Glossary for systematic reviews and meta-analyses

V Nagendrababu[1], P Dilokthornsakul[2], P Jinatongthai[3], SK Veettil[4], SJ Pulikkotil[1], HF Duncan[5], PMH Dummer[6]

[1]Division of Clinical Dentistry, School of Dentistry, International Medical University, Kuala Lumpur, Malaysia, [2]Center of Pharmaceutical Outcomes Research, Department of Pharmacy Practice, Faculty of Pharmaceutical Sciences, Naresuan University, Muang, Phitsanulok, Thailand, [3]Pharmacy Practice Division, Faculty of Pharmaceutical Sciences, Ubon Ratchathani University, Thailand, [4]Department of Pharmacy Practice, School of Pharmacy, International Medical University, Kuala Lumpur, Malaysia, [5]Division of Restorative Dentistry, Dublin Dental University Hospital, Trinity College Dublin, Dublin, Ireland, [6]School of Dentistry, College of Biomedical and Life Sciences, Cardiff University, Cardiff, UK.

Corresponding author

V Nagendrababu BDS, MFDS RCPS (Glasgow), MDS, PhD.

Division of Clinical Dentistry,

School of Dentistry, International Medical University

Bukit Jalil – 57000, Kuala Lumpur, Malaysia.

E mail: venkateshbabu@imu.edu.my

Keywords: Glossary, systematic reviews, meta-analyses

Running title: Glossary for systematic reviews and meta-analyses

The authors deny any conflicts of interest related to this study.

Glossary for systematic reviews and meta-analyses

Abstract

A systematic review aims to answer a focussed research question through a structured review of the evidence, using a predefined methodology, which often includes a meta-analysis. A meta-analysis is a statistical method used to combine the effect estimates from the individual studies included in a systematic review. Systematic reviews and meta-analyses are positioned at the highest level in the hierarchy of clinical evidence. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement was introduced in 2009 to help authors improve the quality and reliability of systematic reviews and meta-analyses. Recently, the volume of systematic reviews and meta-analyses in the field of Endodontology has increased; however, the quality of the published manuscripts has been reported to be sub-optimal, which does not take account of the systematic reviews that were rejected because of more obvious deficiencies. The aim of this paper is to present a comprehensive glossary of terminology commonly used in systematic reviews and meta-analyses in an attempt to provide easily understood definitions and explanations to assist authors when reporting systematic reviews and meta-analyses, and to allow those wishing to read them to become better informed.

Keywords

Glossary, systematic reviews, meta-analyses

Introduction

A systematic review is a well-planned and meticulously executed literature review that analyses the findings from existing studies to answer a focused research or review question (Uman 2011). A meta-analysis, often accompanying a systematic review, is a statistical procedure that combines or pools the results of several studies and provides a more accurate and precise estimate of the effect of a treatment, intervention or drug, the validity of a hypothesis or a risk factor for disease, compared to what an individual study taken in isolation is able to achieve (Haidich 2010). Thus, the main aim of a systematic review and meta-analysis is to summarise accurately, transparently and reliably the available evidence on the efficacy and safety of interventions, techniques or drugs that are used in healthcare (Liberati *et al.* 2009). Ultimately, they aim to provide the most reliable and unbiased resource for developing clinical practice guidelines and recommendations for future research and clinical practice, whilst also updating individual clinicians and other stakeholders on evidence-based care relating to their area of interest (Clarke 2011). In evidence-based clinical medicine/dentistry, it is universally accepted that systematic reviews and meta-analyses are on the top of the hierarchy of evidence (Burns *et al.* 2011).

The number of systematic reviews submitted each year to one journal in the field of Endodontology has increased dramatically in recent years, from approximately 19 in 2016 to 54 in 2018 (Dummer, unpublished data). The quality of systematic reviews and meta-analyses in Endodontology, published between 2001 and 2009, was reported to be "high"

(Suebnukarn *et al*. 2009); however, between 2009 and 2016, it was only categorized as "medium" (Kattan *et al*. 2018). Indeed, Kattan *et al.* (2018) concluded that although the number of systematic reviews and meta-analyses published in Endodontology had increased significantly, their quality was judged to have declined to a sub-optimal level. Shortcomings in the methodological quality of systematic reviews are of concern, as they are likely to have a negative influence on clinical decision-making and have an adverse impact on the progress of evidence-based practice in Endodontology (Kattan *et al*. 2018).

In view of this trend, and for the benefit of authors, readers and other stakeholders, the general methodological and reporting quality of systematic reviews and meta-analyses in Endodontology needs to improve (Kattan *et al*. 2018, Nagendrababu *et al*. 2018). As part of the process, the meaning of key terms used when reporting clinical trials, systematic reviews and meta-analyses needs to be clarified. A glossary is a list of terms related to a specific subject, field, or area of usage, with accompanying definitions. Defining and explaining difficult, discipline-specific or unusual words and expressions used in the text of scientific and clinical papers helps reader understanding and has the potential to result in more standardized and established terminology for dissemination and implementation within a specific field (Rabin *et al*. 2008).

There is a wide variety and increasing range of terms used in systematic reviews and meta-analyses, which can make it difficult for authors and readers to understand their meaning and subsequent use in practice. The aim of this glossary is to provide a

comprehensive list and brief explanation of the common terms used in systematic reviews and meta-analyses. It is designed to highlight the importance and significance of the terms and phrases in an attempt to help readers and also allow researchers and authors to use the terminology effectively while producing high-quality manuscripts. For each term/phrase, the most relevant references have been cited so that readers have the opportunity to engage and learn more if they desire.

Commonly used terms in systematic reviews and meta-analyses (in alphabetical order)

1. *Allocation concealment* – A process to hide information from research participants and investigators regarding the sequence of allocation of participants to the intervention group(s) in a randomised controlled trial, until the moment of assignment. This is done primarily to prevent selection bias by participants and/or investigators who may otherwise allocate participants to a preferred/favoured treatment arm (Forder *et al.* 2005).
2. *AMSTAR* - A MeaSurement Tool to Assess systematic Reviews. AMSTAR is a reliable and valid instrument, which contains 11 items in a checklist that is used to conduct and critically appraise systematic reviews (Shea *et al.* 2007, 2009). The aim of AMSTAR is to improve the methodological quality of the reviews for the benefit of authors, referees, editors, and readers.

3. *Begg's test* - A test for small-study effects (Item 60), that examines the association of standardized treatment effects and their variance using an adjusted rank correlation. Similar to Egger's test (Item 14), when $p<0.05$ it indicates that a small study effect is detected (Begg & Mazumdar 1994, Sterne *et al.* 2008), hence indicating potential bias caused by the influence of small underpowered studies.
4. *Bias* - A term used to indicate the deviation of results or inferences from the actual truth (Last 1995). Bias can occur at any stage when conducting a study, for example, during data collection, data analysis, data interpretation, publication, or review of the data, but is generally classified into 3 categories, namely information bias, selection bias, and confounding bias. Notably, some authors do not consider confounding bias as a type of bias, because is not an error, as it is related to the fact that diseases have many causes, some of which are unknown. Bias leads to the lack of validity of the results reported in a study (Delgado-Rodriguez & Llorca 2004, Olsen *et al.* 2010).
5. *Binary outcome* - An outcome that has two possible endpoints such as dead/alive, occurrence/no occurrence or success/failure. Binary variables are a sub-type of dichotomous variables. This type of outcome is usually used in clinical studies, including those that are interventional and observational (Petrie & Sabin 2000). For instance, the possible endpoint when evaluating the effect of oral premedication on the anaesthetic success of inferior alveolar nerve blocks for mandibular molars, while performing root canal treatment is the occurrence of pain or no pain.

6. *Cochrane* – A leading international organization for evidence synthesis, especially systematic reviews and meta-analyses in relation to health (https://www.cochranelibrary.com/). The organization is legally registered in the United Kingdom. Cochrane's mission is to promote evidence-informed healthcare decision-making by publishing high-quality systematic reviews and other research evidence. To date, Cochrane has 11,000 members with over 68,000 supporters from >130 countries (Cochrane 2019).
7. *Confidence interval* (CI) - A statistical measure for quantifying uncertainty or probability around the sample mean. It gives a range of values in which one can be confident the true population value will lie within. The CI is normally used to indicate the level of precision associated with an estimate from a sample. In other words, large sample sizes usually provide narrow confidence intervals and thus more precision, while small sample sizes provide wide confidence intervals and less precision. The CI is calculated based on a pre-specified level of significance. If the pre-specified level of significance is 95%, we can be 95% confident that the true population value will lie with the calculated interval (Ward *et al*. 2012). For example, a study that determined the mean level of blood haemoglobin A1C (HbA1C) in a sample as 7.2 mg%, with its corresponding 95% confidence interval as 6.5 mg% to 7.9 mg% means we can be 95% confident that the mean of HbA1C in the population (not sample) is within 6.5 mg% to 7.9 mg%.

8. *Confounding factors* – Additional or background factors other than the exposure/treatment/intervention that can have an impact on the outcome(s) of a treatment or intervention. Confounding factors can lead to spurious results and mask the true relationship between the exposure/treatment/intervention of interest and the outcome(s) (Skelly *et al*. 2012).
9. *CONSORT* - CONsolidated Standards Of Reporting Trials. The CONSORT statement is a minimum set of recommendations for reporting randomised clinical trials that comprises a checklist of 25 items and a flow diagram to depict the flow of participants through a trial. The use of CONSORT by the authors of trials facilitates complete and transparent reporting as well as helping during the critical appraisal and interpretation of the results of a trial (Moher *et al*. 2012).
10. *Continuous outcome* - An outcome that has a numeric value on a scale. For instance, the outcome of treating patients with a new diabetic medication is haemoglobin A1C (HbA1C), which is recorded in numeric values such as 7.5 mg%. However, a continuous outcome can be transformed into a binary outcome (Item 5) by grouping it into categories, so-called categorical data (Petrie & Sabin 2000). For example, HbA1C that is a continuous outcome could be categorized into a binary outcome by defining it as <7.0 mg% to indicate the achievement of a treatment goal, while HbA1C≥ 7.0 mg% would indicate not achieving the treatment goal. Thus, patients with a HbA1C of 7.5 mg% would be classified as not achieving the treatment goal.

11. *Direct evidence* – A technical term used in a network meta-analysis (Item 33). It represents the evidence derived from studies that directly compared one intervention with another intervention (Lu & Ades 2004, Lumley 2002). For example, a network meta-analysis that compared interventions A, B, and C could include several randomised controlled trials that revealed comparative evidence of intervention A versus B and intervention A versus C. However, no study compared intervention B versus C. Thus, there was direct evidence of intervention A versus B and direct evidence of intervention A versus C, but no direct evidence of intervention B versus C.
12. *Drop-outs* – The attrition or loss of subjects in one or more arms of a trial, sometimes called, non-participation. Drop-outs can occur for a variety of reasons including the effects of the drug/intervention, the unavailability of participants for review or their refusal to take a further part in the trial. If the number of drop-outs is substantial it can affect the validity of the results and the conclusions (Bell *et al*. 2013). It is imperative that the number and reason for subjects dropping-out are reported in detail within a trial.
13. *Effect size* – The magnitude of difference between groups, when two or more interventions are being compared; it is also called the treatment effect or effect estimate. For continuous outcomes (Item 10), the effect size is the difference between the mean value of the intervention and comparator, also called the mean difference.

