0 (0.0)	11 (6.9)	0 (0.0)
Unclear/Not reported	185 (15.6)	12 (7.4)	31 (16.8)	31 (14.6)	5 (10.0)	6 (11.1)	5 (10.0)	20 (14.5)	24 (15.1)	3 (9.7)
Cochrane Reviews										
Number of included studies, median (IQR)										
	5 (1, 13)	9 (2, 19.75)	12 (3, 30.5)	2 (0, 6)	3 (0.75, 10.25)	3 (0, 4)	7 (3.5, 15)	3 (0, 8)	2 (0, 12)	N/A
Number of included studies, n (% total)										
0	22 (17.5)	1 (4.5)	1 (4.8)	5 (26.3)	2 (25.0)	2 (28.6)	0 (0.0)	4 (26.7)	7 (29.2)	0 (0.0)
1-5	45 (35.7)	8 (36.4)	5 (23.8)	8 (42.1)	4 (50.0)	4 (57.1)	3 (33.3)	6 (40.0)	7 (29.2)	0 (0.0)
6-15	32 (25.4)	6 (27.3)	5 (23.8)	4 (21.1)	1 (12.5)	1 (14.3)	4 (44.4)	5 (33.3)	6 (24.0)	0 (0.0)
16-30	17 (13.5)	4 (18.2)	5 (23.8)	2 (10.5)	0 (0.0)	0 (0.0)	2 (22.2)	0 (0.0)	4 (16.0)	0 (0.0)

>30	10 (7.9)	3 (13.6)	5 (23.8)	0 (0.0)	1 (12.5)	0 (0.0)	0 (0.0)	0 (0.0)	1 (4.0)	0 (0.0)
Unclear/Not reported	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Number of included RCTs, median (IQR)										
	5 (1, 12)	8.5 (2, 19.75)	12 (2, 30.5)	2 (0, 6)	3 (0.75, 10.25)	3 (0, 4)	7 (3.5, 15)	2 (0, 5)	2 (0, 12)	N/A
Number of included RCTs, n (% total)										
0	24 (19)	2 (9.1)	1 (4.8)	5 (26.3)	2 (25.0)	2 (28.6)	0 (0.0)	5 (33.3)	7 (28.0)	0 (0.0)
1-2	27 (21.4)	5 (22.7)	5 (23.8)	6 (31.6)	0 (0.0)	1 (14.3)	1 (11.1)	3 (20.0)	6 (24.0)	0 (0.0)
3-4	11 (8.7)	1 (4.5)	1 (4.8)	0 (0.0)	3 (37.5)	3 (42.9)	2 (22.2)	1 (6.7)	0 (0.0)	0 (0.0)
5-10	27 (21.4)	6 (27.3)	2 (9.5)	5 (26.3)	1 (12.5)	0 (0.0)	3 (33.3)	5 (33.3)	5 (20.0)	0 (0.0)
11-20	18 (14.3)	3 (13.6)	6 (28.6)	1 (5.3)	1 (12.5)	1 (14.3)	2 (22.2)	1 (6.7)	3 (12.0)	0 (0.0)
>20	19 (15.1)	5 (22.7)	6 (28.6)	2 (10.5)	1 (12.5)	0 (0.0)	1 (11.1)	0 (0.0)	4 (16.0)	0 (0.0)
Unclear/Not reported	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Non-Cochrane Reviews										
Number of included studies, median (IQR)										
	15 (8, 29)	17.5 (10, 34.5)	16.5 (8, 33.25)	15.5 (9, 29.75)	14 (6.5, 27)	14 (6.75, 34.75)	15 (9, 29)	12 (7, 21)	15 (8, 29)	15.5 (7, 28)
Number of included studies, n (% total)										
0	3 (0.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (4.3)	0 (0.0)	0 (0.0)	1 (0.7)	0 (0.0)
1-5	121 (11.4)	13 (9.3)	11 (6.7)	17 (9.5)	6 (14.3)	8 (17.0)	22 (10.8)	23 (18.7)	17 (12.7)	4 (12.9)
6-15	401 (37.8)	43 (30.7)	65 (39.9)	65 (36.3)	19 (45.2)	14 (29.8)	77 (37.9)	56 (45.5)	51 (38.1)	11 (35.5)
16-30	244 (23.0)	38 (27.1)	37 (22.7)	42 (23.5)	8 (19.0)	9 (19.1)	50 (24.6)	22 (17.9)	30 (22.4)	8 (25.8)
>30	241 (22.7)	36 (25.7)	45 (27.6)	40 (22.3)	8 (19.0)	13 (27.7)	45 (22.2)	18 (14.6)	29 (21.6)	7 (22.6)
Unclear/Not reported	52 (4.9)	10 (7.1)	5 (3.1)	15 (8.4)	1 (2.4)	1 (2.1)	9 (4.4)	4 (3.3)	6 (4.5)	1 (3.2)
Number of included RCTs, median (IQR)										
	1 (0, 6)	0 (0, 3.75)	4 (0, 9)	1 (0, 6)	0 (0, 5)	3 (0, 6)	3 (0, 9.75)	0 (0, 3)	1 (0, 6)	0 (0, 0)
Number of included RCTs, n (% total)										
0	437 (41.1)	83 (59.3)	52 (31.9)	57 (31.8)	21 (50.0)	14 (29.8)	74 (36.5)	63 (51.2)	48 (35.8)	25 (80.6)
1-2	89 (8.4)	9 (6.4)	9 (5.5)	19 (10.6)	3 (7.1)	6 (12.8)	8 (3.9)	14 (11.4)	19 (14.2)	2 (6.5)
3-4	61 (5.7)	7 (5.0)	9 (5.5)	7 (3.9)	2 (4.8)	6 (12.8)	17 (8.4)	5 (4.1)	8 (6.0)	0 (0.0)
5-10	156 (14.7)	10 (7.1)	32 (19.6)	30 (16.8)	6 (14.3)	13 (27.7)	36 (17.7)	12 (9.8)	17 (12.7)	0 (0.0)
11-20	78 (7.3)	12 (8.6)	20 (12.3)	9 (5.0)	3 (7.1)	2 (4.3)	11 (5.4)	9 (7.3)	11 (8.2)	1 (3.2)
>20	56 (5.3)	7 (5.0)	10 (6.1)	4 (2.2)	2 (4.8)	0 (0.0)	26 (12.8)	0 (0.0)	7 (5.2)	0 (0.0)
Unclear/Not reported	185 (17.4)	12 (8.6)	31 (19.0)	53 (29.6)	5 (11.9)	6 (12.8)	31 (15.3)	20 (16.3)	24 (17.9)	3 (9.7)

NCRs, non-Cochrane Reviews; CRs, Cochrane Reviews; IQR, interquartile range, N/A, not applicable.

