“Assessment of reporting quality in Systematic Reviews in the field of Endodontology”

A MSc thesis submitted in fulfilment of the requirements for the degree of Master of Science in Medical Research Methodology

By

Athanasios Fasoulas

Thessaloniki, December, 2018
“Assessment of reporting quality in Systematic Reviews in the field of Endodontology”

A MSc thesis submitted in fulfilment of the requirements for the degree of Master of Science in Medical Research Methodology

At

The Faculty of Health Sciences

School of Medicine

Aristotle University of Thessaloniki

By

Athanasios Fasoulas

Thessaloniki, December, 2018

Supervisors:

Anna-Bettina Haidich

Dimitrios Goulis

Theodoros Lambrianidis

Word count: 7544
Table of Contents:

1. Abbreviations.................................................................3
2. Abstract.................................................................4
3. Introduction..........................................................5
4. Methods.................................................................7
5. Results.................................................................10
6. Discussion............................................................21
7. Conclusions...........................................................27
8. Funding.................................................................28
9. References.............................................................28
10. References of primary studies.................................33
11. Appendix..............................................................39
**Abbreviations:**

**ABH:** Anna-Bettina Haidich  

**AF:** Athanasios Fasoulas  

**CI:** Confidence interval  

**IQR:** Interquartile range  

**MeSH:** Medical Subject Headings  

**OR:** Odds ratio  

**RCT:** Randomized controlled trial  

**RoB:** Risk of Bias  

**SR:** Systematic review  

**TL:** Theodoros Lambrianidis  

**VB:** Vasileios Bousias
Abstract:

Introduction: Reporting efficiency is critical in systematic reviews (SR) as it ensures that readers may understand the quality of the study and thus, the strengths and limitations of the review. Many studies have already assessed reporting quality of SRs in other medical fields. Recently a study assessed reporting quality of SRs including meta-analysis in Endodontology as well. The common conclusion of these studies was that although crucial, reporting quality remains suboptimal in published SRs. The objective of our study, was to explore reporting of SRs in the field of Endodontology.

Materials and methods: Two reviewers (AF, VB) independently hand-searched the issues of three endodontic journals and Cochrane Database of Systematic Reviews from January 2000 to August 2017. Screening of potential eligible studies was carried out independently by the two reviewers. Data extraction was performed independently as well, by the same two reviewers using pre-determined data extraction sheets, the items of which were drawn from PRISMA statement. Subgroup analysis was performed between studies following certain reporting guidelines (e.g. PRISMA, QUOROM etc.) and those that did not.

Results: One hundred records met the inclusion criteria and were included for analysis. The most common countries of origin of the corresponding authors were USA, China and Brazil. Prospective protocol registration lacks in the vast majority of them. Deficiencies were identified with regard to eligibility criteria, screening, data extraction and quality assessment procedures as well as meta-analysis model and heterogeneity assessment. Subgroup analysis revealed more efficient reporting in certain domains (15/27) in SRs following established reporting guidelines.

Conclusions: In accordance with previous studies, SRs in endodontology exhibited significant reporting limitations. Critical steps need to be taken for advancement of reporting quality. Adherence to reporting guidelines, may improve reporting process and hence, reporting quality *per se.*
Introduction: As an attempt to answer a clinical question, systematic reviews (SR) have gained ground in the medical field (Moher et al. 2007). It is not only clinicians, those, primarily interested in, but researchers and guideline panels too (Moher et al. 2009). Review authors, try to collate all pertinent available evidence, that fits pre-defined eligibility criteria, to answer a research question (Green et al. 2011). The quality of the review, however, is highly dependent on the methods used, to minimize the bias, thus requiring an explicit framework, that might ensure reliable results and conclusions (Green et al. 2011).

Systematic reviews were previously placed on top of the evidence-based pyramid; yet this perception has changed, as they are compared to a spectrum, which enables critical appraisal of the literature (Murad et al. 2016). Their value is not by definition optimal; a SR is as good as the primary studies included (Moher et al. 1998) and the methodology employed (Jadad et al. 2000). But neither the quality of a primary study is strictly reliant on its design per se (Murad et al. 2016). While a randomized controlled trial (RCT) can be rated down for inadequate allocation concealment or deficient random sequence generation, an observational study may be rated up for a very large effect or a dose-response gradient present (Balshem et al. 2011; Guyatt et al. 2011). An outcome-centered approach that takes into consideration additional issues (e.g., indirectness, publication bias etc.) that may affect the strength of the evidence, has been developed by the GRADE group (Guyatt et al. 2008).

The methodology of the SR, on the other hand, concerns every step undertaken, from formulating the research question to selecting the primary studies, data synthesis and drawing conclusions. For the clinician or the guideline panel that will read a given SR, it is essential to confidently trust the findings and the estimate of a subsequent meta-analysis. An estimation of the design and implementation of a SR is primarily reflected in its report (Moher et al. 2007). Extrapolating the results of a study of RCTs (Soares et al. 2004), incomplete reporting of a SR, does not necessarily imply a deficient methodology. Conversely, to ascertain the high quality of a SR, the report of the study needs to adhere to reporting standards (Moher et al. 2009). A plethora of guidelines is available in Equator-Network (http://www.equator-network.org/) , summarizing the items needed to be included in the report. Similar initiatives, including study
registration or post-publication commentary, have been taken in attempt to promote clarity and thoroughness of the reports (Glasziou et al. 2014).

In a recent cross-sectional study of SRs in the biomedical field (Page et al. 2016), adherence to guidelines proved to be suboptimal, yet slightly improved during the last decade (Moher et al. 2007). Barely one third of the studies (29%) reported the review according to established guidelines, while even less (27%) reported working from a protocol or published the protocol (4%) (Page et al. 2016). In a considerable number of studies, information regarding the screening process (29%), data extraction (33%) and risk of bias (37%), were missing (Page et al. 2016). Last but not least, the Cochrane status and reporting of PRISMA guidelines, were linked with a better reporting performance of the 26 reporting items checked (Page et al. 2016).

The PRISMA guidelines, however, do not only concern the full-text of the review. On the contrary, extensions of the guidelines have been published for protocol, abstracts and different types of SRs (Beller et al. 2013; Moher et al. 2015) as it is of great significance, that critical features of the SR are incorporated. Dissemination of these guidelines needs to reach a larger extent, as it seems from the aforementioned study (Page et al. 2016) where a substantial number of studies did not include the terms “systematic review” or/and “meta-analysis” (15%), the number of the participants (50%) and the risk of bias (40%) in their abstracts.

As far as the dental field of endodontics is concerned, studies were conducted the last decade (Suebnukarn et al. 2010; Kattan et al. 2018), assessing the methodological quality of meta-analyses, came in agreement with older studies both in medicine (Jadad et al. 2000; Jadad et al. 1998), revealing various flaws. Recently another study in endodontics (Nagendrababu et al. 2018) assessing the methodological quality of meta-analyses in their inseparable relationship to reporting quality, concluded too, that it is deficient and that measures have to be taken. A gap is therefore revealed in the literature, as to our knowledge, there is not an assessment of reporting quality of systematic reviews in general (not only meta-analyses).

The purpose of our study, was to examine SRs in endodontics, both from an epidemiologic and a reporting point of view. We intended to study in depth their
reporting quality in a similar method that Page et al. (Page et al. 2016) did, incorporating additional items.

Methods:

A protocol of the study, was drafted in advance (21/6/2017), but it is not publicly available, as it is not published. Minor revisions took place, during the process of the study.