However, in some situations within a meta-analysis where studies were conducted using different measurements or scales, the mean difference cannot be used for pooling the effect size. Therefore, the standardised mean difference (Item 61) is used to transform the effect to a scale that is easily understood. For binary outcomes (or categorical outcomes), the effect size can be presented using the odds ratio, risk ratio, rate ratio, hazard ratio, or other measures (Sullivan & Feinn 2012).

14. *Egger's test* - Is a test for small study effects (Item 60). It examines whether the association of treatment effects and their variance is greater than that expected by chance. Egger's test uses the linear regression approach by associating the treatment effect estimate against its standard error weighted by the inverse variance of the treatment effect (Egger *et al.* 1997, 2008). P<0.10 indicates that the association of treatment effects and their variance is greater than chance, which means that a small study effect has been detected. See Begg's test (Item 3).
15. *Eligibility criteria* - Specific conditions that define the criteria for the inclusion/exclusion of studies in a systematic review in order to answer the research question (Item 53). These criteria are framed on the basis of the population, intervention (or exposure), control, outcome and study design (PICOS) in a systematic review. Adherence to the eligibility criteria prevents bias in the selection of studies, which is generally not observed in a narrative review (Higgins & Green 2011, McCrae *et al.* 2015).

16. *Evidence gap map* – A tool for collecting and summarising existing evidence to inform policy decision-making and to prioritize future research. It maps out existing and ongoing primary studies or systematic reviews in a particular sector or subsector in terms of the types of policies or programmes evaluated and the outcomes measured. It is usually presented as a visualized map and highlights the gaps where no (or little) evidence exists (International Initiative for Impact Evaluation [3ie] 2019).
17. *Fixed-effect model* – A statistical method for combining the findings from the included studies. It uses a principle of weighted average to combine the findings across studies using the effect size and variance of each study. However, a fixed-effect model is valid only under the assumption that all effect sizes are estimating the same underlying intervention effect (Deeks *et al.* 2008).
18. *Forest plot* – Provides a graphic summary of the size of the overall effect of all the included studies from a meta-analysis within a systematic review. The weighted effect sizes and statistical heterogeneity are displayed as a plot with the summary effect size depicted as a diamond figure within the plot.

Figure 1 illustrates a forest plot of a hypothetical meta-analysis using random-effects modelling to evaluate the efficacy of intracanal medicaments to prevent the recurrence of apical periodontitis compared to a control (without intracanal medicament) that gives the summary estimates as risk ratio (RR) with 95% confidence interval (CI). Five studies are illustrated in the plot. The x-axis forms the effect size

scale. In each row (horizontal lines), the treatment effect for the individual study is represented by a blue square with the horizontal line representing the corresponding 95% confidence interval (CI). The size of the blue square is proportional to the percentage weight, which is an indication of the influence each study had on the pooled result of the meta-analysis. Therefore, the larger the square the more "weight" the study has in the meta-analysis. The weight (in %, provided in the right-hand column of Figure 1) is the percentage weight given to the study in the pooled meta-analysis. The general rule is that the larger the study in terms of sample size, the greater the percentage of weight (influence) given to that study. The vertical line in the middle of the plot is where the intervention and the placebo have had the same effect on recurrence of apical periodontitis. In other words, there is no difference in the recurrence of apical periodontitis with and without an intracanal medicament. This line is called the "line of no effect". If the horizontal line in each row touches or crosses the line of no effect, it means that the trial found no significant difference between the two treatments (intervention and control). Among the five studies, only one study (AC, 2014) does not touch the line of no effect, which means for that study alone there is a significant difference in the recurrence of apical periodontitis between the intervention and control group. The bottom row (or "summary row" or "overall") in Figure1 represents the result of the meta-analysis. The red diamond along the bottom row represents the pooled quantitative result from the meta-analysis. The dotted vertical line running along the middle of the diamond is the result of the pooled meta-

analysis for the treatment effect, and the horizontal width of the diamond is the certainty of the result, again usually presented as a 95% confidence interval. As in this example, if the CI does not cross the line of no effect, then there is a significant difference between the effectiveness of the intervention and control group.

Interpretation of the results: If the RR is 1 (or the 95% CI includes the value 1), it suggests no difference or little difference in risk (i.e. the incidence in each group is the same); A RR > 1 (and the 95% CI does not include the value 1) suggests an increased risk of that outcome in the intervention group; A RR < 1 (and the 95% CI does not include the value 1) suggests a reduced risk in the intervention group. According to the summary estimates in Figure 1, RR is 0.75 and the 95% CI does not include the value 1 (0.58 to 0.96), that is, in relation to the 95% confidence interval the p-value is <0.05, that is, it is significant. At the same time, the position of the red diamond favours the intervention as it does not touch or cross the line of no effect. In traditional terminology, this means that the meta-analytic effect is statistically significant. The corresponding Z-value and p-value (one-tailed and two-tailed) are also provided below the Forest plot. The corresponding p-value is <0.05 and also indicates statistical significance. Hence the pooled quantitative result of the meta-analysis is statistically significant and a reduced risk of reoccurrence for apical periodontitis is expected from the intervention group (with intracanal medicament) compared to the control

group (without intracanal medicament). In other words, the intervention reduced the risk of apical periodontitis by 25% (% risk reduction= (1-RR) x 100 = (1-0.75) x 100 =25%). Heterogeneity (I^2) (Item 23) is expressed as a percentage with a range between 0 and 100. In the example, I^2 is 0% with p>0.05, suggesting there is no study heterogeneity.

Figure 2 is a forest plot that summarises the efficacy of intracanal medicaments to prevent the recurrence of apical periodontitis compared with controls (without intracanal medicament). The overall summary estimate states that RR is 0.91 and the 95% CI includes the value 1 (0.73 to 1.13). At the same time, the red diamond touches the line of no effect, which in traditional terminology, means that the meta-analytic effect is not statistically significant. Therefore, overall, no difference occurred between the intervention (with intracanal medicament) and control (without intracanal medicament).

In Figure 3, the overall summary estimates, RR is 1.81 and the 95% CI does not include the value 1 (1.37 to 2.39). At the same time, the position of the red diamond favours the control group as it does not touch or cross the line of no effect. This shows that the meta-analytic effect is statistically significant. The inference is that those who were exposed the intervention had an 81% (% increase= (RR-1) x 100 = (1.81-1) x 100 = 81%) increased risk of reoccurrence of apical periodontitis compared to the control. In

other words, the intervention increased significantly the risk of apical periodontitis compared with the control.

19. *Funnel plot* - Provides an indication of publication bias (Item 46) or small-study effects (Item 60) that could affect the validity of the results of a meta-analysis. By plotting the standard error of the effect size on a reversed scale on the y-axis, those studies with large sample sizes are placed at the top of the graph with smaller studies scattered on the lower part of the plot. The plot is bounded by a triangle formed from the summary measure with 95% confidence interval estimated from the fixed-effect meta-analysis. This triangle will include 95% of the studies when publication bias is absent. The symmetry of the smaller studies on both sides of the plot or clustering of larger studies on the top also indicates no / low possibility of publication bias (Sterne *et al*. 2011). Figure 4 is a funnel plot with 15 studies (15 dots). It is close to symmetrical and confirms that the smaller studies were distributed relatively symmetrically.
20. *Generalizability* - The ability, based on a researcher's opinion, to decide to what extent the findings of a study can be extrapolated from the sample to an entire population. It has typically been discussed as one aspect of external validity (Flick 2018).
21. *GRADE* - Grading of Recommendations, Assessment, Development and Evaluations. GRADE is a framework that evaluates the quality, certainty, and level of evidence for developing clinical practice recommendations (Guyatt *et al.* 2008). The evaluation is based on the risk of bias (risks of systematic errors) (Item 54), imprecision (risks of

random errors) (Item 24), inconsistency (Item 25), indirectness (Item 27), and publication bias (Item 46).

22. *Grey literature* - Literature that is not formally published in sources such as books or journals, but includes academic papers, dissertations, research and committee reports, government reports, newspaper, conference abstracts, and ongoing research (Paez 2017). Examples of databases for grey literature are OpenGrey (www.opengrey.com), New York Academy of Medicine Grey Literature Report (http://www.greylit.org/), Bielefeld Academic Search Engine (BASE) (https://www.base-search.net/). Grey literature is unpublished work or work published in a non-commercial manner.
23. *Heterogeneity* - Is a measure of variability among included studies in a systematic review. Diversity among studies included in a systematic review is common, at least in terms of clinical and/or methodological perspectives. Therefore, this effect needs to be assessed and considered when the meta-analysis is performed, in order to estimate the pooled effect from a set of similar studies. Determining statistical heterogeneity can be achieved using a chi-squared test. A low P-value (usually lower than 0.10) or a large chi-squared statistic relative to its degree of freedom, highlights evidence of heterogeneity associated with the effects of an intervention. Another method is to look at the I^2 value which can be derived from the Q statistic (Cochran's Q). According to Cochrane handbook. I^2 of 0% - 40% suggests low heterogeneity that is not

important, 30% - 60% indicates moderate heterogeneity, 50% - 90% indicates substantial heterogeneity, while 75% - 100% indicates considerable heterogeneity (Deeks *et al*. 2019).

24. *Imprecision* - Is one of the criteria for grading the quality of evidence developed by the GRADE (Item 21) working group. The overall criteria consist of the risk of bias, publication bias, imprecision, inconsistency, and indirectness. Imprecision focuses on the variability around the treatment effect measured by a CI. It represents the confidence of reviewers in the observed treatment effect. A wider CI indicates higher imprecision, while a narrower CI indicates lower imprecision. (Guyatt *et al*. 2011a).
25. *Inconsistency* – Confidence in a body of evidence is increased when there are a large number of studies that demonstrate the same consistent effects. Inconsistency means that there is an unexplained heterogeneity in the findings from a meta-analysis. If there is important unexplained heterogeneity in the findings, the confidence in the estimate of the effect for that outcome decreases. The degree of inconsistency could be assessed through the similarity of point estimates between direct evidence and indirect evidence, the overlap of their CIs, and heterogeneity among studies.
26. *Indirect evidence* – A technical term used in a network meta-analysis (Item 33) that indicates the evidence generated by a network meta-analysis through a common comparator (Lumley 2002, Lu & Ades 2004). For example, a network meta-analysis comparing interventions A, B, and C. Several randomised controlled trials revealed

comparative evidence of intervention A versus B and intervention A versus C. However, no study was conducted to compare intervention B versus C. Indirect evidence of intervention B versus C can be estimated through the association of intervention A versus B and intervention A versus C using intervention A as the common comparator.