Appendix 2.G. Number of studies contributed data to the largest meta-analysis in oral health reviews

	Overall (No. Overall=1188: NCRs=1062 & CRs=126)	**Oral Medicine & Oral Pathology** (No. Overall=162: NCRs=140 & CRs=22)	**Dental Public Health** (No. Overall=184: NCRs=163 & CRs=21)	**Prosthodontics & Restorative Dentistry** (No. Overall=198: NCRs=179 & CRs=19)	**Pediatric Dentistry** (No. Overall=50: NCRs=42 & CRs=8)	**Endodontics** (No. Overall=54: NCRs=47 & CRs=7)	**Periodontics** (No. Overall=212: NCRs=203 & CRs=9)	**Orthodontics &Dentofacial Orthopedics** (No. Overall=138: NCRs=123 & CRs=15)	**Oral and Maxillofacial Surgery** (No. Overall=159: NCRs=134 & CRs=25)	**Oral and Maxillo-facial Radiology** (No. Overall=31: NCRs=31 & CRs=0)
Meta-Analysis Conducted, N (% Total)										
Overall (Cochrane & Non-Cochrane Reviews)										
Yes	518 (43.6)	71 (43.8)	82 (44.6)	82 (41.4)	16 (32.0)	31 (57.4)	120 (56.6)	43 (31.2)	62 (39.0)	11(35.5)
No	670 (56.4)	91 (56.2)	102 (55.4)	116(58.6)	34 (68.0)	23 (42.6)	92 (43.4)	95 (68.8)	97 (61.0)	20(64.5)
Cochrane Reviews										
Yes	64 (50.8)	14 (63.6)	16 (76.2)	6 (31.6)	3 (37.5)	2 (28.6)	7 (77.8)	4 (26.7)	12 (48.0)	0 (0.0)
No	62 (49.2)	8 (36.4)	5 (23.8)	13 (68.4)	5 (62.5)	5 (71.4)	2 (22.2)	11 (73.3)	13 (52.0)	0 (0.0)
Non-Cochrane Reviews										
Yes	454 (42.7)	57 (40.7)	66 (40.5)	76 (42.5)	13 (31.0)	29 (61.7)	113 (55.7)	39 (31.7)	50 (37.3)	11(35.5)
No	608 (57.3)	83 (59.3)	97 (59.5)	103 (57.5)	29 (69.0)	18 (38.3)	90 (44.3)	84 (68.3)	84 (62.7)	20(64.5)
Number of Studies Contributed Data to the Largest Meta-Analysis Conducted										
Overall (Cochrane & Non-Cochrane Reviews)										
Total Number	N=518	N=71	N=82	N=82	N=16	N=31	N=120	N=43	N=62	N=11
Number of studies in largest meta-analysis, median (IQR)										
	9 (5, 18)	12 (7, 20.25)	9 (6, 20)	10 (6, 17.25)	9.5 (4.25, 17.5)	8 (5, 13)	7 (5, 13)	7 (3, 13)	10 (5, 19)	15.5 (4.25,34.5)
Number of studies in largest meta-analysis, n (% total)										
2-4	100 (19.3)	11 (15.5)	12 (14.6)	15 (18.3)	4 (25.0)	7 (22.6)	27 (22.5)	11 (25.6)	11 (17.7)	2 (18.2)
5-10	200 (38.6)	21 (29.6)	34 (41.5)	28 (34.1)	5 (31.2)	15 (48.4)	57 (47.5)	16 (37.2)	21 (33.9)	3 (27.3)
11-20	108 (20.8)	21 (29.6)	15 (18.3)	24 (29.3)	4 (25.0)	3 (9.7)	15 (12.5)	9 (20.9)	17 (27.4)	0 (0.0)
>20	104 (20.1)	17 (23.9)	20 (24.4)	15 (18.3)	3 (18.8)	6 (19.4)	19 (15.8)	6 (14.0)	13 (21.0)	5 (45.5)
Unclear/Not reported	6 (1.2)	1 (1.4)	1 (1.2)	0 (0.0)	0 (0.0)	0 (0.0)	2 (1.7)	1 (2.3)	0 (0.0)	1 (9.1)
Total Number	N=518	N=71	N=82	N=82	N=16	N=31	N=120	N=43	N=62	N=11
Number of RCTs in largest meta-analysis, median (IQR)										
	2 (0, 6)	0 (0, 3)	5 (0, 9)	1 (0, 4.25)	2 (0, 5.75)	2.5 (0, 6)	3 (0, 7)	1 (0, 3.25)	3.5 (0, 8)	0 (0, 0)
Number of RCTs in largest meta-analysis, n (% total)										
0	188 (36.3)	45 (63.4)	18 (22.0)	29 (35.4)	6 (37.5)	8 (25.8)	39 (32.5)	16 (37.2)	17 (27.4)	10(90.9)
2-4	107 (20.7)	11 (15.5)	15 (18.3)	18 (22.0)	5 (31.2)	10 (32.3)	23 (19.2)	11 (25.6)	14 (22.6)	0 (0.0)
5-10	104 (20.1)	10 (14.1)	23 (28.0)	12 (14.6)	0 (0.0)	6 (19.4)	33 (27.5)	7 (16.3)	12 (19.4)	1 (9.1)
11-20	27 (5.2)	1 (1.4)	7 (8.5)	3 (3.7)	3 (18.8)	0 (0.0)	5 (4.2)	0 (0.0)	8 (12.9)	0 (0.0)
>20	21 (4.1)	1 (1.4)	8 (9.8)	0 (0.0)	0 (0.0)	2 (6.5)	9 (7.5)	0 (0.0)	1 (1.6)	0 (0.0)