Eligibility criteria: A record to be considered as SR, needs to meet roughly the definition of PRISMA-P (Moher et al. 2015): A data synthesis of all available evidence, that fit pre-specified eligibility criteria to answer a specific research question and the use of explicit, systematic methods to minimize bias in the identification, selection, synthesis and summary of the studies. Three out of the six following procedures needs to be explicitly stated in the report to be included: 1) Inclusion criteria, 2) Search strategy, 3) Screening, 4) Data extraction, 5) Quality assessment, 6) Data Synthesis. Our study focuses on reporting quality of SRs, thus, having the presence of all six of them as a prerequisite for inclusion would minimize the external validity, limiting the sample only to the best-reported studies. Additionally, studies conducted by only one author, or explicitly reported screening or data extraction by only one reviewer without verification by a second reviewer were excluded. Likewise, when information sources included only one electronic database (excluding trial registries or grey literature databases), the study was excluded as well. The rationale behind these eligibility criteria, derives from a recent commentary (Puljak 2017) about the conflicting topic of SRs definition. According to the article, at least two electronic databases have to be searched, and the screening and data extraction procedure has to be undertaken optimally by two reviewers independently. Lastly, a study to be included, needs to include at least one eligible study, so that the reporting quality of the record can be assessed in full extend.

Information sources: Three major endodontic peer reviewed journals (Australian Endodontic Journal, International Endodontic Journal, Journal of Endodontics) and Cochrane Database of Systematic Reviews were hand-searched.
Search strategy: All issues from January 2000 up to August 2017, of the three journals mentioned above and Cochrane database were hand-searched.

Screening: Two independent reviewers, Athanasios Fasoulas (AF) and Vasileios Bousias (VB), initially searched all titles and abstracts of the papers, included in these issues. All records meeting the definition of a knowledge synthesis from Kastner et al. (Kastner et al. 2012), were initially included. Liberal acceleration was used, meaning at least one reviewer is needed for a record to be included, while at least two reviewers are required to exclude it. Subsequently, two independent reviewers screened the full texts of initially included records, against the eligibility criteria. In case of disagreement, a third reviewer, Anna-Bettina Haidich (ABH), resolved the disagreement arbitrating. As far as the Cochrane Database is concerned, the expertise of a fourth reviewer (Theodoros Lambrianidis, TL) in the field of endodontics, was deployed for the inclusion of relevant articles. In cases of disagreement as a result of dental specialties overlap in the topic of the study, TL arbitrated and decided about eligibility. If a study was updated, the latest update was chosen for inclusion. The interobserver agreement during the selection of full-texts was measured using the Gwet’s AC1 measure (Gwet 2008).

Data management: All titles and abstracts initially included during the first phase of screening, were imported in Excel 2016 (Microsoft Office, Microsoft 2015). Subsequently, the studies included for extraction and analysis were imported in a second sheet, including all data items to be extracted from each record.

Data extraction: Predetermined sheets, were used for extraction of the data from the selected studies. The majority of the data items, were identical to the respective items of Page et al. (Page et al. 2016), while minor modifications were made taking into consideration the sample we had. More specifically, epidemiologic characteristics of the studies included studies (number of authors, country of corresponding author, name of journal published, year of publication, primary focus/topic and number of outcomes in the studies) were recorded. Identification of study as systematic review and/or meta-analysis was measured both in abstracts and full texts. Information on funding of the studies and conflict of interest of the authors were recorded as well. Protocol registration and relevant information (database, registration number, presence of hyperlink and manuscript section where registration was reported) were extracted for
individual studies. Information was recorded with regard to reporting guidelines followed and the purpose of using them was reported. Furthermore, Cochrane methodology, use of PICO framework in research question, eligibility of primary studies (study design, publication status, year of publication and language of publication) and information sources (number, names, period of coverage and search strategy of electronic databases, trial registries, additional methods for study identification) were registered. Methods for screening, measurement of interobserver’s agreement, data extraction, risk of bias (assessment of selective outcome reporting and conflict of interest for individual studies were extracted segregated) were recorded as well. Beside the latter, information concerning contact with authors for additional information, number of participants included, measures of diagnostic accuracy, data synthesis (significance of the first reported result, model employed, number of studies included, forest plot, incorporation of risk of bias) and heterogeneity assessment were extracted. Last but not least, publication bias assessment, additional analysis (e.g. subgroup analysis, meta-regression, harm events analysis, cost-effectiveness analysis), GRADE assessment and limitations (both in primary studies and SR level) incorporation in abstract and manuscript were recorded too. Two independent reviewers, AF and VB, independently and in duplicate, extracted the data from all studies. In case of disagreement, a third reviewer (ABH), helped as a referee and consensus was reached after discussion.

**Data analysis:** All analyses were performed using R version 3.3.2. Descriptive statistics were used for summarizing the reporting characteristics of SRs, using frequency and percentage for categorical data and median and interquartile range (IQR) for continuous. The presence and quality of reporting of data items of PRISMA checklist (Moher et al. 2009) were assessed. A post-hoc quantification of the reporting quality was assigned for every record. If the variable was binary (presence/absence, e.g. language eligibility criteria, interrater agreement) then it was treated as such (0,1). In case that an item could take three values (also an intermediate, e.g. screening procedure might be reported but without clearly stating whether it was performed independently, the number of reviewers screened the records, if it was performed both in full-texts and title/abstracts) the quantification was adjusted respectively (0,1,2). Studies in which data synthesis was performed but not in the established methods of meta-analysis (e.g. cochrane methodology, use of PICO framework in research question, eligibility of primary studies (study design, publication status, year of publication and language of publication) and information sources (number, names, period of coverage and search strategy of electronic databases, trial registries, additional methods for study identification) were registered. Methods for screening, measurement of interobserver’s agreement, data extraction, risk of bias (assessment of selective outcome reporting and conflict of interest for individual studies were extracted segregated) were recorded as well. Beside the latter, information concerning contact with authors for additional information, number of participants included, measures of diagnostic accuracy, data synthesis (significance of the first reported result, model employed, number of studies included, forest plot, incorporation of risk of bias) and heterogeneity assessment were extracted. Last but not least, publication bias assessment, additional analysis (e.g. subgroup analysis, meta-regression, harm events analysis, cost-effectiveness analysis), GRADE assessment and limitations (both in primary studies and SR level) incorporation in abstract and manuscript were recorded too. Two independent reviewers, AF and VB, independently and in duplicate, extracted the data from all studies. In case of disagreement, a third reviewer (ABH), helped as a referee and consensus was reached after discussion.
only pooled estimate and not weighted, models that are not accepted), were not considered as meta-analysis.

We examined whether studies following certain reporting guidelines (e.g. PRISMA, QUOROM etc) exhibited better reporting performance than counterparts. A subgroup analysis was performed between the group of SRs following reporting guidelines (including the Cochrane reviews, as they rarely follow additional reporting guidelines but reporting follows the Cochrane Handbook for Systematic Reviews of Interventions (Higgins & Green 2011). The level of statistical significance was set at $\alpha=0.05$.

**Results:**

**Search results:** During the first phase of screening, 127 titles were identified and included for further consideration. After full text evaluation twenty-seven records did not meet the inclusion criteria, thus, were excluded. The Gwet’s AC1 was calculated: 0.78 (95%CI: 0.67-0.88, standard error = 0.052) Finally, the remaining 100 SRs were included in our study. The full text of the manuscripts, including the appendix, where provided, were scrutinized and predetermined data items were extracted from each review.

**Epidemiological characteristics:** From the sample of our 100 studies, an 8% of them were Cochrane reviews, while the remaining 92% were published in peer reviewed journals in the field of endodontics. More specifically, almost a third (29%) was published in International Endodontic Journal (IEJ), the majority (61%) in Journal of Endodontics (JOE) and a tiny amount (2%) of the records in Australian Endodontic Journal (AEJ). These three journals have the highest impact factor among the journals in endodontics.