27. *Indirectness* – Is one of the criteria for grading the quality of evidence developed by the GRADE (Item 21) working group. The overall criteria include risk of bias, publication bias, imprecision, inconsistency, and indirectness. Indirectness focuses on the existence of research that directly compares the interventions of interest in the population with important outcomes for patients. Evidence with high indirectness means the population, interventions, and outcomes of research under consideration, differ from those in which we are interested. In addition, evidence with substantial indirectness occurs when there are no head-to-head or direct comparisons between the interventions of interest (Guyatt *et al.* 2011b).

28. *Intention-to-treat analysis* – An analytical approach that aims to compare outcomes based on a group of patients allocated after randomisation, even though they may not have fully complied with the indicated treatment as the study progressed (e.g. a patient who was not taken their medication for the duration of the study, patient who dropped-out, etc.). The results obtained from an intention-to-treat analysis will estimate what happens in the "real world", where the situation is not perfect. This

approach is commonly used when assessing the results of clinical trials, preferably when analysing a superiority trial (Higgins *et al*. 2019).

29. *Meta-analysis* – A statistical technique that pools the findings, results or data of several independent studies included in a systematic review (Haidich 2010). A meta-analysis mimics the conduct of a larger study by combining the data from several individual studies included a systematic review and, as a consequence, provides more precise summary estimates (Moher *et al*. 2015). In health sciences, meta-analyses are cited more often than other study designs (Patsopoulos *et al*. 2005).
30. *Meta-regression* – *Meta-regression* is a statistical method to investigate heterogeneity (Item 23) within a meta-analysis. It allows the meta-analyst to investigate the effect of potential effect modifiers on the pooled effect estimates, whether the potential effect modifiers are continuous or categorical variables. Meta-regression is similar to a simple regression. The treatment effect is the outcome variable, while study' characteristics are considered as explanatory variables that might affect the treatment effect. However, there are two differences between a meta-regression and a simple regression. Firstly, a meta-regression considers the sample size of each study within its analysis. Studies with large sample sizes have more influence on the relationship between the outcome variable and explanatory variables than studies with a small sample sizes because studies are weighted by the variation of their treatment effect. Secondly, a meta-regression does not consider heterogeneity among treatment effects within its analysis. However, a meta-regression should not be

considered when the number of included studies is fewer than 10 because the statistical power is insufficient to detect significant differences (Deeks *et al*. 2008, Harbord & Higgins 2016).

31. *Mixed treatment comparison* - Mixed treatment comparison is an infrequently used term used interchangeably with a network meta-analysis. Multiple treatment meta-analysis has a similar meaning (Chaimani *et al*. 2019).
32. *Narrative review* – A traditional literature review designed to (critically) describe and discuss the information available from books and journals on a specific topic from a theoretical and contextual point of view. Although the norm in science, narrative reviews are not particularly useful in clinical research as they are not reproducible, often biased, nor do they answer a specific quantitative research questions as they do not follow a strict methodological approach such as using inclusion/exclusion criteria, specific literature databases, search processes, nor use a quality assessment tool to evaluate the included studies.
33. *Network meta-analysis* – Compares multiple interventions (more than two) by evaluating and analysing direct and indirect comparisons among trials. Direct comparison is similar to a conventional pair-wise comparison when studies comparing interventions are available. Indirect comparisons are undertaken against a common comparator (usually placebo or an existing gold standard intervention). A network meta-analysis ranks the various interventions from the most to least effective. The basic concept behind a network meta-analysis is as follows: consider

three individuals, X, Y and Z. If Y is 7 kg heavier than X, and Z is 10 kg heavier than X, then we know that Z is 3 kg heavier than Y, and is therefore the heaviest of the three. The individuals can also be ranked in terms of who is heaviest as 1st =Z, 2nd =Y and 3rd =X. So, by taking the weight of individual X as a reference and measuring the weights of the others compared with theirs, the weight of everyone can be compared and how they are ranked (in order) by their individual weight can be deduced.

34. *Network plot* - Plots a network of interventions as a visual presentation of the evidence-base and offers a concise explanation about its characteristics. Figure 5 is a network plot of a hypothetical situation, which aimed to find the answer for the research question: what is the most effective oral medication in reducing post-operative pain in adults following root canal treatment? For example, in the endodontic literature, three drugs (A, B and C) have been studied. However, so far, no clinical trials have been conducted to compare the effectiveness of drug C with drugs B and A. In this situation, a network meta-analysis compares the effectiveness of drugs A, B and C by evaluating and analysing direct and indirect comparisons among trials to find the answer for the research question above. The network plot consists of nodes representing the interventions being compared. In Figure 5, four interventions (nodes) shown as coloured circles were compared including a placebo, while the lines connecting the nodes represent the existing direct evidence which means there is at least one study comparing the effect of that pair of interventions (also called contrast). The size of each node usually represents the sample size within the intervention with

larger nodes indicating larger sample sizes. The thickness of the line represents the number of studies comparing the pairs of interventions with thicker lines indicating larger numbers of studies.

The edge thickness (the connecting line between two interventions) is proportional to the number of studies comparing the two interventions, thus illustrating which interventions were more frequently compared. In Figure 5, the line between B and the placebo is thicker, which means more clinical trials had been conducted between these two interventions. On Figure 6, there is direct evidence available (line) for studies comparing A versus placebo, B versus placebo, and A versus B. However, no direct evidence (line missing) is available for C versus B and C versus A. In the absence of direct comparison among interventions, a network meta-analysis synthesizes both direct and indirect evidence in a network of trials that compares the numerous interventions and has the ability to rank the drugs.

35. *Odds ratio* – Is the ratio between the odds of exposure to a factor among people with a condition or disease (cases) and the odds of exposure to a factor among those who do not have the condition or disease (controls). This ratio is a statistic that quantifies and represents the strength of the association between an exposure and a disease in a case-control study. For interpretation, the odds of exposure in the two groups is the same if the odds ratio is equal to 1. The odds of exposure are considered to be lower

or greater in the diseased group if the ratio is less than 1 or more than 1, respectively (Bonita *et al*. 2006).

36. *Outcome* - The effect of a treatment or drug or technique that can be measured reliably. Within randomised clinical trials there are generally primary and secondary outcomes. Examples: postoperative pain assessed using a visual analogue scale or periapical healing assessed using standardised periapical radiographs.

37. *Per-protocol analysis* – A comparison of interventions that includes only those participants who completed the treatment they were originally allocated. Example: a randomised clinical trial designed to compare the effect of a newly developed irrigant with chlorhexidine on the outcome of root canal treatment after 24 months. If some participants drop-out of the clinical trial before examiners measured the outcome, a per-protocol analysis would not combine their results with those participants who completed treatment. In non-inferiority trials, both intention-to-treat (item 28) and per-protocol analysis are recommended; both approaches should support non-inferiority.

38. *PICOS format* - A structured approach for developing research questions within a systematic review that includes five components. The acronym ''PICOS'' where each letter stands for: the patient population or the disease being addressed (P), the interventions or exposure (I), the comparator (C), the outcome or endpoint (O), and the study design chosen (S). Example: Does mineral trioxide aggregate (I) compared to

formocresol (C), result in better clinical success (O) in primary molar teeth undergoing pulpotomy (P) in randomised clinical trials (S)?

39. *Power* - From a statistical perspective, power is the probability of rejecting the null hypothesis when the null hypothesis is false (correct decision). To simplify, power is the probability of detecting a statistically significant difference when two interventions are truly different. However, there might be a chance that two interventions are deemed not to be different when they are actually truly different (called Type II error). Thus, statistical power is not 100% perfect. In the medical literature, a statistical power of 80% or more is accepted as sufficient, which means that the acceptable level of probability to make a correct decision should be at least 80% or a 20% error is accepted to determine that two interventions are not different even though they are different (Petrie & Sabin 2000).
40. *Precision* - The effect estimate from a particular study may carry a degree of uncertainty and less power to detect a significant result. Therefore, by using a meta-analysis the pooled effect estimates from similar studies can improve the precision of the treatment effect, which can be determined by the confidence intervals. Wider intervals indicate greater uncertainty (less precision). However, there is no exact value to determine whether the precision is too low. Readers need to make a judgement to consider whether the interval is too wide for the degree of precision (Schünemann *et al.* 2019).

41. *PRISMA* – Preferred Reporting for Systematic Reviews and Meta-Analyses (http://www.prisma-statement.org/). PRISMA provides a minimum set of items that are required to ensure the quality of reports describing systematic reviews and meta-analyses. It consists of a checklist with 27-items and one four-phase flow diagram (Liberati *et al.* 2009). PRISMA is recommended by numerous journals in the medical and dental fields.
42. *PRISMA-P* – Are guidelines to facilitate the development of a protocol prior to carrying out a systematic review. The protocol *a priori* reports on the purpose and methodology of the systematic review, which avoids any bias that could occur during its conduct. PRIMSA-P was developed from items within the PRISMA checklist and PROSPERO register (Moher *et al.* 2015).
43. *PROSPERO* – An international database of prospectively registered systematic reviews. It is a free online portal that facilitates the publication of the protocol of systematic reviews related to health. The primary aim is to document *a priori* protocols, avoid duplication of reviews at the developmental stage, provide transparency in the conduct of reviews and aid in the evaluation of reporting bias (Moher *et al.* 2015).
44. *Protocol* - A protocol of a systematic review and meta-analysis should be established *a priori* to describe the purpose, research question and methods of the systematic review in advance of it being conducted. Protocols should be published (e.g. Cochrane (Item 6), PROSPERO (Item 43) and made available before, during and after a

systematic review is carried out. Protocols of reviews should contain a rationale, *a priori* methodology, quality assessment and the statistical approach to be adopted during the review (Moher *et al*. 2015).

45. *P-score* - A statistical method used to rank treatment hierarchy in a network meta-analysis, based on a frequentist approach (a method to undertake a network meta-analysis). This value is derived from the frequentist point estimate and its standard error. P-scores can be considered as equivalent to a SUCRA value (Item 63), which both allow treatments to be ranked on a continuous 0-1 scale (Rucker & Schwarze 2015).
46. *Publication bias* – A form of reporting bias resulting from the preferred publication of papers with positive or acceptable results. This can lead to negative results being under-represented in the published literature and subsequent systematic reviews. A systematic search leads to the selection of all results, positive or negative, and should be reflected in the results of a systematic review. Publication bias can be evaluated through a funnel plot (*Item* 19).
47. *Quasi-randomized controlled trials* – In this study design, the participants are allocated to different arms of the trial (to receive the intervention or placebo) using a method of allocation that is not truly randomised. In quasi-randomisation there is a greater chance of risk that the investigator will be aware of which participant is in which group, in other words, the risk of selection bias will be introduced.