Unclear/Not reported	71 (13.7)	3 (4.2)	11 (13.4)	20 (24.4)	2 (12.5)	5 (16.1)	11 (9.2)	9 (20.9)	10 (16.1)	0 (0.0)
Cochrane Reviews										
Total Number	N=64	N=14	N=16	N=6	N=3	N=2	N=7	N=4	N=12	N=0
Number of studies in largest meta-analysis, median (IQR)										
	5.5 (3, 9)	6.5 (3, 9)	9 (6.25, 31.25)	3 (2, 4.25)	2 (2, 11)	5 (2, 8)	3 (2, 9)	2.5 (2, 3)	4 (2.25, 13)	N/A
Number of studies in largest meta-analysis, n (% total)										
2-4	31 (48.4)	6 (42.9)	2 (12.5)	5 (83.3)	2 (66.7)	1 (50.0)	4 (57.1)	4 (100)	7 (58.3)	0 (0.0)
5-10	20 (31.2)	8 (57.1)	7 (43.8)	1 (16.7)	0 (0.0)	1 (50.0)	2 (28.6)	0 (0.0)	1 (8.3)	0 (0.0)
11-20	7 (10.9)	0 (0.0)	2 (12.5)	0 (0.0)	1 (33.3)	0 (0.0)	1 (14.3)	0 (0.0)	3 (25.0)	0 (0.0)
>20	6 (9.4)	0 (0.0)	5 (31.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (8.3)	0 (0.0)
Unclear/Not reported	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Total Number	N=64	N=14	N=16	N=6	N=3	N=2	N=7	N=4	N=12	N=0
Number of RCTs in largest meta-analysis, median (IQR)										
	4.5 (2, 9)	6.5 (3, 9)	8 (6, 31.25)	2.5 (2, 4.25)	2 (2, 11)	5 (2, 8)	3 (2, 9)	2.5 (2, 3)	4 (2.25, 13)	N/A
Number of RCTs in largest meta-analysis, n (% total)										
0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
2-4	32 (50.0)	6 (42.9)	3 (18.8)	5 (83.3)	2 (66.7)	1 (50.0)	4 (57.1)	4 (100)	7 (58.3)	0 (0.0)
5-10	19 (29.7)	8 (57.1)	6 (37.5)	1 (16.7)	0 (0.0)	1 (50.0)	2 (28.6)	0 (0.0)	1 (8.3)	0 (0.0)
11-20	7 (10.9)	0 (0.0)	2 (12.5)	0 (0.0)	1 (33.3)	0 (0.0)	1 (14.3)	0 (0.0)	3 (25.0)	0 (0.0)
>20	6 (9.4)	0 (0.0)	5 (31.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (8.3)	0 (0.0)
Unclear/Not reported	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Non-Cochrane Reviews										
Total Number	N=454	N=57	N=66	N=76	N=13	N=29	N=113	N=39	N=50	N=11
Number of studies in largest meta-analysis, median (IQR)										
	9 (6, 19)	15 (9, 22.75)	9 (6, 19)	11 (7, 18)	10 (5.5, 20)	8 (5, 16)	7 (5, 13)	8 (5, 13)	11.5 (6.75, 20.5)	15.5 (4.25, 34.5)
Number of studies in largest meta-analysis, n (% total)										
2-4	69 (15.2)	5 (8.8)	10 (15.2)	10 (13.2)	2 (15.4)	6 (20.7)	23 (20.4)	7 (17.9)	4 (8.0)	2 (18.2)
5-10	180 (39.6)	13 (22.8)	27 (40.9)	27 (35.5)	5 (38.5)	14 (48.3)	55 (48.7)	16 (41.0)	20 (40.0)	3 (27.3)
11-20	101 (22.2)	21 (36.8)	13 (19.7)	24 (31.6)	3 (23.1)	3 (10.3)	14 (12.4)	9 (23.1)	14 (28.0)	0 (0.0)
>20	98 (21.6)	17 (29.8)	15 (22.7)	15 (19.7)	3 (23.1)	6 (20.7)	19 (16.8)	6 (15.4)	12 (24.0)	5 (45.5)
Unclear/Not reported	6 (1.3)	1 (1.8)	1 (1.5)	0 (0.0)	0 (0.0)	0 (0.0)	2 (1.8)	1 (2.6)	0 (0.0)	1 (9.1)
Total Number	N=454	N=57	N=66	N=76	N=13	N=29	N=113	N=39	N=50	N=11

Number of RCTs in largest meta-analysis, median (IQR)										
	1 (0, 6)	0 (0, 0)	4 (0, 8)	0 (0, 4.75)	0 (0, 4)	2.5 (0, 5.75)	3 (0, 7)	0 (0, 2.25)	2.5 (0, 7)	0 (0, 0)
Number of RCTs in largest meta-analysis, n (% total)										
0	188 (41.4)	45 (78.9)	18 (27.3)	29 (38.2)	6 (46.2)	8 (27.6)	39 (34.5)	16 (41.0)	17 (34.0)	10(90.9)
2-4	75 (16.5)	5 (8.8)	12 (18.2)	13 (17.1)	3 (23.1)	9 (31.0)	19 (16.8)	7 (17.9)	7 (14.0)	0 (0.0)
5-10	85 (18.7)	2 (3.5)	17 (25.8)	11 (14.5)	0 (0.0)	5 (17.2)	31 (27.4)	7 (17.9)	11 (22.0)	1 (9.1)
11-20	20 (4.4)	1 (1.8)	5 (7.6)	3 (3.9)	2 (15.4)	0 (0.0)	4 (3.5)	0 (0.0)	5 (10.0)	0 (0.0)
>20	15 (3.3)	1 (1.8)	3 (4.5)	0 (0.0)	0 (0.0)	2 (6.9)	9 (8.0)	0 (0.0)	0 (0.0)	0 (0.0)
Unclear/Not reported	71 (15.6)	3 (5.3)	11 (16.7)	20 (26.3)	2 (15.4)	5 (17.2)	11 (9.7)	9 (23.1)	10 (20.0)	0 (0.0)
NCRs, non-Cochrane Reviews; CRs, Cochrane Reviews; IQR, interquartile range, N/A, not applicable.										

Appendix 3.A. Search strategy used in the study

Database	Search Strategy
PubMed	((systematic review* OR meta-analys*)) AND (dent* OR tooth OR teeth OR orthodon* OR oral surg* OR endodon* OR periodon* OR prosthodon* OR pedodon* OR pediatric* AND dentistry OR paediatric* AND dentistry OR dent* AND public health OR oral pathology)
EMBASE	(systematic review* or meta-analys*).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui] AND (dent* OR tooth OR teeth OR orthodon* OR oral surg* OR endodon* OR periodon* OR prosthodon* OR pedodon* OR pediatric* dentistry OR paediatric* dentistry OR dent* public health OR oral pathology).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui]
MEDLINE	(systematic review* or meta-analys*).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui] AND (dent* OR tooth OR teeth OR orthodon* OR oral surg* OR endodon* OR periodon* OR prosthodon* OR pedodon* OR pediatric* dentistry OR paediatric* dentistry OR dent* public health OR oral pathology).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui]
ISI Web of Science	Topic=(dent* OR tooth OR teeth OR orthodon* OR oral surg* OR endodon* OR periodon* OR prosthodon* OR pedodon* OR pediatric* AND dentistry OR paediatric* AND dentistry OR dent* AND public health OR oral pathology) AND Topic=(systematic review* OR meta-analys*)
Cochrane Database of Systematic Reviews	(systematic review* or meta-analys*).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui] AND (dent* OR tooth OR teeth OR orthodon* OR oral surg* OR endodon* OR periodon* OR prosthodon* OR pedodon* OR pediatric* dentistry OR paediatric* dentistry OR dent* public health OR oral pathology).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui]
HealthSTAR	(systematic review* or meta-analys*).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui] AND (dent* OR tooth OR teeth OR orthodon* OR oral surg* OR endodon* OR periodon* OR prosthodon* OR pedodon* OR pediatric* dentistry OR paediatric* dentistry OR dent* public health OR oral pathology).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, ui]