An increase in the number of SRs produced each year was revealed from 2001 up to 2017 (Figure 1). The number of authors ranged from 2-12 with a median of four authors (IQR 3-6). The most predominant country of corresponding author was USA (25/100) with Brazil (12/100) and China (10/100) in the second and third place respectively.
More than half of the studies (54%) addressed a treatment question. The remaining studies concerned epidemiologic (18%), prognosis (12%), diagnosis (10%), mixed (4%) or other (3%) topic. Studies that reported the source of funding were a minority (23%). All 23 studies reported a non-profit sponsor. The total number of participants included in the study was reported only in 30% of the reviews while it is worth underlining that in most cases (19%, 19/100) the unit of measure was not patients but instead the number of teeth or restorations of each patient (units potentially correlated). The median number of patients included in the studies was 206 (IQR 50-900). Quantitative synthesis was performed in almost half of the studies (47%) with a median of 9 (IQR 5-14) studies included in the largest meta-analysis; yet in three of them, unknown (unreferenced) or inappropriate methods were applied. An analysis of harms was reported only in 17% of the studies and in another 3 studies this pre-specified
analysis was not performed due to missing data. An even lower percentage of studies reported the intention to perform cost-effectiveness analysis (4%) while in only one it was possible to be performed due to lack of eligible studies. The most important characteristics are summarized in Table 1.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Category</th>
<th>Number (Percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of journals</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>Year of publication</td>
<td>2001-2005</td>
<td>4 (4%)</td>
</tr>
<tr>
<td></td>
<td>2006-2010</td>
<td>18 (18%)</td>
</tr>
<tr>
<td></td>
<td>2011-2015</td>
<td>45 (45%)</td>
</tr>
<tr>
<td></td>
<td>2016-2017</td>
<td>33 (33%)</td>
</tr>
<tr>
<td>Number of authors</td>
<td>2-3</td>
<td>38 (38%)</td>
</tr>
<tr>
<td></td>
<td>4-6</td>
<td>51 (51%)</td>
</tr>
<tr>
<td></td>
<td>≥7</td>
<td>11 (11%)</td>
</tr>
<tr>
<td>Country of corresponding author</td>
<td>USA</td>
<td>25 (25%)</td>
</tr>
<tr>
<td></td>
<td>Brazil</td>
<td>12 (12%)</td>
</tr>
<tr>
<td></td>
<td>China</td>
<td>10 (10%)</td>
</tr>
<tr>
<td></td>
<td>Israel</td>
<td>7 (7%)</td>
</tr>
<tr>
<td></td>
<td>UK</td>
<td>7 (7%)</td>
</tr>
<tr>
<td></td>
<td>Sweden</td>
<td>6 (6%)</td>
</tr>
<tr>
<td></td>
<td>Other (&lt;5 SRs/country 18</td>
<td>33 (33%)</td>
</tr>
<tr>
<td></td>
<td>countries)</td>
<td></td>
</tr>
<tr>
<td>Focus of review</td>
<td>Treatment</td>
<td>54 (54%)</td>
</tr>
<tr>
<td></td>
<td>Epidemiology</td>
<td>18 (18%)</td>
</tr>
<tr>
<td></td>
<td>Prognosis</td>
<td>12 (12%)</td>
</tr>
<tr>
<td></td>
<td>Diagnosis</td>
<td>10 (10%)</td>
</tr>
<tr>
<td></td>
<td>Other (e.g. mixed, genetics etc.)</td>
<td>6 (6%)</td>
</tr>
<tr>
<td>Journals</td>
<td>Journal of Endodontics</td>
<td>61 (61%)</td>
</tr>
<tr>
<td></td>
<td>International Endodontic Journal</td>
<td>29 (29%)</td>
</tr>
<tr>
<td></td>
<td>Cochrane Database of Systematic Reviews</td>
<td>8 (8%)</td>
</tr>
<tr>
<td>Update of previous SR</td>
<td></td>
<td>5 (5%)</td>
</tr>
<tr>
<td></td>
<td>Australian Endodontic Journal</td>
<td>2 (2%)</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>Yes</td>
<td>44 (44%)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>56 (56%)</td>
</tr>
<tr>
<td>Harms considered</td>
<td></td>
<td>20 (20%)</td>
</tr>
<tr>
<td>Cost-effectiveness considered</td>
<td></td>
<td>4 (4%)</td>
</tr>
</tbody>
</table>

Table 1. Epidemiologic characteristics of studies
Reporting characteristics:

As far as the title and abstract of the SRs, it is noteworthy that 22% and 27% respectively did not identify the study as a SR or meta-analysis. This percentage is slightly elevated if we take into consideration that these two terms are barely included in the title of Cochrane reviews. Among the studies selected only 5% were updates of previously conducted SRs while the rest 95% were identified as originals. Only a small minority (22%) of the studies reported working from a protocol, fifteen out of which published a protocol in advance. The databases that the protocols were published were either PROSPERO (7%) or Cochrane Database (8%). The number of registration was available in all (seven) studies registered in PROSPERO while the Cochrane studies reported the issue of Cochrane journal that the protocol was published; a hyperlink, however, was not available in any of the studies. After omitting the Cochrane SRs that are reported according to the Cochrane Handbook for Systematic Reviews of interventions, it was only 18% of studies (17/92) that cited established guidelines (e.g. PRISMA, MOOSE etc.) for the report of the review. One more study reported using AMSTAR and Oxford systematic review appraisal sheet. Cochrane methods were reported in twenty (22%) non-Cochrane studies either for a specific procedure, or as a general statement citing the Cochrane Handbook for Systematic Reviews of Interventions.

Study eligibility criteria:

In this section, we will describe the reporting efficiency of the studies regarding the eligibility criteria that were defined for the identification of the primary studies. A research question in terms of PICO or PPP-ICP-TR was reported in about half (46/100) of our sample. Even more studies (95%) reported study designs that were eligible for inclusion. Among them only the half (46/95) defined both eligible and ineligible designs, and the rest 42 and 7 defined solely eligible and ineligible designs respectively. A relatively lower percentage (63%) reported the publication status with regard to eligibility for the review, with 27% considering both published and unpublished for inclusion while the rest 36% searched only for published. A wide heterogeneity was revealed in study designs considered for inclusion. Inclusion was restricted exclusively
to randomized controlled trials in a 10% (7% of which were Cochrane reviews) and to experimental clinical designs in general (e.g. RCTs, quasi-RCTs, non-randomized controlled trials e.t.c.) in 17% of the studies. Lastly, the majority (86%) of reviews reported the language of the primary studies as an inclusion criterion. English only-written studies (43%) and no language restriction (34%) were the primary preferences in the reviews that were included. The authors of the remaining studies (9%) reported a combination of English and other languages as an inclusion criterion.

**Search methods:**

Excluding registries or grey literature databases, a median of four electronic databases (IQR 3-5) were searched for primary studies identification. Years of coverage for the databases was reported in 79% of the studies but only in half of them (41%) both start and end dates were reported for all databases while in the rest (38%) it was reported partially stating either start or end dates for one or more databases. Search strategy was available in 92% of reviews but only in 70% a full Boolean strategy was reported; in the remaining 22% either free text or MeSH terms were provided. Study registries were searched only in 12% of the studies (8/12 were the Cochrane studies) with Clinicaltrials.gov (11/12) and World Health Organization International Clinical Trials Registry Platform (WHO ICTRP) (8/12) being the most common. Additionally, in most trials (77%), the references of the primary studies were searched for extra records. Alternative sources of information were reviewing reference lists of relevant reviews, studies or textbooks (51%), handsearching of particular journal (37%), searching for grey literature and reviewing abstracts/proceedings of specific conferences (9%). In case of missing data or ambiguity, it was reported in 21% of the studies that efforts to contact the authors of the primary studies were made while in one study, it was explicitly stated that no effort was made. In the rest of the reviews (78%) no relevant statement was made.