48. *Random effects model* - A meta-analysis approach that incorporates an assumption that the various included studies are estimating the effects of different interventions (Deeks *et al*. 2019). Opposite to fixed-effect model, a random effects model assumes that studies included in a meta-analysis are from a random sample of existing studies and the observed treatment effect is the average of the true treatment effects within a relevant distribution. The random effects model will give a similar result to a fixed-effect model when heterogeneity among the studies is not observed.
49. *Randomisation* – A mandatory step in a randomised clinical trial that provides an equal chance that all selected participants can be allocated to the available intervention groups. It is an accepted method to avoid selection bias. Various types of randomisation methodology can be used in studies, including simple, stratified and block techniques.
50. *Randomised clinical trial* – A prospective experiment (usually medical or health-related) that compares the effectiveness of two or more interventions/treatments. Normally, the test intervention being evaluated is compared against a standard treatment (gold standard) or no treatment (placebo).
51. *Rankogram* - Is a graphical tool to present the possible ranking of each intervention among all the comparisons in the network meta-analysis (https://methods.cochrane.org/cmi/glossary). This two-dimensional graph is plotted between the probability for each intervention to have a specific rank in the vertical

axis and the number of possible ranks in the horizontal axis (Chaimani *et al.* 2013, Salanti *et al.* 2011).

52. *Realist review* - This type of qualitative review also uses a systematic review approach, but is designed to review whether social or policy interventions work (or not) in a 'real world' or in a practical sense. It aims to provide an explanatory analysis of how and why they work in particular 'real world' contexts or settings. Not tasked with necessarily establishing the most academic-based solution, rather a realist review aims to answer "what works for whom in what circumstances, in what respects and how?", with respect to community intervention programs (Pawson *et al.* 2005). The literature included in realist reviews should focus on real-world studies and pragmatic trials. These patterns predict which aspects of the intervention make it pragmatically effective, and which are required to replicate that success across a range of contexts (Gwyther *et al.* 2018)
53. *Research question* – A well-formulated research question will guide many aspects of the review process, such as inclusion/exclusion criteria, searching for studies, data collection from included studies, and presenting the findings. The research question should be clear and focused - not vague, too specific or too broad. The review question should specify the types of population (participants), types of interventions with comparisons and the types of outcomes that are of interest. The acronym PICOS helps

to serve as a reminder of these. Good research questions may have a narrow or broad focus, depending on the overall objectives of the review.

54. *Risk of bias tool* - A tool for assessing the risk of bias in studies included in systematic reviews. Briefly, the biases which might occur in studies consist of selection bias, performance bias, attrition bias, detection bias, and reporting bias. The tool is used to assess those biases within studies. Several versions of the tool have been introduced based on the design of the original studies included in a systematic review such as risk of bias tool for randomised controlled trials of intervention (ROB tool) (Higgins *et al.* 2008, Sterne *et al.* 2016), risk of bias tool for non-randomised studies of intervention (ROBINS-I tool) (Sterne *et al.* 2016), risk of bias tool for systematic review (ROBIS tool) (Whiting *et al.* 2016), and risk of bias tool for animal studies (SYRCLE tool) (Hooijmans *et al.* 2014).
55. *Risk ratio* - Is the ratio of the risk of occurrence of a disease among exposed individuals to that among unexposed individuals. It is used when assessing the likelihood that an association represents a causal relationship. The interpretation of risk ratio is similar to the interpretation of an odds ratio (Item 35) but should be considered as "risk" instead of "odds" (Bonita & Kjellström 2006).
56. *Scoping review* - To determine the scope or coverage of a body of literature on a given topic and provide clear indication of the volume of the literature and studies available as well as an overview of its focus. A scoping review provides an overview of a

potentially large and diverse body of literature related to a broad topic without critically appraising individual studies, whereas systematic reviews combine empirical evidence from a relatively small number of studies associated with a focused review/research question and assesses the quality of the included studies (Colquhoun *et al*. 2014, Munn *et al*. 2018).

57. *Search strategy* – Used *a priori* to search, identify and select the appropriate and relevant evidence to answer the research question of a systematic review. Search terms are identified and arranged to retrieve eligible studies and information from electronic databases, websites and reference lists of the included studies, and previously published reviews.
58. *Selective outcome reporting bias* – The bias due to the selection of a subset of outcomes among recorded outcomes to be reported (Hutton 2000). The selective outcome reporting of some outcomes but not others, depending on the nature and direction of the results. For example, reporting bias can occur due to the tendency for journals to publish only positive results.
59. *Sensitivity analysis* - An analysis of the main outcome scrutinised under alternative assumptions to determine the robustness or stability of the observed outcomes, such as analysis with the exclusion of studies with a high risk of bias, studies with small sample size, or of poor quality. This analysis is useful for synthesising the evidence according to the observed results (Chaimani *et al.* 2019, Deeks *et al*. 2019).

60. *Small-study effects* - The phenomenon that occurs when smaller studies occasionally are associated with different, often larger, treatment effects than large studies, which may indicate publication bias. This effect may be evident when studies with small sample sizes are more likely to be published if the finding are positive. Determining if there is evidence of small-study effects can be achieved using several methods, such as inspection of the funnel plots, Egger's test (Chaimani *et al*. 2017, Debray *et al*. 2018).
61. *Standardized mean difference* – Is used as a summary statistic in a meta-analysis when the included studies report the same outcomes using different measurements or different scales. For example, studies reported the effect of an intervention of pain reduction may use a visual analogue scale with a scale of 0-10, while others may use a visual analogue scale with a scale of 0-100. A meta-analysis could not combine the findings using their original treatment effect. Thus, standardized mean difference (SMD) is an option for combining the findings of studies'. SMD is calculated by using the mean difference between groups divided by its standard deviation. Notably, SMD should not be directly interpreted as a simple mean difference because SMD is reported in a unit of standard deviation. A simple way to interpret the SMD is that 0.2 represents a small effect, 0.5 a moderate effect, and 0.8 a large effect.
62. *Subgroup analysis* - An analysis to investigate whether the observed effect of an intervention is consistent across well-defined types of participant. It is generally performed to determine the effect between the particular patient group, types of

intervention, or types of study. However, subgroup analyses of subsets of participants within studies are sometimes impossible to perform in systematic reviews because of insufficient details within the published literature that prevent the data being extracted (Deeks *et al.* 2019, Tierney *et al.* 2019).

63. *SUCRA* - The "Surface Under the Cumulative RAnking" curve or SUCRA is the area under the curve of cumulative rankograms. SUCRA is usually presented in terms of numbers ranging between 0 and 1, but can be reported as a percentage. For interpretation, the larger the SUCRA number, the better is the position within the treatment hierarchy according to the outcome (https://methods.cochrane.org/cmi/glossary).
64. *Summary measures* – A measure of treatment effect used to report the results of a meta-analysis. In general, the treatment effect can be presented as a summary measure in terms of odds ratio, rate ratio, or relative risk for a dichotomous variable or mean difference for a continuous variable.
65. *Systematic review* – Locates and reviews all the relevant literature that fits pre-specified inclusion/exclusion criteria to answer a specific research question. It uses a clear and systematic methodology to reduce bias in the identification, selection, synthesis, and summary of studies. It provides reliable findings from which conclusions can be used to inform clinical decision making (Moher *et al.* 2015).
66. *Trial sequential analysis (TSA)* - Calculates the required information size (RIS) for a meta-analysis similar to a sample size calculation for a clinical trial. TSA can evaluate

the effect of random error, false positive, and false negative values on the results of a meta-analysis. The RIS that is calculated can be used to monitor the sufficiency of studies to provide a valid and reliable result when an update of a meta-analysis is undertaken.

Figure 6 compares the effectiveness of intracanal medicaments to prevent the recurrence of apical periodontitis compared with controls (without intracanal medicament). TSA were calculated with type 1 error of 5% and type II error of 20% and a required information size (n= 4560) based on the intervention effect suggested by the included trials using a random-effects model for intervention (RRR of 25% (as per meta-analysis provided in Figure 1, the pooled risk ratio is 0.75. hence, risk reduction (RRR in %) is 25% (1-RR) and control group event proportion of 8.87% (Figure 1, (number of patients exposed to events in control group (119)/total number of patients in control group (1284)). The number of patients included in the meta-analysis (n=2954) did not exceed the required information size (red dotted vertical line) and the cumulative z-curve (blue line) did not cross the trial sequential boundary (red-dotted curve), which indicates that the evidence demonstrated by the meta-analysis in Figure 1 is not conclusive.

Figure 7 compares the effectiveness of intracanal medicaments to prevent the recurrence of apical periodontitis compared to controls (without intracanal medicament). TSA were calculated with type 1 error of 5% and type II error of 20% and

a required information size (n= 3730) based on the intervention effect suggested by the included trials using a random-effects model for intervention (RRR of 27%) and control group event proportion (number of patients exposed to events in control group (117)/total number of patients in control group (1284)) of 9.1% (Supplementary Figure 1). Although the number of patients included in the meta-analysis did not exceed the required information size, the cumulative evidence is conclusive for a 27% reduction of risk for reoccurrence of apical periodontitis because the cumulative z-curve has crossed the trial sequential boundary. This indicates that the evidence demonstrated by the meta-analysis (Supplementary Figure 1) is conclusive.

67. *Umbrella review* - Provides the highest levels of evidence as they combine the results of published systematic reviews or/and meta-analyses (Fusar-Poli & Radua 2018). An umbrella review is able to demonstrate whether the evidence-base around a topic or question is consistent or whether it is contradictory or whether discrepant findings exist, and if they do, explores and details the reasons why. Umbrella reviews allow assessment and consideration of whether each included systematic review addressed similar research or review questions, observed similar results independently and arrived at similar conclusions. This form of review is a summary syntheses of the evidence that exists (Aromataris *et al.* 2015).

68. *Weighted mean difference* - Weighted mean difference (WMD), or "mean difference", or "difference in means" is a standard statistic when a meta-analysis is performed to

determine the pooled effects of continuous variables. The absolute difference will be computed between the mean values of the intervention and control group. The weight given to each study is determined by the precision of its estimate of effect. If the outcome measurements from all of the studies are made on a similar scale, then the weighted mean difference can be used as a summary statistic in the meta-analysis (Higgins *et al*. 2019).

Conclusion

This article provides a comprehensive glossary, which contains definitions and explanations for the most commonly used terms employed when conducting and reporting systematic reviews. These terms will benefit authors when planning and producing high quality systematic reviews and meta-analysis in Endodontology. High quality systematic reviews and meta-analyses will minimize the possibility that their results or findings are influenced by bias, which provides confidence to clinicians when applying them in their day-to-day clinical practice.

References

Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P (2015) Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. *International Journal of Evidence-Based Healthcare* **13**, 132-40.

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

Bell ML, Kenward MG, Fairclough DL, Horton NJ (2013) Differential dropout and bias in randomised controlled trials: when it matters and when it may not. *BMJ*. **346**, e8668.