Appendix 5.A. Tools and items to assess quality of randomized trials [12-20]

Items included in the scales	Jadad [13]	MAT* [14]	Delphi [15]	Van Tulder [16]	MAT-AM£ [17]	PeDro [18, 19]	Bizzini [20]	Total items (n)	FREQ %	R†	C‡	BIAS
PATIENT SELECTION (INCLUSION AND EXCLUSION AND DESCRIPTION OF SUBJECTS)												
Inclusion criteria clearly defined/eligibility criteria specified		X	X		X	X	X	5	71.4	X		Selection Bias
Exclusion criteria defined		X					X	2	28.6	X		Selection Bias
Baseline comparability (group equivalence, homogeneity) regarding the most important prognostic indicators		X	X	X	X	X	X	6	85.7		X	Selection Bias
ASSIGNMENT, RANDOMIZATION, AND ALLOCATION CONCEALMENT												
Study described as randomized	X					X	X	3	42.9	X		Selection Bias
Method of randomization described and appropriate	X	X		X	X	X		5	71.4		X	Selection Bias
Method of randomization concealed		X	X	X	X	X		5	71.4		X	Selection Bias
BLINDING												
Study described as double blind	X							1	14.3	X		Performance Bias/Detection Bias (outcome assessment)
Method of blinding described	X							1	14.3	X		Performance Bias/Detection Bias (outcome assessment)
Blinding of investigator		X	X	X	X	X	X	6	85.7		X	Detection Bias (outcome assessment)
Observer blinding evaluated and successful		X						1	14.3		X	Detection Bias (outcome assessment)
Blinding of subjects/patients		X	X	X	X	X		5	71.4		X	Performance Bias/Detection Bias (self-reported outcome assessment)
Blinding of therapist/care provider		X	X	X	X	X		5	71.4		X	Performance Bias
Blinding of the outcome (data analyst)		X						1	14.3		X	Detection Bias

INTERVENTIONS												
Treatment protocol adequately described for the treatment group regarding type of intervention, duration of each intervention, frequency, intensity, and dosage		X			X		X	3	42.9	X		Performance Bias
Treatment protocol adequately described for the control group regarding type of intervention, duration of each intervention, frequency, intensity, and dosage		X			X			2	28.6	X		Performance Bias
Treatment protocol adequately described for the comparison group regarding type of intervention, duration of each intervention, frequency, intensity (if applicable) *		X			X			2	28.6	X		Performance Bias
Control adequate (presence of a control group)							X	1	14.3		X	Performance Bias
Placebo adequate (presence of a placebo group)							X	1	14.3		X	Performance Bias
Cointerventions avoided/or comparable		X		X	X		X	4	57.1		X	Performance Bias
Cointerventions reported for each group separately					X			1	14.3	X		Performance Bias
Testing of subject compliance to treatment protocol (report of compliance)		X						1	14.3	X	X	Performance Bias/Compliance bias
Compliance acceptable in all groups (80% of treatment received)				X	X			2	28.6		X	Performance Bias/Compliance bias
ATTRITION, FOLLOW UP, AND PROTOCOL DEVIATION												
Report of withdraws and dropouts (rate)	X	X		X	X		X	5	71.4	X		Attrition Bias
Withdrawal/dropouts rate acceptable (less than 20%)		X (< 5%)		X	X	X (15%)		4	57.1		X	Attrition Bias
Reasons for withdraws and dropouts reported	X	X					X	3	42.9	X		Attrition Bias
Adverse effects described		X			X			2	28.6	X		Reporting Bias
Short follow-up measurement performed					X			1	14.3		X	Attrition Bias

Long term follow-up measurement performed					X			1	14.3		X	Attrition Bias
OUTCOMES												
Outcome measures described							X	1	14.3	X		Reporting Bias
Validity for main outcome measures reported							X	1	14.3	X	X	Information Bias
Responsiveness for main outcome measures reported							X	1	14.3	X	X	Information Bias
Reliability for main outcome measures reported							X	1	14.3	X	X	Information Bias
STATISTICAL ANALYSIS												
Descriptive measures (point estimates and measures of variability) identified and reported for the primary outcome		X	X		X	X	X	5	71.4	X		Reporting Bias
Appropriate statistical analysis used		X				X	X	3	42.9		X	Statistical Bias
Sample size calculation performed prior to initiation of the study			X						14.3		X	Threats to precision
Adequate sample size		X					X	2	28.6		X	Threats to precision
Sample size described for each group					X			1	14.3	X		Threats to precision
Intention to treat analysis used		X	X	X	X	X	X	6	85.7		X	Selection bias/attrition bias

* Maastricht
£ Maastricht Amsterdam
† Reporting
‡ Conducting

Appendix 5.B. Guidelines for the quality assessment of trials based on the tools most commonly used in health research [12-20]

Item No.	Items/Definitions	Yes (High Quality)	No (Low Quality)	Unclear Quality
1	**Inclusion, eligibility criteria for participants** (e.g., pathology of interest, age, gender, and special characteristics)	The authors describe inclusion criteria of the study participants. They clearly show the characteristics of the study population.	The authors do not describe the inclusion criteria of the study participants.	There is insufficient information about inclusion and exclusion criteria to permit a judgment.
2	**Exclusion, eligibility criteria for participants** (e.g., pathology of interest, age, gender, and special characteristics)	The authors describe exclusion criteria of the study participants. There is clear information regarding the population under study.	The authors do not describe the exclusion criteria of the study participants.	There is insufficient information about inclusion and exclusion criteria to permit a judgment.
3	**Baseline (group equivalence of participants, homogeneity) regarding the most important prognostic indicators** (the groups are similar at the start of the trial).	The authors state that the groups were comparable or had an equal prognostic factor baseline. They analyzed this by comparing groups through a statistical test in all variables of interest. Or the authors state that groups are not comparable and they adjusted statistically (e.g., by using ANCOVA). Groups must be comparable with regard to (for example) pain, global perceived effect, participation in daily activities; at least one of the main outcomes must be described, age; sex; and pre-existing participation problems.	The authors state that groups are not equal at baseline and they did not adjust for any difference.	There is insufficient information to permit a judgment.
4	**Study is described as randomized**	The authors use the word randomized, randomization, random, or minimization as derived within the title, abstract, or text.	The word randomized or randomization or any similar word does not appear in the title, abstract, or text.	There is insufficient information to permit a judgment.
5	**Randomization process performed**	The authors use the word randomized or randomization or a similar word as derived within the title, abstract, or text, to describe the method performed in the trial, such as random number tables, computer program, etc.	The authors do not describe the method performed in the trial, such as random number tables, computer program.	There is insufficient information about the sequence generation process to permit a judgment.
6	**Method of randomization described and appropriate**	The authors use the word randomized or randomization or a similar word as derived within	The authors do not describe the method used for doing the	There is insufficient information about the