**Study selection, Data extraction and Risk of Bias:**

The screening method was not reported in 19 (19%) studies at all. All titles and abstracts, and full-texts were evaluated by two (17%) or more (4%) reviewers independently in 21% of the SRs. A different method was used between titles/abstracts and full-texts evaluation in 2 studies (2%). In the remaining records (58%), study
selections was incompletely reported either not reporting whether it was performed independently or not reporting both titles/abstracts and full-text phases. Agreement between the reviewers was evaluated and reported in 16% of the studies. Cohen’s $\kappa$ was the most common statistic used (12%) while Gwet’s AC1 (2%) and descriptive percentage agreement (2%) were the alternatives.

Data extraction was reported adequately in over a third of the studies (36%). In 24 cases the procedure was performed by two reviewers independently extracting data from all studies. Extraction by one reviewer and verification by a second (4%) or independently by more than 2 reviewers (7%) were the alternatives. In 5 reviews, data extraction was reported insufficiently either not reporting the number of reviewers that extracted data and/or whether it was performed independently and/or in all records. In over half of the studies (59%), however, the data extraction phase was not reported at all.

Assessment of the primary studies’ quality was reported in 66 studies. The method of implementation, though, was only reported in 36/66 reviews. Independent risk of bias assessment by two reviewers, applied in all studies, was reported in a quarter (23%) while in 4% it was implemented independently by three reviewers. Assessment only by one reviewer and verification by a second was performed in 2 studies. In the remaining 7 studies reporting of RoB assessment was incomplete not mentioning the number of reviewers or/and whether it was performed independently and/or in all records. As far as the tool used is concerned, it was reported in the majority of studies (63/66). The most common tool (15/66) used was the Cochrane Risk of Bias tool or a modification. Newcastle-Ottawa (3/66) and Jadad scale (2/66) were used for assessment of observational studies while Quadas (5/66) and Quadas-2 (2/66) were used for assessment in DTA systematic reviews. Reporting guidelines (e.g. CONSORT, STROBE etc.) (8/66), self-developed tools (4/66) or other tools for RoB (e.g. ARRIVE modification, SORT, CASP etc.) (26/66) were the alternatives. Selective outcome reporting assessment was reported only in 9 reviews.

*Reporting of results (Records included/excluded, participants):*

Review flow was reported in 95% of the cases, either only in text/table (32%), only in a PRISMA/QUOROM-like flowchart (4%) or as a combination of these two (59%). In a similar way, studies that were excluded together with the reason for exclusion, were
reported only in text/table (50%), only in a PRISMA/QUOROM-like flowchart (2%), in a combination of these two (2%) or partially (either selectively reporting reasons of exclusion of some studies or aggregating studies with reasons for exclusion) (8%). In the rest 38 SRs, no reason for exclusion was reported for specific studies. The total number of participants was reported in 27% in the main text and in 22% in the abstract.

**Reporting of study outcomes:**

The outcome of the studies was reported in the methods section most of the studies (95%), mostly vaguely though. Mainly, the outcome of the studies was reported by stating the objective of the study (e.g. the purpose of our study was to compare the effectiveness of intervention A vs comparator B etc.) instead of clearly identifying the outcomes of interest. In the results section of the reviews, a median of 2 outcomes was reported (IQR 1-4). Outcomes were identified as primary/secondary in 56 studies. Among them, there were 43 reviews that did not specify which outcome was the primary, as they included and reported only one. Seven were SRs of diagnostic test accuracy, thus specification of primary and secondary outcome was not reported. No identification of primary/secondary outcome was revealed in 37 studies. The most common types of primary outcomes were binary (44%), continuous (23%) or categorical (22%). In almost half of the studies (20/44), a comparator was not considered while the results of the rest were either favourable non-statistically significant (12/44) or favourable significant (7/44). Only 4 cases of unfavourable (1 statistically significant and 3 not significant) and one case of equipoise were reported.

**Meta-analyses:**

Meta-analysis was performed -properly- in 44 cases. The effect model was reported in 40 of them. Random effects model was more often used (21/44) in comparison to fixed effects model (12/44). The use of both of them in the same review was reported seven times. In a case that was self-identified as meta-analysis (in our study it was not treated as such), a model was reported that do not correspond to the standard models of meta-analysis neither was it accompanied by a pertinent reference. A median of nine studies (IQR 5-13.5) were synthesized in the largest meta-analysis as reported in 43/44 of the cases. A forest plot was used in 35 of the meta-analyses. Regardless of whether quantitative synthesis was performed, a formal analysis of heterogeneity was reported.
in 40 studies while in 27 other studies, it was taken into account qualitatively. The methods mostly used was chi-squared test (or Cochran’s Q) (27/40) and in over half of them the inconsistency index ($I^2$) was reported (24/40). Other common methods such tau ($\tau^2$) and L’abbe plot (3/40 and 5/40, respectively) were used as well. In sixteen reviews (16%), the meta-analytic model, was selected based on a cutoff value of heterogeneity (e.g. if heterogeneity was less than 50% a fixed-effect model was chosen, if it was larger than 50%, random-effect model was the model of choice). The intention to assess publication bias was reported in 41% of the studies (18/44) but it was eventually evaluated only in ten reviews as a result of low number of studies. Funnel plot was the most used method (9/10). In one case, the combination of funnel plot with Egger’s test was employed and in another the Orwin’s method (Orwin 1983) for publication bias assessment. A discussion and a possible correlation of the SR’s results to potential publication bias was reported in sixteen studies (16%). Additional analyses were reported in twenty studies (20%). Meta-regression was performed in only eight and subgroup analysis in fourteen reviews. In one of the studies (Panitvisai et al 2010), however, subgroup analysis was improperly conducted as data were correlated yet treated as independent.

**GRADE and conclusions**

The strength of evidence was evaluated with the GRADE approach in only a few studies (10%) as reported. In 4/10 the GRADE assessment was reported in the text whilst a summary of evidence table was available in the remaining 6/10. In a significantly higher number of studies (70/100), limitations were reported as part of discussion or conclusion sections. Shortcomings were mostly attributed to the primary studies (56/70), and in minority of them to the review (4/70) or a combination of both of them (10/70). Limitations were incorporated in results or conclusion of the abstract in fewer studies (36/100).
Conflict of interest

Disclosure of the SR’s authors was included in 69% of the studies. Conflict of interest of the primary studies, though, was investigated only in nine cases as reported.

Subgroup analysis:

Subgroup analysis was performed between studies that cited reporting guidelines (e.g. PRISMA, QUOROM) and Cochrane studies as well (they necessarily adhere to the reporting norms of Cochrane handbook of systematic reviews) from the one side, and studies not cited any guidelines for the report of their study from the other side. Twenty-five studies were included in the first group and seventy-five in the second.

Significant differences were detected in fifteen out of 27 items. These items were more frequently reported in the guidelines group. More detailed, the results of the analysis are displayed in Table 2:

<table>
<thead>
<tr>
<th>Categories</th>
<th>Characteristic</th>
<th>Levels</th>
<th>Subgroups</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>No guidelines (75)</td>
<td>Guidelines (25)</td>
</tr>
<tr>
<td>Administrative information</td>
<td>Systematic review/ Meta-analysis used in title/abstract</td>
<td>Yes</td>
<td>66 (88%)</td>
<td>22 (88%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>9 (12%)</td>
<td>3 (12%)</td>
</tr>
<tr>
<td>Protocol of the systematic review</td>
<td>Registered/Published</td>
<td>Yes but neither registered nor published</td>
<td>2 (3%)</td>
<td>13 (52%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes but neither registered nor published</td>
<td>6 (8%)</td>
<td>1 (4%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>67 (89%)</td>
<td>11 (44%)</td>
</tr>
<tr>
<td>Eligibility criteria</td>
<td>PICO framework</td>
<td>Primary outcome reported</td>
<td>Study design</td>
<td>Publications status</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>--------------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>27 (36%)</td>
<td>19 (76%)</td>
<td>p&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>48 (64%)</td>
<td>6 (24%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>42 (56%)</td>
<td>14 (56%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>28 (44%)</td>
<td>9 (44%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>72 (96%)</td>
<td>22 (88%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>3 (4%)</td>
<td>3 (12%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>47 (63%)</td>
<td>16 (64%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>28 (37%)</td>
<td>9 (36%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>64 (85%)</td>
<td>22 (88%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>11 (15%)</td>
<td>3 (12%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>27 (36%)</td>
<td>14 (56%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Partially</td>
<td>35 (47%)</td>
<td>3 (12%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>13 (17%)</td>
<td>8 (32%)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study identification, selection, data extraction and quality assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search strategy</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Screening method</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Reasons for exclusion</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Review flow</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Inter-observer's agreement</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Category</td>
</tr>
<tr>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>Data extraction method</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Contact with the authors of primary studies</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Number of participants included in the main text</td>
</tr>
<tr>
<td>Number of participants included in the abstract</td>
</tr>
<tr>
<td>Risk of Bias method</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Risk of Bias tool</td>
</tr>
<tr>
<td>Selective reporting</td>
</tr>
<tr>
<td>Conflict of interest in primary studies</td>
</tr>
<tr>
<td>Metanalysis Metanalysis model *</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Table 2. Summary of subgroup analysis.

*The number of studies included meta-analysis was 33 and eleven in “No guidelines” and “Guidelines” group respectively.

**Discussion:**

Paraphrasing (Moher et al. 2009), the clarity of reporting reflects what was done and found after conducting a SR. In this study, the reporting quality of SRs in Endodontology was assessed. Regardless whether a study was identified as SR or not, studies including specific criteria, were included. It is a vicious circle trying to evaluate reporting on the one side and including actual SRs on the other side as the first informs us about the methodology of the second. Eventually, a balance between them was achieved requiring three out of six procedures described in the record. A consideration regarding studies eligibility arose with regard to methods pre-specification. It is not,
however, within the limits of the present review to investigate methodologic accuracy but rather study the reporting efficacy of systematic reviews.

Identification and inclusion of all (or nearly all) pertinent studies is a critical step of SR, that ensures reduction of random error in meta-analysis and minimization of selection bias (Clarke & Stewart 1994). Therefore, electronic search in minimum two electronic databases (Gehanno et al. 1998) and screening of records by at least two reviewers independently (Edwards et al. 2002) may improve the quality of the study. Data extraction procedure, might have an impact on study findings as well. Trying to minimize bias, optimally is carried out by at least two reviewers independently, since as (Buscemi et al. 2006) found, errors and omissions during the procedure are thus reduced.

In our study, a recent editorial (Puljak 2017) with regard to search strategy, screening and data extraction had an impact on primary studies eligibility. Reviews authored by only one reviewer were a priori excluded as both screening and data extraction could not be performed in duplicate. Likewise, if electronic search was restricted in only one database (excluding trial registries or grey literature databases) the study was not considered a SR. In cases, however, that one reviewer carried out screening and/or data extraction while a second verified the results, the study was included, despite the potential impact on quality. Accordingly, the authors of the present study are aware of the potential selection bias introduced, as only the highest quality cluster -from a methodological point of view- was included.

SRs of the last seventeen years were identified from three major endodontic journals and Cochrane database. The vast majority of the studies was published either in JOE or in IEJ while almost in half of the studies the corresponding author was either from USA, Brazil or China while production of SRs was increased since 2007 but steeper since 2010.

The findings of the present study as far as reporting quality is concerned are in agreement with previous studies (Moher et al. 2007; Page et al. 2016), as overall, reporting quality of SRs proved to be suboptimal. A considerable number of studies were not self-identified as SR or Meta-analysis, labeled narrative or literature review. All these studies, though, employed systematic methods during review implementation.
Primary focus of more than half of the studies (54%) was on treatment. Only a quarter of the studies (23%) reported the funding of the SR. All 23 studies was funded by non-profit sponsors, mainly university or state grants. On the contrary, 69% of the studies reported conflict of interest of the authors. As far as conflict of interest of the primary studies is concerned though, it was assessed only in nine studies, eight out of which were the Cochrane studies included. The importance of conflict of interest has been recently highlighted (Jørgensen et al. 2018) as it may impact both the primary studies and the evidence synthesis per se, and thus affect the effect estimate. It is already some years, that all three journals included in the study have incorporated in author guidelines disclosure statement as a requirement for publication. Cochrane collaboration has a standard policy, free from conflict of interest related to industry or manufacturers.

The PICO framework is yet not widely accepted (46%) in defining the research question. Formulating the research question is deemed as the first most important step in a SR. It will guide all procedures of the review, thus “the acronym of PICO helps to serve as a reminder” of all critical components of the question (Higgins & Green 2011). Solely randomized controlled trials or interventional studies in general were considered only in a minority of the trials (17%), eight out of which were the Cochrane studies that typically restrict the review in these types of studies. A considerable number of studies (86%) reported eligibility with regard to the language of the paper, yet only 34% of the whole sample, stated no language restrictions. Mainly English (43%) or a combination of English and other languages (9%) were considered eligible introducing potential language bias in the studies; statistically significant findings are predominantly published in English instead of other languages (Egger et al. 1997).

According to the inclusion criteria, at least two major electronic databases were searched in the studies. A number of four databases were at median searched with the majority of the studies (79%) reporting the years coverage, though partially in half (38%) of them. Search strategy of at least one database was reported in most of the studies (92%), but some of them (22%) did not cite the whole strategy; MeSH or free text terms were reported instead. Among the four data sources, three endodontic journals and Cochrane collaboration, only the latter requires detailed citation of search strategy of all databases searched in the reviews, as transparency and repetition are highly valued. Contrary, only 12% searched for unpublished data in trials registries and
another 9% in conference proceedings or grey literature databases. Either reporting or actual methodology of the reviews requires improvement, as considerable number of studies may be identified in these sources only (Mallet et al. 2002; Sterne et al. 2011), and due to publication bias distort the effect estimate (Hopewell et al. 2007). Further reporting/methodological development is necessary in cases of missing data or reporting ambiguity of the primary studies; only 21% of the studies reported the attempt to contact authors, one study reported that they were not contacted while for the remaining studies no information are given. Occasionally what was actually done differs from what was reported (Devereaux et al. 2004) and omission of outcomes with statistically non-significant findings is common as well (Chan & Altman 2005). In order to minimize reporting bias (Furukawa et al. 2007), reviewers should strive to collect further information from authors and report it as well (Mullan et al. 2009).

Screening of the records was reported in the majority of SRs (81%). An incomplete specification of the method used was revealed in most of them (58%) as they omitted reporting important components of the procedure (independent screening, whether it was performed accordingly in both titles/abstracts and full records). Agreement of reviewers in primary studies inclusion was stated in 16% of the records with Cohen’s $\kappa$ coefficient (12%) predominantly most frequent. Augmentation of screening and inter-observer’s agreement is perceived crucial as study selection may impact on precision and validity of the SR’s findings (Edwards et al. 2002). For the purpose of more accurately assessed inter-rater’s agreement, integration of alternative statistics, such as Gwet’s AC1 (used only in 2/16 studies) should be considered (Wongpakaran et al. 2013).