Bonita R, Beaglehole R, Kjellström T. Basic epidemiology: World Health Organization; 2006 [cited August 1, 2019]. 2nd: [Available from: https://apps.who.int/iris/handle/10665/43541

Brok J, Thorlund K, Wetterslev J, Gluud C (2009) Apparently conclusive meta-analyses may be inconclusive--Trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. *International Journal of Epidemiology* **38**, 287-98.

Burns PB, Rohrich RJ, Chung KC (2011) The levels of evidence and their role in evidence-based medicine. *Plastic and Reconstructive Surgery* **128**, 305-10.

Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G (2017) Additional considerations are required when preparing a protocol for a systematic review with multiple interventions. *Journal of Clinical Epidemiology* **83**, 65-74.

Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Chapter 11: Undertaking network meta-analysis. 2019 [cited August 1, 2019]. In: Cochrane Handbook for Systematic Reviews of

Interventions [Internet]. London: Cochrane, [cited August 1, 2019]. Available from: https://training.cochrane.org/handbook/version-6/chapter-10-draft.

Chaimani A, Higgins JPT, Mavridis D, Spyridonos P, Salanti G (2013) Graphical tools for network meta-analysis in STATA. *PLoS One* **8**, e76654-e.

Clarke J (2011) What is a systematic review? *Evidence Based Nursing* **14**, 64.

Cochrane (2019) About us. Retrieved from https://www.cochrane.org/about-us. Accessed 6th September 2019

Colquhoun HL, Levac D, O'Brien KK, et al. (2014) Scoping reviews: time for clarity in definition, methods, and reporting. *Journal of Clinical Epidemiology* **67**, 1291-4.

Debray TPA, Moons KGM, Riley RD (2018) Detecting small-study effects and funnel plot asymmetry in meta-analysis of survival data: A comparison of new and existing tests. *Research Synthesis Methods* **9**, 41-50.

Deeks JJ, Higgins JP, Altman DG (2008) Analysing data and undertaking met-analyses. In J. P. Higgins & S. Green (Eds.), Cochrane Handbook for Systematic Reviews of Intervention. NJ, USA: John Wiley & Son Ltd.

Deeks JJ, Higgins JPT, Altman DG. Chapter 10: Analysis data and undertaking meta-analysis. Draft version (29 January 2019). 2019 [cited August 1, 2019]. In: Cochrane Handbook for

Systematic Reviews of Interventions [Internet]. London: Cochrane, [cited August 1, 2019]. Available from: https://training.cochrane.org/handbook/version-6/chapter-10-draft.

Delgado-Rodriguez M, Llorca J (2004) Bias. *Journal of Epidemiology and Community Health* **58,** 635-41.

Egger M, Davey Smith G, Schneider M, Minder C (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ* **315**, 629-634.

Flick U. The SAGE Handbook of Qualitative Data Collection. 2018 2019/08/21. 55 City Road, London: SAGE Publications Ltd. Available from: https://methods.sagepub.com/book/the-sage-handbook-of-qualitative-data-collection

Forder PM, Gebski VJ, Keech AC (2005). Allocation concealment and blinding: when ignorance is bliss. *Medical Journal of Australia* **182,** 87-9.

Fusar-Poli P, Radua J (2018). Ten simple rules for conducting umbrella reviews. *Evidence-Based Mental Health* **21**, 95-100.

Guyatt GH, Oxman AD, Vist GE, Kunz R, et al (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **336**, 924–6.

Guyatt GH, Oxman AD, Kunz R et al. (2011a) GRADE guidelines 6. Rating the quality of evidence--imprecision. *Journal of Clinical Epidemiology* **64**, 1283-93.

Guyatt GH1, Oxman AD, Kunz R, et al. (2011b). GRADE guidelines: 8. Rating the quality of evidence--indirectness. *Journal of Clinical Epidemiology* **64,** 1303-10.

Gwyther H, Bobrowicz-Campos E, Luis Alves Apóstolo J, Marcucci M, Cano A, Holland C (2018) A realist review to understand the efficacy and outcomes of interventions designed to minimise, reverse or prevent the progression of frailty. *Health Psychology Review* **12,** 382-404.

Haidich AB (2010) Meta-analysis in medical research. *Hippokratia* **14**, 29-37.

Haidich AB (2010) Meta-analysis in medical research. *Hippokratia* ***14***, 29–37.

Harbord RM, Higgins JP (2016). Meta-regression: metareg. In T. M. Palmer & J. Sterne (Eds.), Meta-analysis in Stata: An updated colilection from the Stata Journal. Texas, USA: Stata Press.

Higgins JP, Altman DG, Sterne J (2008) Assessing risk of bias in included studies. In J. P. Higgins & S. Green (Eds.), Cochrane Handbook for Systematic Reviews of Interventions (pp. 187-235). NJ, USA: John Wiley & Son.

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

Higgins JPT, Li T, Deeks JJ. Chapter 6: Choosing effect measures and computing estimates of effect. 2019 [cited August 1, 2019]. In: Cochrane Handbook for Systematic Reviews of

Interventions [Internet]. London: Cochrane, [cited August 1, 2019]. Available from: https://training.cochrane.org/handbook/version-6/chapter-10-draft.

Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial. 2019 [cited August 1, 2019]. In: Cochrane Handbook for Systematic Reviews of Interventions [Internet]. London: Cochrane, [cited August 1, 2019]. Available from: https://training.cochrane.org/handbook/version-6/chapter-10-draft.

Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M, Langendam MW. (2014). SYRCLE's risk of bias tool for animal studies. *BMC Medical Research Methodology* **14**, 43.

Imberger G, Thorlund K, Gluud C, Wetterslev J (2016). False-positive findings in Cochrane meta-analyses with and without application of trial sequential analysis: an empirical review. *BMJ Open* **6**, e011890.

International Initiative for Impact Evaluation (3ie). (2019). What are evidence gap maps? Retrieved from https://www.3ieimpact.org/evidence-hub/evidence-gap-maps

Kattan S, Lee SM, Kohli MR, Setzer FC, Karabucak B (2018) Methodological Quality Assessment of Meta-analyses in Endodontics. *Journal of Endodontics* **44**, 22-31.

Last JM (1995). A dictionary of epidemiology (3rd ed.). NY, USA: Oxford University Press.

Liberati A, Altman DG, Tetzlaff J *et al*. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine* **6**, e1000100.

Liberati A, Altman DG, Tetzlaff J, et al. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine* **6,** e1000100.

Lu G, Ades AE. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* **23**, 3105-24.

Lumley T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine* **21**, 2313-24.

McCrae N, Blackstock M, Purssell E (2015) Eligibility criteria in systematic reviews: A methodological review. *International Journal of Nursing Studies* **52**,1269-76.

Moher D, Hopewell S, Schulz KF et al. (2012) CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *International Journal of Surgery* **10**, 28-55.

Moher D, Shamseer L, Clarke M (2015) Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Systematic Review* **4,** 1.

Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E (2018) Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology* **18,** 143.

Olsen J, Christensen K, Murray J, Ekbom A (2010). An introduction to epidemiology for health professionals. NY, USA: Springer Science+Business Media

Paez A (2017) Gray literature: An important resource in systematic reviews. *Journal of Evidence-Based Medicine* **10**, 233-240.

Patsopoulos NA, Analatos AA, Ioannidis JP (2005) Relative citation impact of various study designs in the health sciences. *JAMA* **293**, 2362-6.

Pawson R, Greenhalgh T, Harvey G, Walshe K (2005) Realist review--a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy* ***10*****,** 21-34.

Petrie A, Sabin C. (2000). Medical Statistics at a Glance. Oxford, UK: Blackwell Sciences Ltd.

Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA (2014) A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Research Synthesis Methods* **5**, 371-85.

Rabin BA, Brownson RC, Haire-Joshu D, Kreuter MW, Weaver NL (2008) A glossary for dissemination and implementation research in health. *Journal of Public Health Management & Practice* **14**, 117-23.

Rucker G, Schwarzer G (2015). Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC medical research methodology* **15**, 58.

Salanti G, Ades AE, Ioannidis JP (2011) Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of Clinical Epidemiology* **64**, 163-71.

Schünemann HJ, Vist GE, Higgins JPT, Santesso N, Deeks JJ, Glasziou P, et al. Chapter 15: Interpreting results and drawing conclusions. 2019 [cited August 1, 2019]. In: Cochrane Handbook for Systematic Reviews of Interventions [Internet]. London: Cochrane, [cited August 1, 2019]. Available from: https://training.cochrane.org/handbook/version-6/chapter-10-draft.

Shea BJ, Grimshaw JM, Wells GA, et al. (2007) Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology* **7**, 10.

Shea BJ, Hamel C, Wells GA (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology* **62,** 1013-20.

Skelly AC, Dettori JR, Brodt ED (2012) Assessing bias: the importance of considering confounding. *Evidence Based Spine Care Journal* **3**, 9-12.

Sterne J, Egger M, Moher D (2008). Addressing reporting biases. In J. Higgins & S. Green (Eds.), Cochrane Handbook for Systematic Reviews of Intervention (pp. 298-325). NJ, USA: John Wily & Son Ltd.

Sterne JAC, Savović J, Page MJ (2019) RoB 2: a revised tool for assessing risk of bias in randomised trials. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ **366**, l4898

Sterne JA, Hernán MA, Reeves BC et al. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* **355**, i4919.

Sterne JA, Sutton AJ, Ioannidis JP et al. (2011) Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* **343**, d4002.

Suebnukarn S, Ngamboonsirisingh S, Rattanabanlang A (2010) A systematic evaluation of the quality of meta-analyses in endodontics. *Journal of Endodontics* **36**, 602-8.

Sullivan GM, Feinn R (2012). Using Effect Size-or Why the P Value Is Not Enough. *Journal of Graduate Medical Education* **4**, 279-282.

The Cochrane Comparing Multiple Interventions Methods Group (CMIMG). A glossary of key terms Bristol: Cochrane; [August 1, 2019]. Available from: https://methods.cochrane.org/cmi/glossary.

Tierney JF, Stewart LA, Clarke M. Chapter 26: Individual participant data. 2019 [cited August 1, 2019]. In: Cochrane Handbook for Systematic Reviews of Interventions [Internet]. London: Cochrane, [cited August 1, 2019]. Available from: https://training.cochrane.org/handbook/version-6/chapter-10-draft.

Uman LS (2011) Systematic reviews and meta-analyses. *Journal of the Canadian Academy of Child and Adolescent Psychiatry* **20**, 57-9.

Ward H, Toledano MB, Shaddick G, Davies B, Elliott P (2012). Oxford Handbook of Epidemiology for Clinicians. Oxford: Oxford University Press.

Whiting P, Savovic J, Higgins JP, et al. (2016). ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology* **69**, 225-234.

Legends

Figure 1: Forest plot evaluating the efficacy of intracanal medicaments to prevent the reoccurrence of apical periodontitis compared to controls (without intracanal medicament) revealing a significant difference between the groups with and without an intracanal medicament in favour of medicaments.