		the title, abstract, or text and described the method used for doing the randomization such as random number tables, computer program, etc. The investigators describe a random component in the sequence generation process such as: • Referring to a random number table; • Using a computer random number generator; • Minimization; • Coin tossing; • Shuffling cards or envelopes; • Throwing dice; • Drawing of lots;	randomization, such as random number tables, computer program. Other nonrandom methods were used such as: • hospital records numbers, • time of presentation, • alternate numbers, • date of birth	sequence generation process to permit a judgment.
7	**Method of randomization concealed**	Allocation was done and is appropriate Assignment is generated by an independent person not responsible for determining the eligibility of the patients: To score yes this person: must have no information about patients included in the trial; and must have no influence on the assignment sequence or decision about the eligibility of the patients. Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, is used to conceal allocation: • Central allocation (including telephone, web-based, and pharmacy-controlled randomization); • Sequentially numbered drug containers of identical appearance; provided by a different person who did the randomization allocation; • Sequentially numbered, opaque, sealed envelopes.	Participants or investigators enrolling participants could possibly foresee assignments and thus introduce selection bias.	There is insufficient information to permit a judgment.
8	**Study described as double blind**	"Double blind" is the description in the study related to "blindness." Also, it should be stated that neither the person doing the assessments nor the study participants could identify the intervention being assessed.	Not described as double blind.	There is insufficient information to permit a judgment.

9	**The method of blinding was appropriate**	The authors use the blinding method appropriately. Blinding of participants/patients is a "must" when outcomes are subjective or self-reported. When outcomes are measured by an assessor, the assessors should be blinded to group allocation. Also, score "completely done" when it is unlikely that the blinding could have been broken and the nonblinding of others is unlikely to introduce bias. No blinding, but the review authors judge that the outcome and the outcome measurement are not likely to be influenced by lack of blinding. Objectives automatized outcomes coming from databases or hospital register office.	There is no blinding or incomplete blinding is performed, and the outcome or outcome measurement is likely to be influenced by lack of blinding.	There is insufficient information to permit a judgment.
10	**Blinded investigator**	The study describes in the title, abstract, or text that the investigator was blinded. The blinding was appropriate.	The study describes in the title, abstract, or text that the investigator was not blinded.	There is insufficient information to permit a judgment.
11	**Blinded assessor**	The study describes in the title, abstract, or text that the assessor was blinded. The blinding was appropriate. When outcomes are measured by an assessor, the assessors should be blinded to group allocation.	The study describes in the title, abstract, or text that the assessor was not blinded.	There is insufficient information to permit a judgment.
12	**Blinded subjects/patients**	The study describes in the title, abstract, or text that subjects/patients were blinded. The blinding was appropriate.	The study describes in the title, abstract, or text that subjects/patients were not blinded.	There is insufficient information to permit a judgment.
13	**Blinded therapist/care-provider**	The study describes in the title, abstract, or text that the therapists/care-providers were blinded. The blinding was appropriate.	The study describes in the title, abstract, or text that the therapists/care-providers were not blinded, or because of the nature of the intervention (e.g., exercise prescription or supervision, etc.), the therapist could not be blinded.	There is insufficient information to permit a judgment.

14	**Blinded statistician**	The study describes in the title, abstract, or text that the statistician was blinded. The blinding was appropriate.	The study describes in the title, abstract, or text that the statistician was not blinded.	There is insufficient information to permit a judgment.
15	**Treatment protocol for experimental group**	The authors describe doses, frequency, intensity, of the treatment protocol for the experimental group (repetition, days per week, length of time) in enough detail to reproduce the intervention. At least three of the five points below are described for the experimental intervention; 1. type of intervention; 2. intensity of the intervention; 3. duration and site of each treatment session; 4. frequency of treatment sessions; and 5. total number of treatment sessions.	The authors do not describe the treatment protocol.	The authors do not describe enough aspects of the treatment protocol for the experimental group that would allow reproducibility of the intervention.
16	**Treatment protocol for control group**	The authors describe doses, frequency, intensity, position of treatment protocol for the comparison group (repetition, days per week, length of time) in enough detail to reproduce the intervention. At least three of the five points below are described for the control intervention; if more than two types of intervention are compared, take only two of them): 1. type of intervention; 2. intensity of the intervention; 3. duration and site of each treatment session; 4. frequency of treatment sessions; and 5. total number of treatment sessions.	The authors do not describe the treatment protocol.	The authors do not describe enough aspects of the treatment protocol for the control group that would allow reproducibility of the intervention.
17	**Treatment protocol for the control or comparison group #2** (if applicable)	The authors describe doses, frequency, intensity, position, of treatment protocol for the comparison group (repetition, days per week, length of time) in enough detail to reproduce the intervention. At least three of five are described for second control intervention; if more than two types of interventions. 1. type of intervention; 2. intensity of the intervention;	The authors do not describe the treatment protocol.	The authors do not describe enough aspects of treatment protocol for control group #2 that would allow reproducibility of the intervention.