Data extraction and quality assessment implementation on the other hand, were deficiently reported. In the majority of the studies (59%), information with regard to implementation method of data extraction were missing. It is noteworthy that among the 36% of the studies that adequately reported the method applied, in four studies (4%) data extraction was carried out only by one reviewer and confirmed by a second. Although study selection of the present study was guided by (Puljak 2017), and thus the aforementioned studies were also considered, future reviewers and readers need to bear in mind that data extraction by at least two independent reviewers is deemed superior compared to single data extraction (Buscemi et al. 2006). As far as risk of bias
assessment is concerned, 34% of the studies did not report evaluating the quality of the studies at all. From the remaining 66% that reported assessing the quality of primary studies, another 30% did not report the method of implementation. Although in our study, we mainly focused on the number of reviewers that performed risk of bias assessment and whether it was performed independently or not, quality of the assessment may be influenced by other factors as well (blinding of reviewers, methodological training and experience etc.) (Jadad et al. 1996; Kjaergard et al. 2001). Beside the method of implementation, of great importance is the tool for quality assessment. It was reported in 63/66 studies together with a great variety of tools used. The most used tool (15/66) was Cochrane tool for risk of bias. Mistakenly reporting guidelines and checklists were used for assessing quality as well as scales and quantitative self-developed tools that are not suggested for that purpose (Moher et al. 1995; Higgins & Green 2011). The assessment, however, was incorporated (using subgroup analysis, meta-regression or sensitivity analysis) in only five (5/44) studies included meta-analysis, according to what was reported. The risk of bias and other limitations were included in the abstract in 36% of the studies. In 70% of the studies, limitation related both to primary studies and the conduct of the SR were reported in the discussion/conclusion section of the main text. It is worth noting, that only in 14/70 studies, possible sources of bias at the review level were noted, that is highly unlikely. The GRADE approach was applied in 10% of the studies.

The review flow was reported in the majority (95%) of the studies. Reporting of the reasons of exclusion though was deficient (62%) as well as the total number of participants included (27% in main text and 22% in abstracts). It was not only the reporting of the number of participants though that was problematic. In nineteen (19/30) of the studies that the number of participants was identified, the unit of measurement was not patients but other units (e.g. teeth, restorations, implants etc.). This might have led to violation of independence of observations, a common pitfall in clinical research that is rarely taken into account properly (Fleming et al. 2013). Statistical problems, however, were not restricted only to the latter. Self-developed meta-analysis models (Peterson & Gutmann 2001), deficient statistical analysis such as pooling all observations without weighting them (De Pablo et al. 2010; Abella et al. 2012; Ahmad & Alenezi 2016; Pak et al. 2012) or improper statistical test choice (Tsesis et al. 2013; Tsesis et al. 2009; Tsesis et al. 2015) were the most frequent limitations of the studies.
Outcomes of the studies, were specified in the big majority of the method’s section of studies (95%). The most important drawback, though, was that in many cases (37%) the primary outcome(s) was not identified in contrast to secondary outcomes. The group of the 56 studies that did identify them on the other hand, was skewed as in most of them (44/56) only one outcome was considered, thus it was not that reporting was per se superior. Patient-important outcomes such as harm or adverse events (20%) or cost-effectiveness (4%) were infrequently considered.

Among the 44 (44%) studies that meta-analysis was properly performed, most of them (40/44) specified the model employed. In 35% of meta-analysis, a forest plot was used for visualization of the analysis and effect estimate. A former assessment of heterogeneity was performed in 40 (40%) studies and the most used methods where chi-squared (or Cochran’s Q) (27/40) and I-squared test (24/40). In sixteen of them (16/40), the magnitude of heterogeneity, determined the effect model used in meta-analysis. Our findings are in agreement with (Page et al. 2018) despite the fact that this approach is discouraged (Higgins & Green 2011). Researchers might opt predefining the effect model choice on the basis of effect expected or alternatively employ both methods and present results accordingly (IOM 2011; Higgins & Green 2011). Statistical variance was considered qualitatively as well in another 27 (27%) studies. Assessment of publication bias, was performed in ten studies (10%) while slightly more (18%) reported the intention to do. The most commonly used method (9/10) was a funnel plot. Primarily, it was due to insufficient number of studies that most of the aforementioned studies did not proceed to publication bias evaluation; as a rule of thumb, ten studies need to be at least included for adequate test power (IOM 2011; Higgins & Green 2011). It is worth noting that in two studies (Zhang et al. 2015; Li et al. 2015), publication bias was assessed with a funnel plot while nine and five studies were included respectively. Contrary, a Cochrane review (Nadin et al. 2008) did not proceed with funnel plot, although ten studies were included in one meta-analysis. Within-study selective outcome reporting was stated in nine (9%) studies.

Prospective protocol development was reported in one fifth of the studies, while only fifteen per cent of the studies registered their protocol either in Cochrane Database of Systematic Reviews or in PROSPERO. An even lower number of studies (8/15, all of them Cochrane SRs) published their protocols as well. Publication of the protocol is
deemed not only as a methodological quality indicator of SRs (Sideri et al. 2018) but as a marker of reporting capacity as well (Allers et al. 2018). In our study, studies reported the presence of a protocol were associated with reporting guidelines reporting and thus, with better overall reporting quality. Caution for deficient reporting is required though, as presence of a published protocol does not per se ensure adherence to pre-determined methods; on the contrary actual methods frequently deviate from what was planned (Pandis et al. 2015; Silagy et al. 2002).

Commonly established reporting guidelines were followed in one fourth of the studies. In accordance with (Page et al. 2016), apart from the Cochrane studies, most of them (16/17), mistakenly reported the use of the guidelines for conducting the review while the purpose of them is in the reporting procedure of the study. Only IEJ among the three journals requires reporting of the SR according PRISMA statement. Journal of Endodontics and Australian Endodontic Journal have not yet incorporated specific guidelines in author guidelines. As far as Cochrane collaboration is concerned, reporting of SRs follows Cochrane handbook of Systematic reviews. Certain reporting domains of studies following reporting guidelines seems to be superior than counterparts, in accordance with (Page et al. 2016) as it was showed in the subgroup analysis performed.

**Strengths and limitations:** Several strengths characterize our study. Study identification was performed in duplicate and independently applying hand-search. Study selection and data extraction was performed in duplicate and independently as well. We did not include only studies self-identified as SRs but eligibility was assessed based on presence of at least 3/6 methods cited in PRISMA-P.

Eligibility, however, was also determined according to (Puljak 2017), thus including only the best cluster of SRs. Only the latest edition of the Cochrane reviews were included, that might further improved the quality of the studies. Moreover, the three endodontic journals considered, were them with the highest impact factor. Last but not least, our findings and analysis were based on what reported and not what was done, thus methodological conclusions must be cautiously interpreted.
Conclusions:

Information about methodology of a SR, is drawn from the report of the study. Overall reporting quality of SRs in endodontics proved to be suboptimal. Future researchers, might opt to follow certain reporting guidelines such as PRISMA that have been associated with better reporting performance. Endodontic journals need to incorporate reporting guidelines in author guidelines and perhaps increase the word limit for a SR as it might had an impact on reporting quality.

Funding: There was no funding

References:


Allers, K. et al., 2018. Systematic reviews with published protocols compared to those without: more effort, older search. *Journal of Clinical Epidemiology*.


Buscemi, N. et al., 2006. Single data extraction generated more errors than double data extraction in systematic reviews. *Journal of Clinical Epidemiology*.

Journal.


Devereaux, P.J. et al., 2004. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. *Journal of Clinical Epidemiology*.


IOM, (Institute of Medicine), 2011. Finding What Works in Health Care: Standards for Systematic Reviews,


Kastner, M. et al., 2012. What is the most appropriate knowledge synthesis method to conduct a review? Protocol for a scoping review. BMC Medical Research Methodology, 12, p.114.


Mallet, S., Hopewell, S. & Clarke, M., 2002. The use of grey literature in the first 1000 Cochrane reviews. 4th Symposium on Systematic Reviews: Pushing the
Boundaries; 2002 Jul 2-4; Oxford, UK.


Mullan, R.J. et al., 2009. Systematic reviewers commonly contact study authors but do so with limited rigor. *Journal of Clinical Epidemiology*.


Page, M.J. et al., 2018. Flaws in the application and interpretation of statistical analyses in systematic reviews of therapeutic interventions were common: a cross-sectional analysis. *Journal of Clinical Epidemiology*.


Pandis, N. et al., 2015. Discrepancies in outcome reporting exist between protocols and published oral health Cochrane systematic reviews. *PLoS ONE*.