Figure 2: Forest plot evaluating the efficacy of intracanal medicaments to prevent the reoccurrence of apical periodontitis compared to controls (without intracanal medicament) revealing no significant difference between the groups with and without an intracanal medicament.

Figure 3: Forest plot evaluating the efficacy of intracanal medicaments to prevent the reoccurrence of apical periodontitis compared to controls (without intracanal medicament) revealing a significant difference between the groups in favour of the control. Thus, the intervention increased significantly the risk of apical periodontitis occurring compared to the control.

Figure 4: Funnel Plot showing that smaller studies (smaller dots) were distributed relatively symmetrically. (RR – relative risk)

Figure 5: Network plot comparing the effectiveness of three drugs (A, B, C) for reducing postoperative pain following root canal treatment and the demonstrated relationship between them and the placebo.

Figure 6: Trial sequential analysis comparing the effectiveness of intracanal medicaments to prevent the recurrence of apical periodontitis compared with controls (without intracanal medicament). The number of patients included in the meta-analysis (n=2954) did not exceed the required information size (red dotted vertical line on the extreme right) and the cumulative z-curve (blue line) did not cross the trial sequential boundary (red-dotted curve on the upper half of the diagram), which indicates that the evidence derived from the meta-analysis (Figure 1) was not conclusive.

Figure 7: Trial sequential analysis comparing the effectiveness of intracanal medicaments to prevent the recurrence of apical periodontitis compared with controls (without intracanal medicament). The number of patients included in the meta-analysis (n=2954) did not exceed the required information size (red dotted vertical line on the extreme right). However, the cumulative z-curve (blue line) has crossed the trial sequential boundary (red-dotted curve in upper half of diagram). This demonstrates that the cumulative evidence obtained by the meta-analysis (Supplementary Figure 1) is conclusive for a 27% reduction of risk for reoccurrence of apical periodontitis.

Supplementary Figure 1: Forest plot evaluating the efficacy of intracanal medicaments to prevent the reoccurrence of apical periodontitis compared with controls (without intracanal medicament). The summary estimates, RR is 0.73 and the 95% CI does not include the value 1 (0.57 to 0.93). At the same time, the position of the diamond favours the intervention group as it does not touch or cross the line of no effect, which means the meta-analytic effect is statistically significant. The inference is that those who were exposed the intervention had a 27% (% decrease= (1-RR) x 100 = (1-0.73) x 100 = 27%) reduced risk of reoccurrence of apical periodontitis compared to the control.

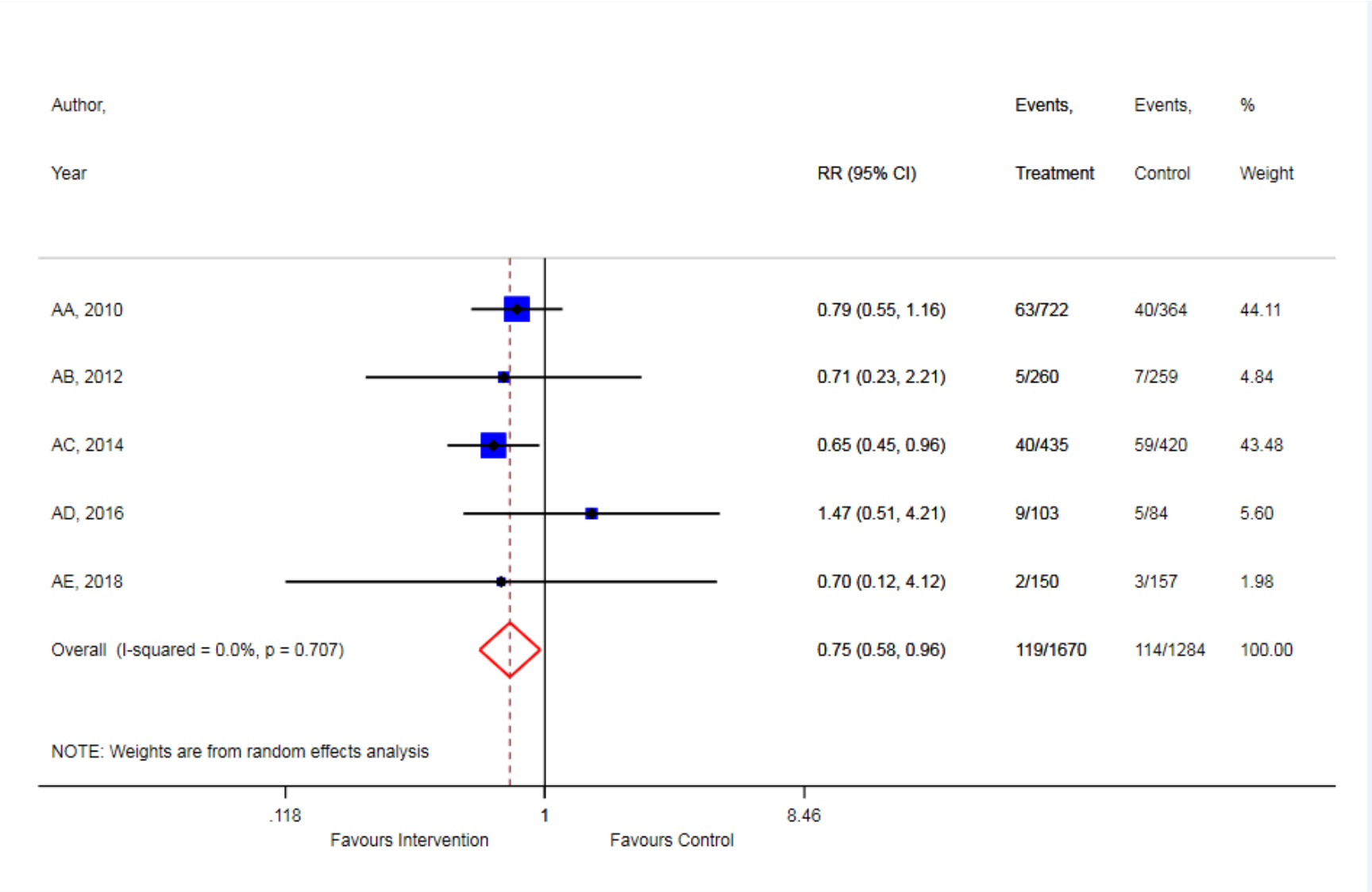
Author,
Year
RR (95% CI)
Events, Treatment
Events, Control
% Weight
AA, 2010
0.79 (0.55, 1.16)
63/722
40/364
44.11
AB, 2012
0.71 (0.23, 2.21)
5/260
7/259
4.84
AC, 2014
0.65 (0.45, 0.96)
40/435
59/420
43.48
AD, 2016
1.47 (0.51, 4.21)
9/103
5/84
5.60
AE, 2018
0.70 (0.12, 4.12)
2/150
3/157
1.98
Overall (I-squared = 0.0%, p = 0.707)
0.75 (0.58, 0.96)
119/1670
114/1284
100.00
NOTE: Weights are from random effects analysis
.118
1
8.46
Favours Intervention
Favours Control

Figure 1

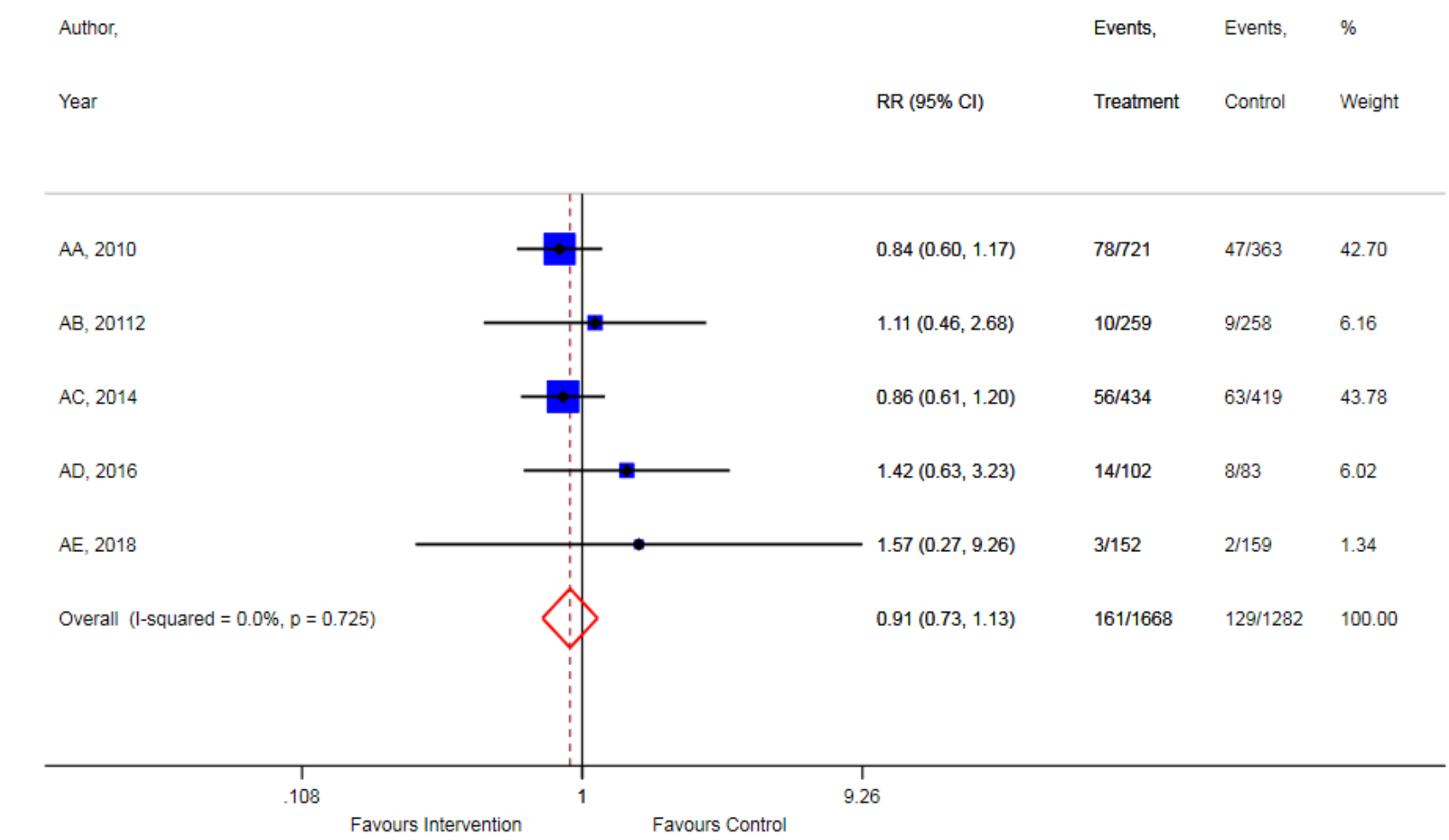
Author,
Year
RR (95% CI)
Events, Treatment
Events, Control
% Weight
AA, 2010
0.84 (0.60, 1.17)
78/721
47/363
42.70
AB, 20112
1.11 (0.46, 2.68)
10/259
9/258
6.16
AC, 2014
0.86 (0.61, 1.20)
56/434
63/419
43.78
AD, 2016
1.42 (0.63, 3.23)
14/102
8/83
6.02
AE, 2018
1.57 (0.27, 9.26)
3/152
2/159
1.34
Overall (I-squared = 0.0%, p = 0.725)
0.91 (0.73, 1.13)
161/1668
129/1282
100.00
.108
1
9.26
Favours Intervention
Favours Control