		3. duration and site of each treatment session; 4. frequently of treatment sessions; and 5. total number of treatment sessions.		
18	**Control group**	The study employs a control group (i.e., no-treatment/waiting list/standard care).	A control group was not used.	There is insufficient information to permit a judgment.
19	**Placebo**	The authors describe the use of a placebo group. The authors used a credible sham and there is certainty that this sham was good and was not discovered by the patients.	A placebo group was not used.	There is insufficient information to permit a judgment.
20	**Cointerventions avoided or comparable.** Cointerventions are interventions other than the treatment under study.	The authors state that subjects did not receive an additional intervention, or that cointerventions were balanced between treatment and control groups. Data about cointerventions are presented and comparable between treatment and control groups.	Subjects received additional interventions besides the intervention under study. The cointerventions were not balanced between treatment and control groups.	There is insufficient information to permit a judgment. **N/A:** Treatment and control groups did not receive an intervention in addition to the intervention under study.
21	**Cointerventions reported for treatment and control groups.**	The authors describe cointerventions for treatment and control groups separately (type of intervention, frequency, dosage, etc.).	The authors do not explain the type or process of cointervention.	There is insufficient information to permit a judgment.
22	**Subject compliance to treatment protocol.** *Compliance* means that the subjects follow the treatment as planned; that is, the subjects attend at least 80% of the treatment sessions.	The authors describe that they registered the compliance of the subjects (e.g., through logs or diaries), or they say that subjects were compliant with the treatment because they attended at least 80% of the treatment sessions. Compliance monitoring is assumed for a one-time intervention.	The authors did not test subject compliance to treatment protocol.	There is insufficient information to permit a judgment.
23	**Acceptable compliance**	There is ≥ 80% compliance in treatment and control groups. The control group might have to be "compliant" as well. For example, in an exercise intervention, the control group would have to comply by doing no exercise.	There is less than 80% compliance in treatment and control groups.	There is insufficient information to permit a judgment.

24	**Report of withdrawals and dropouts**	There is clear reporting of all withdrawals and dropouts. Generally, this is done by using a flowchart. 1. Number of dropouts. 2. If there were no withdrawals, this fact should be stated in article.	Withdrawals and dropouts are not reported.	
25	**Acceptable withdrawal/dropout rate**	The withdrawal/dropout rate in the study was less than or equal to 20%. Or with multiple time points, at any point there must be at least 85% patients included in analysis	The withdrawal/dropout rate was > 20% when only one-time point was evaluated.	
26	**Reasons for dropouts**	There is clear reporting of all dropouts and the reason for each dropout is given.	Reasons for dropouts are not reported.	
27	**Adverse effects of the intervention**	Adverse effects of the intervention are reported.	Adverse effects of the intervention are not reported.	
28	**Short follow-up measurement of the intervention**	An outcome assessment of the intervention was performed at the end the of intervention period.	The outcome was measured before the treatment was finished and there was no outcome evaluation after the treatment was completed.	
29	**Long term follow-up of the intervention**	An outcome assessment was performed three or more months after the treatment was completed.	The outcome was assessed less than three months after the treatment was completed.	**N/A:** There is no assessment performed in the study.
30	**Description of outcome measures of the intervention**	The authors describe all the treatment outcomes, primary and secondary, and they explain how to score them.	The authors do not describe the outcome(s) of the treatment.	There is insufficient information to permit a judgment.
31	**Validity of the main outcome**	The authors report the validity of the measure of the main outcome of the intervention (this can be done by references).	The authors do not report the validity of the main outcome measure of the intervention.	There is insufficient information to permit a judgment.
32	**Responsiveness of the main outcome**	Authors report the responsiveness of the main outcome. This can be done by references. Responsiveness means sensitive to change, or able to detect change.	The authors do not report the responsiveness of the main outcome.	There is insufficient information to permit a judgment.
33	**Reliability of the main outcome**	The authors report the reliability of the main outcome of the intervention.	The authors do not report the reliability of the main outcome of the intervention.	There is insufficient information to permit a judgment.

34	**Descriptive measures (point estimates and measures of variability) reported for the primary outcome of the treatment.** Point estimates include means, medians, modes, and measures of variability and include standard deviation, 95% confidence interval, and quartile.	The authors describe both point estimates (e.g. mean) and variability measures (e.g. SD or CI "confidence interval") for the main outcome of the intervention.	The authors do not describe either the point estimates or the measures of variability for the main outcome of the intervention.	
35	**Appropriate statistical analysis**	The authors describe the analysis for each outcome and the alpha level chosen, and it seems that the chosen analysis was a good approach to the research question. Statistical comparisons and variability between or among groups in the trial must be provided. Authors need to provide estimates and variability data.	The statistical analysis is not appropriate.	There is insufficient information to permit a judgment.
36	**Sample size calculation performed prior to initiation of the study**	The authors describe a sample size calculation prior to start the study and calculate how many participants need to be recruited for the study to have an acceptable power.	The authors did not perform a sample size calculation prior to the start of the study.	
37	**Adequate sample size**	The sample size calculated is the same as the sample size recruited and maintained throughout the trial.	The sample size calculated is not the same as the sample size recruited, or the dropout rate is more than 20%, or the sample size was insufficient to show a significant treatment effect (acceptable power).	There is insufficient information to permit a judgment.
38	**Sample size described for each group**	The sample size is described for each group in the study.	The sample size is not described for each group in the study.	
39	**Intention to treat (ITT) analysis used (patients are analyzed in the groups to which they were randomized)** All randomized patients have to be analyzed for the most	The authors used an intention to treat (ITT) analysis principle and, according to their evaluation, they analyzed the subjects as randomized. There are no missing data so it is assumed that the ITT principle was followed if no other protocol deviations occurred.	The authors did not use the ITT principle or the authors said that they used ITT but an evaluation of the study indicates that subject analysis was not randomized.	There is insufficient information to permit a judgment.

	important outcome measures and at the most important moments of intervention effect measurement or of whether there are no withdrawals or loss to follow-up measurements of intervention.			

Appendix 5.C. The Cochrane Collaboration's tool for assessing risk of bias [4, 5]

DOMAIN	DESCRIPTION	RISK OF BIAS	CONSENSUS (CIRCLE)
Random sequence generation		Was the allocation sequence adequately generated?	Low/High/Unclear
Allocation concealment		Was allocation adequately concealed?	Low/High/Unclear
Blinding of participants and personnel	*Subjective outcomes* *Objective outcomes*	Was knowledge of the allocated intervention adequately prevented during the study?	Low/High/Unclear
Blinding of outcome assessment	*Subjective outcomes* *Objective outcomes*	Was knowledge of the allocated intervention adequately prevented during the study?	Low/High/Unclear
Incomplete outcome data	*Subjective outcomes* *Objective outcomes*	Were incomplete outcome data adequately addressed?	Low/High/Unclear
Selective outcome reporting		Are reports of the study free of suggestion of selective outcome reporting?	Low/High/Unclear
Other sources of bias		Was the study apparently free of other problems that could put it at a high risk of bias?	Low/High/Unclear
Overall risk of bias	Low/High/Unclear		Low/High/Unclear

Appendix 5.D. Guidelines for evaluating the risk of bias of trials

<table>
<tr><td colspan="2">RANDOM SEQUENCE GENERATION

SELECTION BIAS (BIASED ALLOCATION TO INTERVENTIONS) DUE TO INADEQUATE GENERATION OF A RANDOMISED SEQUENCE.</td></tr>
<tr><td>Criteria for a judgement of

‘Low risk’ of bias.</td><td>The investigators describe a random component in the sequence generation process such as:

• Referring to a random number table;
• Using a computer random number generator;
• Coin tossing;
• Shuffling cards or envelopes;
• Throwing dice;
• Drawing of lots;
• Minimization*.