Puljak, L., 2017. If there is only one author or only one database was searched, a study should not be called a systematic review. *Journal of Clinical Epidemiology*, 91, pp.4–5.

Sideri, S., Papageorgiou, S.N. & Eliades, T., 2018. Registration in the international prospective register of systematic reviews (PROSPERO) of systematic review protocols was associated with increased review quality. *Journal of Clinical Epidemiology*.


Sterne, J.A.C. et al., 2011. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ (Online)*.


References of primary studies:


Allers, K. et al., 2018. Systematic reviews with published protocols compared to those without: more effort, older search. *Journal of Clinical Epidemiology*.


Buscemi, N. et al., 2006. Single data extraction generated more errors than double data extraction in systematic reviews. *Journal of Clinical Epidemiology*.

Chan, A.W. & Altman, D.G., 2005. Identifying outcome reporting bias in randomised


Devereaux, P.J. et al., 2004. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. *Journal of Clinical Epidemiology*.


IOM, (Institute of Medicine), 2011. Finding What Works in Health Care: Standards for Systematic Reviews,


Kastner, M. et al., 2012. What is the most appropriate knowledge synthesis method to conduct a review? Protocol for a scoping review. BMC Medical Research Methodology, 12, p.114.


Mallet, S., Hopewell, S. & Clarke, M., 2002. The use of grey literature in the first 1000 Cochrane reviews. 4th Symposium on Systematic Reviews: Pushing the


Mullan, R.J. et al., 2009. Systematic reviewers commonly contact study authors but do so with limited rigor. *Journal of Clinical Epidemiology*.


Page, M.J. et al., 2018. Flaws in the application and interpretation of statistical
analyses in systematic reviews of therapeutic interventions were common: a cross-sectional analysis. *Journal of Clinical Epidemiology*.


Pandis, N. et al., 2015. Discrepancies in outcome reporting exist between protocols and published oral health cochrane systematic reviews. *PLoS ONE*.


Puljak, L., 2017. If there is only one author or only one database was searched, a study should not be called a systematic review. *Journal of Clinical Epidemiology*, 91, pp.4–5.

Sideri, S., Papageorgiou, S.N. & Eliades, T., 2018. Registration in the international prospective register of systematic reviews (PROSPERO) of systematic review protocols was associated with increased review quality. *Journal of Clinical Epidemiology*.


Sterne, J.A.C. et al., 2011. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ (Online)*.


Appendix:

R script:

```r
mor$pri <- factor(mor$pri, levels = c(0,1), labels = c("No guidelines","Guidelines"))
mor$term <- factor(mor$term, levels = c(0,1), labels = c("No","Yes"))
chisq.test(with(mor,table(pri,prot)))$expected
fisher.test(with(mor,table(pri,prot)), alternative = "two.sided")
tb1 <- with(mor,table(pri,term))
chisq.test(tb1)$expected
fisher.test(tb1, alternative = "two.sided")
mor$prot <- factor(mor$prot, levels = c(0,1,2), labels = c("No","Partially","Yes"))
tb2 <- with(mor,table(pri,prot))
chisq.test(tb2)$expected
fisher.test(tb2, alternative = "two.sided")
mor$pico <- factor(mor$pico, levels = c(0,1), labels = c("No","Yes"))
tb3 <- with(mor,table(pri,pico))
chisq.test(tb3)$expected
chisq.test(tb3, correct = F)
mor$design <- factor(mor$design, levels = c(0,1), labels = c("No","Yes"))
tb4 <- with(mor,table(pri,design))
chisq.test(tb4)$expected
fisher.test(tb4, alternative = "two.sided")
```
mor$elpub <- factor(mor$elpub, levels = c(0,1), labels = c("No", "Yes"))
tb5 <- with(mor, table(pri, elpub))
chisq.test(tb5)$expected
chisq.test(tb5, correct = F)

mor$ellan <- factor(mor$ellan, levels = c(0,1), labels = c("No", "Yes"))
tb6 <- with(mor, table(pri,ellan))
chisq.test(tb6)$expected
fisher.test(tb6, alternative = "two.sided")

mor$coda <- factor(mor$coda, levels = c(0,1,2), labels = c("No", "Partially", "Full"))
tb7 <- with(mor, table(pri,coda))
chisq.test(tb7)$expected
chisq.test(tb7, correct = F)

mor$codab <- factor(mor$codab, levels = c(0,1), labels = c("No", "Yes"))
tb8 <- with(mor, table(pri,codab))
chisq.test(tb8)$expected
chisq.test(tb8, correct = F)

mor$strat <- factor(mor$strat, levels = c(0,1,2), labels = c("No", "Partially", "Yes"))
tb9 <- with(mor, table(pri,strat))
chisq.test(tb9)$expected
fisher.test(tb9, alternative = "two.sided")

mor$stratb <- factor(mor$stratb, levels = c(0,1), labels = c("No", "Yes"))
tb9.5 <- with(mor, table(pri,stratb))
chisq.test(tb9.5)$expected
fisher.test(tb9.5, alternative = "two.sided")

mor$scre <- factor(mor$scre, levels = c(0,1,2), labels = c("No", "Partially", "Yes"))
tb10 <- with(mor, table(pri,scre))
chisq.test(tb10)$expected
fisher.test(tb10, alternative = "two.sided")
mor$screb<-factor(mor$screb,levels = c(0,1), labels = c("No","Yes"))
tb11<-with(mor,table(pri,screb))
chisq.test(tb11)$expected
fisher.test(tb11,alternative = "two.sided")
mor$extr<-factor(mor$extr,levels = c(0,1,2), labels = c("No","Partially","Yes")))
tb12<-with(mor,table(pri,extr))
chisq.test(tb12)$expected
fisher.test(tb12, alternative = "two.sided")
mor$extrb<-factor(mor$extrb,levels = c(0,1), labels = c("No","Yes"))
tb13<-with(mor,table(pri,extrb))
chisq.test(tb13)$expected
chisq.test(tb13,correct = F)
mor$agr<-factor(mor$agr,levels = c(0,1),labels = c("No","Yes"))
tb14=with(mor,table(pri,agr))
chisq.test(tb14)$expected
fisher.test(tb14,alternative = "two.sided")
mor$rob<-factor(mor$rob,levels = c(0,1),labels = c("No","Yes"))
tb15=with(mor,table(pri,rob))
chisq.test(tb15)$expected
chisq.test(tb15,correct = F)
mor$robm<-factor(mor$robm,levels = c(0,1,2),labels = c("No","Partially","Yes"))
tb16<-with(mor,table(pri,robm))
chisq.test(tb16)$expected
fisher.test(tb16,alternative = "two.sided")
mor$robmb <- factor(mor$robmb, levels = c(0, 1), labels = c("No", "Yes"))

tb16.5 <- with(mor, table(pri, robmb))

chisq.test(tb16.5)$expected
chisq.test(tb16.5, correct = F)

mor$robt <- factor(mor$robt, levels = c(0, 1), labels = c("No", "Yes"))

tb17 <- with(mor, table(pri, robt))

chisq.test(tb17)$expected
chisq.test(tb17, correct = F)

mor$sere <- factor(mor$sere, levels = c(0, 1), labels = c("No", "Yes"))

tb18 <- with(mor, table(pri, sere))

chisq.test(tb18)$expected

fisher.test(tb18, alternative = "two.sided")

mor$cont <- factor(mor$cont, levels = c(0, 1), labels = c("No", "Yes"))

tb19 <- with(mor, table(pri, cont))

chisq.test(tb19)$expected
chisq.test(tb19, correct = F)

mor$flow <- factor(mor$flow, levels = c(0, 1), labels = c("No", "Yes"))

tb20 <- with(mor, table(pri, flow))

chisq.test(tb20)$expected

fisher.test(tb20, alternative = "two.sided")