Figure 2

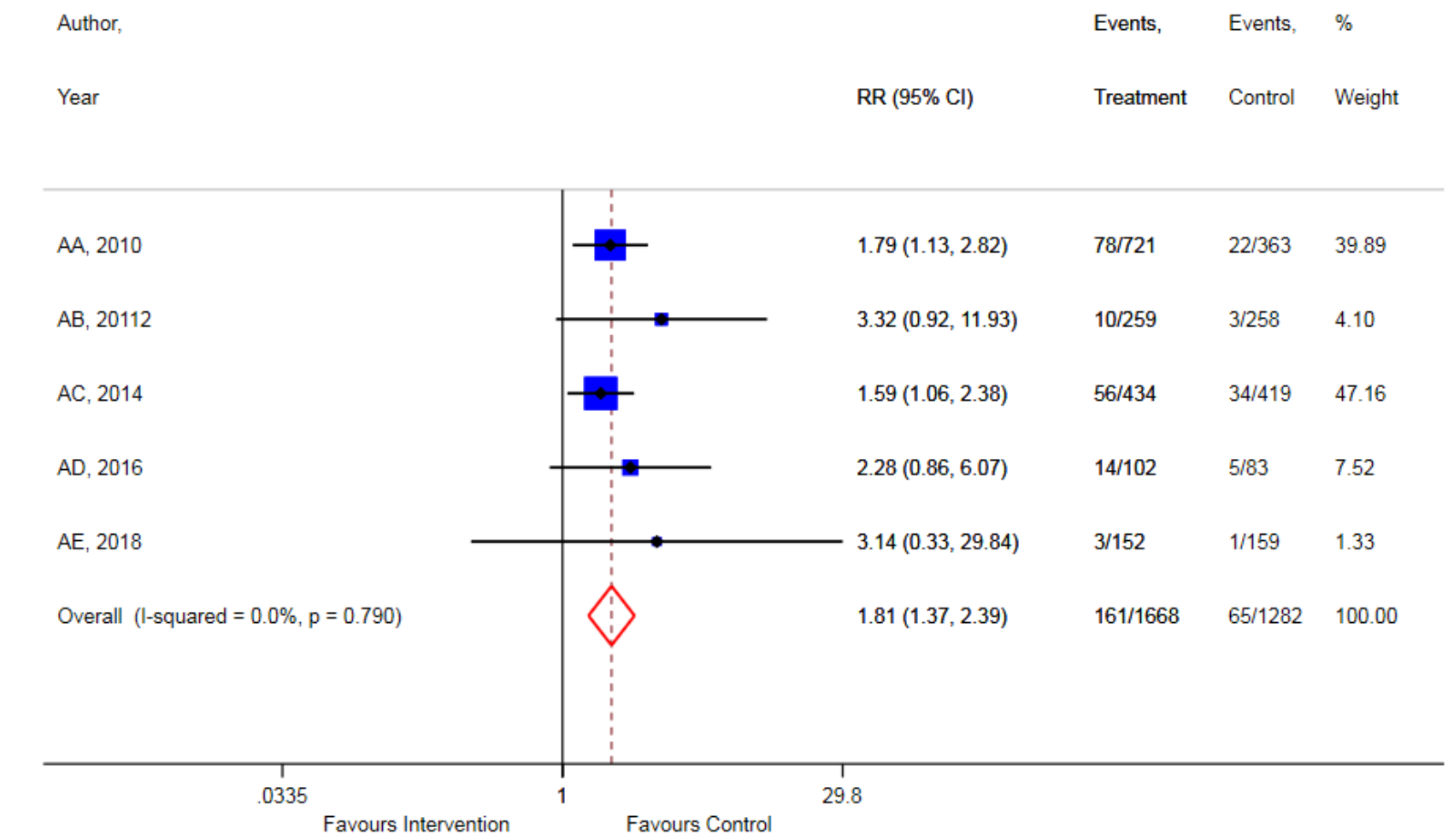
Author,
Year
RR (95% CI)
Events, Treatment
Events, Control
% Weight
AA, 2010
1.79 (1.13, 2.82)
78/721
22/363
39.89
AB, 20112
3.32 (0.92, 11.93)
10/259
3/258
4.10
AC, 2014
1.59 (1.06, 2.38)
56/434
34/419
47.16
AD, 2016
2.28 (0.86, 6.07)
14/102
5/83
7.52
AE, 2018
3.14 (0.33, 29.84)
3/152
1/159
1.33
Overall (I-squared = 0.0%, p = 0.790)
1.81 (1.37, 2.39)
161/1668
65/1282
100.00
.0335
1
29.8
Favours Intervention
Favours Control

Figure 3

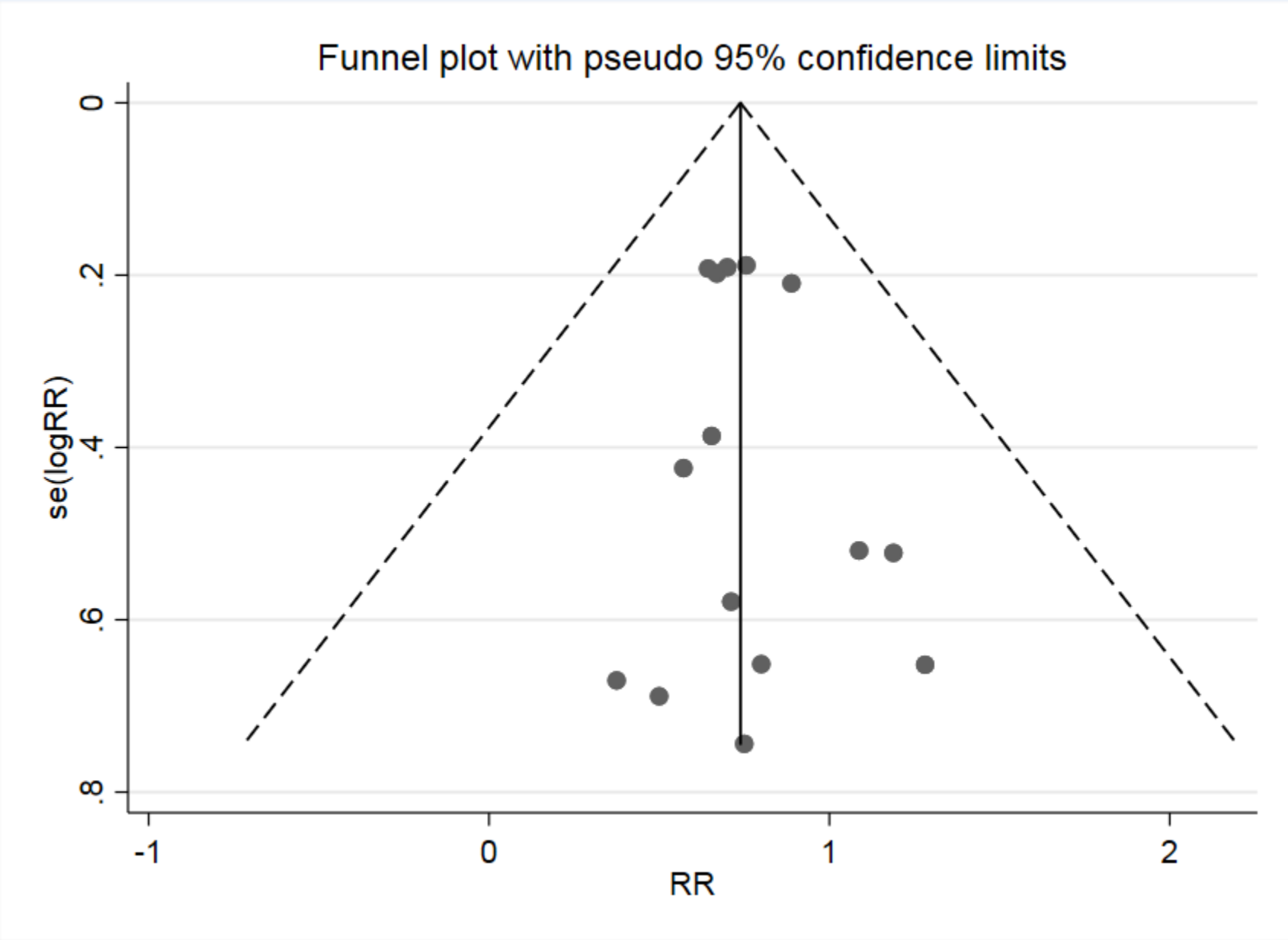
Funnel plot with pseudo 95% confidence limits
se(logRR)
0
.2
.4
.6
.8
-1
0
1
2
RR