*Minimization may be implemented without a random element, and this is considered to be equivalent to being random.</td></tr>
<tr><td>Criteria for the judgement of ‘High risk’ of bias.</td><td>The investigators describe a nonrandom component in the sequence generation process. Usually, the description would involve some systematic, nonrandom approach, for example:

• Sequence generated by odd or even date of birth;
• Sequence generated by some rule based on date (or day) of admission;
• Sequence generated by some rule based on hospital or clinic record number.

Other nonrandom approaches happen much less frequently than the systematic approaches mentioned above and tend to be obvious. They usually involve judgement or some method of nonrandom categorization of participants, for example:

• Allocation by judgement of the clinician;
• Allocation by preference of the participant;
• Allocation based on the results of a laboratory test or a series of tests;
• Allocation by availability of the intervention.</td></tr>
<tr><td>Criteria for the judgement of ‘Unclear risk’ of bias.</td><td>Insufficient information about the sequence generation process to permit judgement of ‘Low risk’ or ‘High risk.’</td></tr>
<tr><td colspan="2">ALLOCATION CONCEALMENT

Selection bias (biased allocation to interventions) due to inadequate concealment of allocations prior to assignment.</td></tr>
</table>

Criteria for a judgement of 'Low risk' of bias.	Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation: • Central allocation (including telephone, web-based and pharmacy-controlled randomization); • Sequentially numbered drug containers of identical appearance; • Sequentially numbered, opaque, sealed envelopes (3 characteristics need to be present)
Criteria for the judgement of 'High risk' of bias.	Participants or investigators enrolling participants could possibly foresee assignments and thus introduce selection bias, such as allocation based on: • Using an open random allocation schedule (e.g. a list of random numbers); • Assignment envelopes were used without appropriate safeguards (e.g. if envelopes were unsealed or nonopaque or not sequentially numbered); • Alternation or rotation; • Date of birth; • Case record number; • Any other explicitly unconcealed procedure.
Criteria for the judgement of 'Unclear risk' of bias.	Insufficient information to permit judgement of 'Low risk' or 'High risk.' This is usually the case if the method of concealment is not described or not described in sufficient detail to allow a definite judgement – for example, if the use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque, and sealed. (Not all of the 3 characteristics are present).
BLINDING OF PARTICIPANTS AND PERSONNEL (Main Outcome) • Blinding of participants/patients is a "must" when outcomes are subjective or self-reported. • When outcomes are measured by an assessor, then assessors should be blinded to group allocation. • When outcomes are automated [database] (there is no assessor involved) then, blinding of participants or assessors is not an issue Performance bias due to knowledge of the allocated interventions by participants and personnel during the study.	
Criteria for a judgement of 'Low risk' of bias.	Any one of the following: • No blinding or incomplete blinding, but the review authors judge that the outcome is not likely to be influenced by lack of blinding (Automated outcome or administrative) • Blinding of participants and key study personnel ensured, and it is unlikely that the blinding could have been broken

	• Objectives automatized outcomes coming from databases or hospital register office.
Criteria for the judgement of 'High risk' of bias.	Any one of the following: • No blinding or incomplete blinding, and the outcome is likely to be influenced by lack of blinding; • Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding.
Criteria for the judgement of 'Unclear risk' of bias.	Any one of the following: • Insufficient information to permit a judgement of 'Low risk' or 'High risk'; • The study did not address the issue of blinding.
BLINDING OF OUTCOME ASSESSMENT Detection bias due to knowledge of the allocated interventions by outcome assessors.	
Criteria for a judgement of 'Low risk' of bias.	Any one of the following: • No blinding of outcome assessment, but the review authors judge that the outcome measurement is not likely to be influenced by lack of blinding; • Blinding of outcome assessment ensured, and unlikely that the blinding could have been broken.
Criteria for a judgement of 'High risk' of bias.	Any one of the following: • No blinding of outcome assessment, and the outcome measurement is likely to be influenced by lack of blinding; • Blinding of outcome assessment, but likely that the blinding could have been broken and the outcome measurement is likely to be influenced by lack of blinding.
Criteria for the judgement of 'Unclear risk' of bias.	Any one of the following: • Insufficient information to permit judgement of 'Low risk' or 'High risk'; • The study did not address the issue of blinding.
INCOMPLETE OUTCOME DATA Attrition bias due to amount, nature, or handling of incomplete outcome data.	
Criteria for a judgement of	Any one of the following:

'Low risk' of bias.	• No missing outcome data (All patients were accounted for in the analysis) • Reasons for missing outcome data are unlikely to be related to the outcome (for survival data, censoring is unlikely to introduce bias); • Missing outcome data are balanced in numbers across intervention groups, with similar reasons for missing data across groups; • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate; • For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a clinically relevant impact on observed effect size; • Missing data have been imputed using appropriate methods. • If authors claimed that an intention-to-treat analysis was performed, raters should confirm that all patients entered were accounted for in the analysis (i.e., do not assume that a true intention-to-treat analysis was done). • If the numbers and reasons for withdrawal/drop-out were described and comparable across groups and the authors performed an intention to treat analysis with ≤ 20% drop outs, then score low risk of bias • If the numbers and reasons for withdrawal/drop-out were described and comparable across groups but the authors did not perform an ITT and the dropout rate was less than or equal to 10%, then score low risk of bias
Criteria for the judgement of 'High risk' of bias.	Any one of the following: • Reason for missing outcome data is likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups; • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk is enough to induce clinically relevant bias in the intervention effect estimate; • For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes is enough to induce clinically relevant bias in observed effect size; • 'As-treated' analysis is done with a substantial departure of the intervention received from the intervention assigned at randomization; • Potentially inappropriate application of simple imputation. • No Intention to treat (ITT) or ITT performed with > 20% drop outs
Criteria for the judgement of 'Unclear risk' of bias.	Any one of the following: • Insufficient reporting of attrition/exclusions to permit judgement of 'Low risk' or 'High risk' (e.g. number randomized participants is not stated, no reasons for missing data are provided); • The study did not address this outcome.