mor$exre <- factor(mor$exre, levels = c(0, 1, 2), labels = c("No", "Partially", "Yes"))

tb21 <- with(mor, table(pri, exre))

chisq.test(tb21)$expected

fisher.test(tb21, alternative = "two.sided")

mor$nopar <- factor(mor$nopar, levels = c(0, 1), labels = c("No", "Yes"))

tb22 <- with(mor, table(pri, nopar))
chisq.test(tb22)$expected
chisq.test(tb22,correct = F)
mor$nopara=factor(mor$nopara,levels = c(0,1),labels = c("No","Yes"))
tb23=with(mor,table(pri,nopara))
chisq.test(tb23)$expected
chisq.test(tb23,correct = F)
mor$prio=factor(mor$prio,levels = c(0,1),labels = c("No","Yes"))
tb24=with(mor,table(pri,prio))
oddsratio(tb24)
meta$pri=factor(meta$pri,levels = c(0,1),labels = c("No","Yes"))
meta$forest=factor(meta$forest,levels = c(0,1),labels = c("No","Yes"))
meta$model=factor(meta$model,levels = c(0,1),labels = c("No","Yes"))
tb25=with(meta,table(pri,forest))
tb26=with(meta,table(pri,model))
chisq.test(tb25)$expected
chisq.test(tb26)$expected
fisher.test(tb25,alternative = "two.sided")
fisher.test(tb26,alternative = "two.sided")
mor$limit<-factor(mor$limit,levels = c(0,1,2),labels = c("No","Partially","Yes"))
tb27=with(mor,table(pri,limit))
chisq.test(tb27)$expected
chisq.test(tb27,correct = F)
mor$limita<-factor(mor$limita,levels = c(0,1),labels = c("No","Yes"))
tb28=with(mor,table(pri,limita))
chisq.test(tb28)$expected
chisq.test(tb28,correct = F)
mor$coi=factor(mor$coi,levels = c(0,1),labels=c("No","Yes"))
mor$coip=factor(mor$coip,levels = c(0,1),labels=c("No","Yes"))
tb29=with(mor,table(pri,coi))
tb30=with(mor,table(pri,coip))
chisq.test(tb29)$expected
chisq.test(tb30)$expected
chisq.test(tb29,correct=F)
fisher.test(tb30,alternative = "two.sided")
mor$fund<-factor(mor$fund,levels = c(0,1),labels = c("No","Yes"))
tb31<-with(mor,table(pri,fund))
chisq.test(tb31)$expected
chisq.test(tb31,correct = F)
barplot(with(mor,table(pri,prot)),main="Use of a protocol VS Reporting guidelines",beside=T,col=c("blue","red"),xlab="Protocol development",ylim = c(0,80))
barplot(with(mor,table(pri,coda)),main="Years of coverage VS Reporting guidelines",col=c("blue","red"),xlab="Years of coverage",ylim = c(0,60),beside=T)
legend(x=8,y=45,c("Guidelines","No Guidelines"),cex=1.1,col=c("red","blue"),pch=c(15,15))
barplot(with(mor,table(pri,strat)),main="Search strategy VS Reporting guidelines",col=c("blue","red"),xlab="Search strategy",ylim = c(0,80),beside=T)
barplot(with(mor,table(pri,limit)),main="Limitations VS Reporting guidelines",col=c("blue","red"),xlab="Limitations",ylim = c(0,50),beside=T)
barplot(with(mor,table(pri,scre)),main="Screening method VS Reporting guidelines",col=c("blue","red"),xlab="Screening method",ylim = c(0,60),beside=T)
barplot(with(mor,table(pri,extr)),main="Data extraction method VS Reporting guidelines",col=c("blue","red"),xlab="Data extraction method",ylim = c(0,60),beside=T)
barplot(with(mor,table(pri,robm)),main="Risk of bias method VS Reporting guidelines",col=c("blue","red"),xlab="Risk of bias method",ylim = c(0,60),beside=T)

barplot(with(mor,table(pri,exre)),main="Reasons for exclusion VS Reporting guidelines",col=c("blue","red"),xlab="Reasons for exclusion",ylim = c(0,60),beside=T)

barplot(with(mor,table(pri,term)),main="Terms in abstract and title VS Reporting guidelines",col=c("blue","red"),xlab="Terms in abstract and title",ylim = c(0,100),beside=T)

barplot(with(mor,table(pri,pico)),main="PICO framework VS Reporting guidelines",col=c("blue","red"),xlab="PICO framework",ylim = c(0,60),beside=T)

barplot(with(mor,table(pri,design)),main="Study design eligibility VS Reporting guidelines",col=c("blue","red"),xlab="Study design eligibility",ylim = c(0,100),beside=T)

barplot(with(mor,table(pri,elpub)),main="Publication status VS Reporting guidelines",col=c("blue","red"),xlab="Publication status",ylim = c(0,80),beside=T)

barplot(with(mor,table(pri,ellan)),main="Publication language VS Reporting guidelines",col=c("blue","red"),xlab="Publication language",ylim = c(0,100),beside=T)

barplot(with(mor,table(pri,agr)),main="Interobserver’s agreement VS Reporting guidelines",col=c("blue","red"),xlab="Interobserver’s agreement",ylim = c(0,80),beside=T)

barplot(with(mor,table(pri,robt)),main="Risk of bias tool VS Reporting guidelines",col=c("blue","red"),xlab="Risk of bias tool",ylim = c(0,80),beside=T)

barplot(with(mor,table(pri,sere)),main="Selective reporting VS Reporting guidelines",col=c("blue","red"),xlab="Selective reporting",ylim = c(0,100),beside=T)

barplot(with(mor,table(pri,cont)),main="Contact with authors VS Reporting guidelines",col=c("blue","red"),xlab="Contact with authors",ylim = c(0,80),beside=T)

barplot(with(mor,table(pri,flow)),main="Review flow VS Reporting guidelines",col=c("blue","red"),xlab="Review flow",ylim = c(0,100),beside=T)

barplot(with(mor,table(pri,nopar)),main="Number of participants in main text VS Reporting guidelines",col=c("blue","red"),xlab="Number of participants in main
barplot(with(mor,table(pri,nopara)),main="Number of participants in abstract VS Reporting guidelines",col=c("blue","red"),xlab="Number of participants in abstract",ylim = c(0,80),beside=T)

barplot(with(mor,table(pri,prio)),main="Outcomes prioritization VS Reporting guidelines",col=c("blue","red"),xlab="Outcomes prioritization",ylim = c(0,60),beside=T)

barplot(with(mor,table(pri,limita)),main="Limitations in abstract VS Reporting guidelines",col=c("blue","red"),xlab="Limitations in abstract",ylim = c(0,60),beside=T)

barplot(with(mor,table(pri,fund)),main="Source of funding VS Reporting guidelines",col=c("blue","red"),xlab="Source of funding",ylim = c(0,80),beside=T)

barplot(with(mor,table(pri,coi)),main="Conflict of interest VS Reporting guidelines",col=c("blue","red"),xlab="Conflict of interest",ylim = c(0,80),beside=T)

barplot(with(mor,table(pri,coip)),main="Conflict of interest in primary studies VS Reporting guidelines",col=c("blue","red"),xlab="Conflict of interest in primary studies",ylim = c(0,80),beside=T)

barplot(with(meta,table(pri,model)),main="Meta-analysis model VS Reporting guidelines",col=c("blue","red"),xlab="Meta-analysis model",ylim = c(0,40),beside=T)

barplot(with(meta,table(pri,forest)),main="Forest plot VS Reporting guidelines",col=c("blue","red"),xlab="Forest plot",ylim = c(0,40),beside=T)

Agr$T<-factor(Agr$T)
Agr$V<-factor(Agr$V)
rel::gac(data = Agr,kat=2,weight = c("unweighted"),conf.level = 0.95)