Figure 4

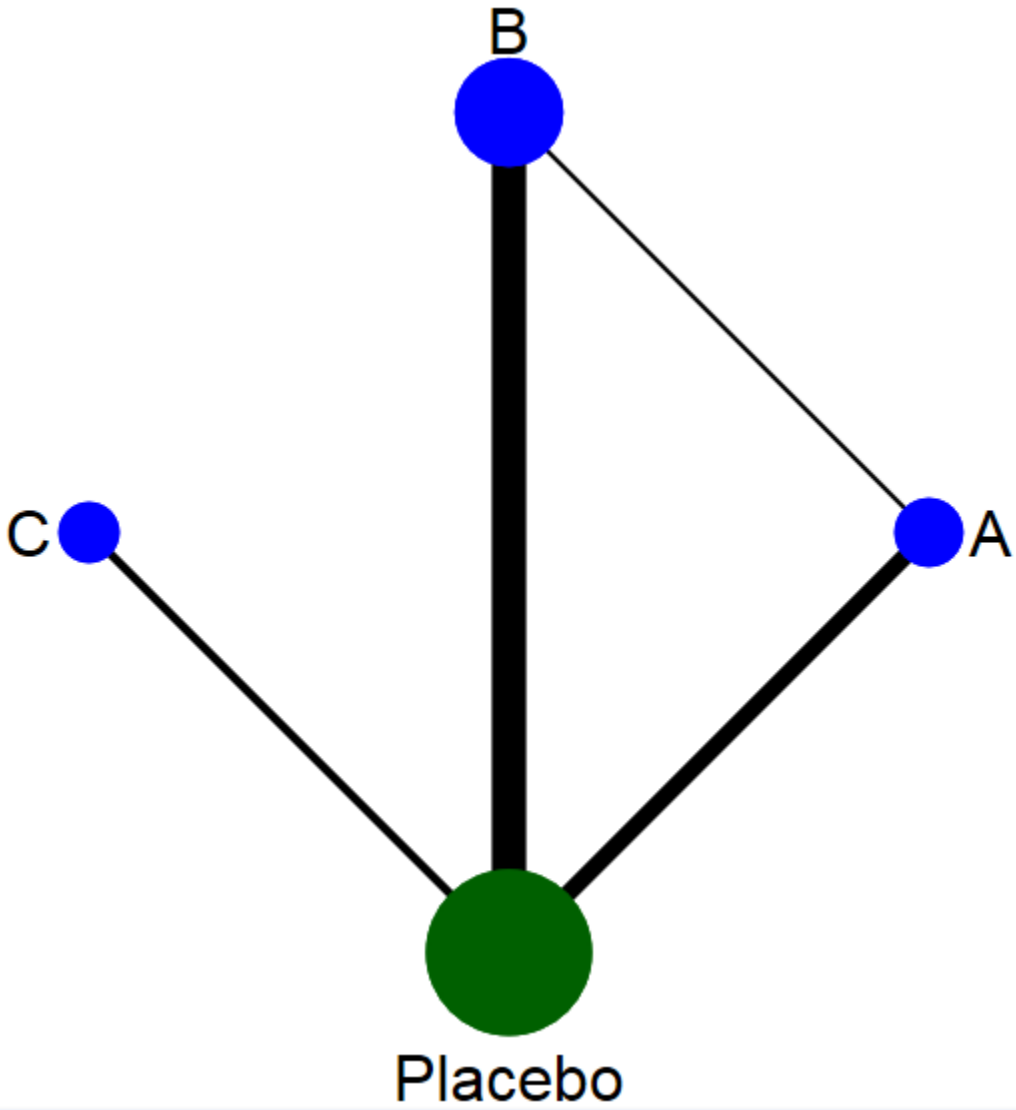

Figure 5

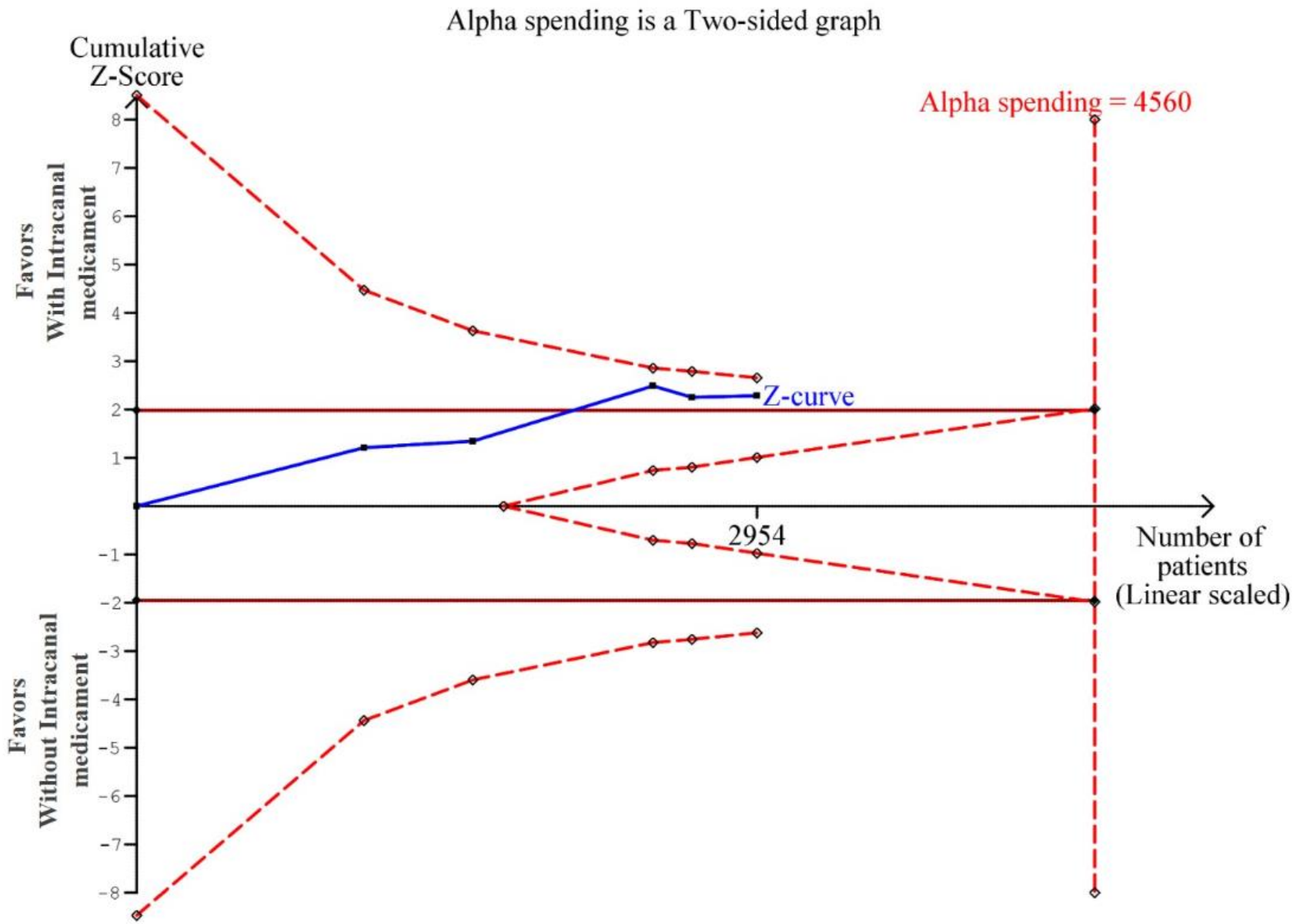
Alpha spending is a Two-sided graph
Cumulative
Z-Score
Alpha spending = 4560
Favors
With Intracanal
medicament
Z-curve
2954
Number of
patients
(Linear scaled)
Favors
Without Intracanal
medicament
8
7
6
5
4
3
2
1
-1
-2
-3
-4
-5
-6
-7
-8

Figure 6

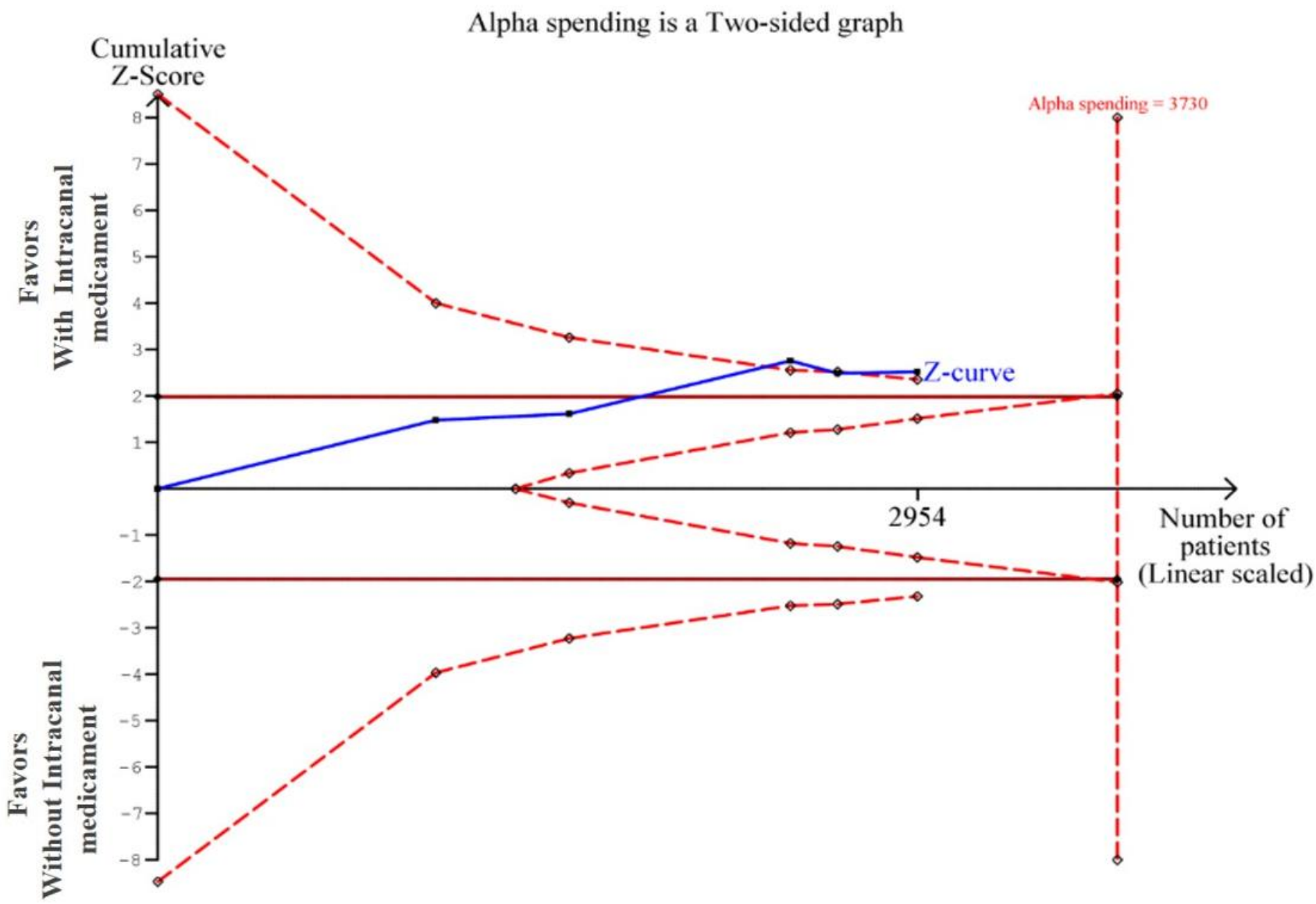
Alpha spending is a Two-sided graph
Cumulative
Z-Score
Alpha spending = 3730
Favors
With Intracanal
medicament
Favors
Without Intracanal
medicament
Z-curve
2954
Number of
patients
(Linear scaled)
8
7
6
5
4
3
2
1
-1
-2
-3
-4
-5
-6
-7
-8

Figure 7

Supplementary Figure 1: Forest plot evaluating the efficacy of intracanal medicaments to prevent the reoccurrence of apical periodontitis compared with controls (without intracanal medicament). The summary estimates, RR is 0.73 and the 95% CI does not include the value 1 (0.57 to 0.93). At the same time, the position of the diamond favours the intervention group as it does not touch or cross the line of no effect, which means the meta-analytic effect is statistically significant. The inference is that those who were exposed the intervention had a 27% (% decrease= (1-RR) x 100 = (1-0.73) x 100 = 27%) reduced risk of reoccurrence of apical periodontitis compared to the control.