	• No Intention to treat (ITT) but >10% and ≤ 20% drop outs.
SELECTIVE REPORTING (Are outcomes reported in methods and results?) Reporting bias due to selective outcome reporting.	
Criteria for a judgement of 'Low risk' of bias.	Any of the following: • The study protocol is available and all of the study's prespecified (primary and secondary) outcomes that are of interest in the review have been reported in the manner prespecified; • The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were prespecified (convincing text of this nature may be uncommon). • Since we are not searching for protocols, outcomes reported in the methods section need to match those reported in the results section • If 0-30% of the secondary outcomes are not reported, score low risk. • The main outcome has to be included in both methods and results sections.
Criteria for the judgement of 'High risk' of bias.	Any one of the following: • Not all of the study's prespecified primary outcomes have been reported; • One or more primary outcomes is reported using measurements, analysis methods, or subsets of the data (e.g. subscales) that were not prespecified; • One or more reported primary outcomes were not prespecified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); • One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; • The study report fails to include results for a key outcome that would be expected to have been reported for such a study. • ≥ 70% of secondary outcomes were unreported (combining methods or results sections) • If the main outcome was not reported in the study, score a high risk of bias
Criteria for the judgement of 'Unclear risk' of bias.	• Insufficient information to permit judgement of 'Low risk' or 'High risk.' It is likely that the majority of studies will fall into this category. • If between 31%–69% of secondary outcomes are UNREPORTED (combining methods or results sections) score as unclear risk of bias
OTHER BIAS Bias due to problems not covered elsewhere in the table.	

Criteria for a judgement of 'Low risk' of bias.	The study appears to be free of other sources of bias.
Criteria for the judgement of 'High risk' of bias.	There is at least one important risk of bias. For example, the study: • Had a potential source of bias related to the specific study design used; or • Has been claimed to have been fraudulent; or • Had some other problem. • If the study has baseline imbalances regarding demographic factors, duration and severity of complaints, and value of main outcome measure(s) [21]. • Imbalances in co-interventions: if the co-interventions were imbalances between groups or they were not similar between groups [21]. • Compliance with treatment was not acceptable (very poor adherence with actual treatment: e.g. exercises performed) based on the reported intensity, duration, number and frequency of sessions for both the index intervention and control intervention(s). For example, physiotherapy treatment is usually administered over several sessions; therefore, it is necessary to assess how many sessions each patient attended. For single session interventions this item is irrelevant [21].
Criteria for the judgement of 'Unclear risk' of bias.	There may be a risk of bias, but there is either: • Insufficient information to assess whether an important risk of bias exists; or • Insufficient rationale or evidence that an identified problem will introduce bias.
Total Scoring for RoB tool: If any High Risk = High Risk of Bias If any Unclear and NO High Risk = Unclear Risk of Bias If all of the items are Low Risk = Low Risk of Bias	

Appendix 8.A. A common classification scheme for bias [5]

Type of bias	Description	Relevant domains in the Collaboration's 'Risk of bias' tool
Selection bias	Systematic differences between baseline characteristics of the groups that are compared.	• Sequence generation. • Allocation concealment.
Performance bias	Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest.	• Blinding of participants and personnel. • Other potential threats to validity.
Detection bias	Systematic differences between groups in how outcomes are determined.	• Blinding of outcome assessment. • Other potential threats to validity.
Attrition bias	Systematic differences between groups in withdrawals from a study.	• Incomplete outcome data.
Reporting bias	Systematic differences between reported and unreported findings.	• Selective outcome reporting.

References for Appendices

[1] Smail-Faugeron V, Fron-Chabouis H, Durieux P. Clinical trial registration in oral health journals. J Dent Res 2015; 94:8S-13S.

[2] Hardt JL, Metzendorf MI, Meerpohl JJ. Surgical trials and trial registers: a cross-sectional study of randomized controlled trials published in journals requiring trial registration in the author instructions. Trials 2013; 14:407.

[3] Wager E, Williams P. "Hardly worth the effort"? Medical journals' policies and their editors' and publishers' views on trial registration and publication bias: quantitative and qualitative study. BMJ 2013; 347:f5248.

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

[5] Higgins J, Altman D. Chapter 8: Assessing risk of bias in included studies. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions version 50. Chichester, UK: John Wiley & Sons, Ltd.; 2008.

[6] Turpin DL. Challenges faced by journal editors. Am J Orthod Dentofacial Orthop 2015; 147:285-9.

[7] Pandis N, Shamseer L, Kokich V, Moher D. Implementation of adherence to the CONSORT guidelines by the American Journal of Orthodontics and Dentofacial Orthopedics (AJO-DO). Peer Review Congress, Chicago (2013).

[8] Pandis N, Turpin DL. Enhancing CONSORT compliance for improved reporting of randomized controlled trials. Am J Orthod Dentofacial Orthop 2014; 145:1.

[9] Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias – an updated review. PloS One 2013; 8:e66844.

[10] Hooft L, Korevaar DA, Molenaar N, Bossuyt PM, Scholten RJ. Endorsement of ICMJE's Clinical Trial Registration Policy: a survey among journal editors. Neth J Med 2014; 72:349-55.

[11] Giannobile WV. Improving clinical trials in dentistry. J Dent Res 2015; 94:6S-7S.

[12] Armijo Olivo S, Fuentes CJ, Ospina M, Saltaji H, Hartling L. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: A descriptive analysis. BMC Med Res Methodol 2013; 13:1-19.

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

[14] De Vet HCW, De Bie RA, Van Der Heijden GJMG, Verhagen AP, Sijpkes P, Knipschild PG. Systematic reviews on the basis of methodological criteria. Physiotherapy 1997; 83:284-9.

[15] Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 1998; 51:1235-41.

[16] Van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. Spine 2003; 28:1290-9.

[17] Van Tulder MW, Assendelft WJJ, Koes BW, Bouter LM, Bombardier C, Nachemson AL, et al. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for spinal disorders. Spine 1997; 22:2323-30.

[18] Moseley AM, Herbert RD, Sherrington C, Maher CG. Evidence for physiotherapy practice: a survey of the Physiotherapy Evidence Database (PEDro). Aust J Physiother 2002; 48:43-9.

[19] Sherrington C, Herbert RD, Maher CG, Moseley AM. PEDro. A database of randomized trials and systematic reviews in physiotherapy. Man Ther 2000; 5:223-6.

[20] Bizzini M, Childs JD, Piva SR. Systematic review of the quality of randomized controlled trials for patellofemoral pain syndrome. J Orthop Sports Phys Ther 2003 Jan; 33(1):4-20.

[21] Furlan AD, Pennick V, Bombardier C, Van Tulder M. 2009 Updated method guidelines for systematic reviews in the cochrane back review group. Spine 2009; 34:1929-41.