Do systematic reviews and meta-analyses, published in the dental literature, comply with the QUOROM and PRISMA statements?

Dissertation submitted in accordance
with the requirements of the
University of Liverpool for the degree of
Doctorate of Dental Science
by
Noor Mohammed Al-Ramadhan

Supervisors:
Dr. Jayne Harrison
and
Professor Neil Pender

October 2011

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Jayne Harrison for making this project possible. I am very grateful for all the long hours spent meticulously reading and rereading my dissertation on weekdays, weekends and even holidays abroad. I will miss the late night emails back and forth, (barely ten minutes between each one!) with comments, corrections, advice and answers to all my questions.

I also owe a great debt to my wonderful family for all their support in the last three years, without which I would I not be where I am today. Nothing calmed me more than my dad's long phone conversations, giving me advice based on his PhD experience in the US. I wouldn't have been able to finish my write-up so quickly either without my mother's hot meals and constant support as I typed up my thesis back home. She truly created a stress free environment for me to work in and for that I am deeply grateful. My amazing brother Abdulaziz and my wonderful sisers Duha, Razan and Dalal were also there for me to provide words of encouragement and to lift me whenever I was down. I love you all!

I would also like to thank my good friend Fatima (SF) Fakhra for her support during the data collection process. I would never have managed to read all those articles so quickly if it hadn't been for her constant nagging and her staying up late with me to provide encouragement.

Finally, I would like to thank Dr. Phil Benson and Dr. Anthony Preston for the numerous corrections they made to my dissertation.

Noor Al-Ramadhan

Table of Contents

Abstract

Background: The QUOROM and PRISMA guidelines were created in an attempt to improve the standard of reporting systematic reviews. At present there are no studies in the dental literature that have assessed the compliance of papers with these two sets of guidelines.

Aims: To determine whether the reports of systematic reviews in four dental specialities comply with the QUOROM and PRISMA statements, whether there has been an improvement in standard over time and whether Cochrane reviews differ from other reviews.

Design: Retrospective observational study

Method: A search of the Cochrane library identified 181 systematic reviews and meta-analyses for inclusion across four dental specialities (orthodontics, periodontics, preventive dentistry and endodontics). Each review was scored using a 63-item checklist developed from the QUOROM guidelines and a 63-item checklist developed from the PRISMA guidelines.

Results: The mean QUOROM score for the whole sample was 70.86% (SD 11.36%, 95% CI 69.20%, 70.86%) and the mean PRISMA score for the whole sample was 74.07% (SD 10.48%, 95% CI 72.53%, 75.61%). The mean PRISMA score for Cochrane reviews was 85.19% (SD 5.03%, 95% CI 83.79%, 86.59%) and the mean PRISMA score for non-Cochrane reviews was 69.59% (SD 8.60%, 95% CI 68.09%, 71.09%). This difference was statistically significant (mean difference 15.50% (95% CI 13.58%, 17.62%; $p<0.00001$). The mean PRISMA score for orthodontic papers was 75.07% (SD 10.36%, 95% CI 72.32%, 77.82%), for periodontic papers it was 74.91% (SD 7.96%, 95% CI 72.80%, 77.03%), for preventive dentistry papers the means score was 71.50% (SD 13.73%, 95% CI 67.22%, 75.78%) and for endodontic papers the mean score was 74.20% (SD 9.37%, 95% CI 70.33%, 78.07%). The differences between these scores was not statistically significant ($p = 0.851$). There was a weak negative linear relationship between the age of a

Cochrane review and its PRISMA score, indicating a small improvement in compliance with the PRISMA guidelines over time. This was statistically significant ($p = 0.019$). There was also a weak negative linear relationship between the age of a non-Cochrane review and its PRISMA score but this was not statistically significant ($p=0.422$). The age of a paper, the speciality it belonged to and the type of review (Cochrane versus non-Cochrane) accounted for 46.5% of the variability in the final PRISMA score.

Conclusions: The compliance of systematic reviews and meta-analyses with the QUOROM and the PRISMA guidelines was highly variable. There were significant differences between the PRISMA scores of non-Cochrane reviews and Cochrane reviews with the latter scoring more highly. There was also a slight increase in the compliance of Cochrane reviews over time, which was statistically significant. Although the speciality of orthodontics had the highest mean PRISMA score, there were no significant differences between the four specialities.

Introduction

In the literature there are two types of research that are recognised: primary research and secondary studies. Systematic reviews and meta-analyses fall into the category of secondary research and can be described as overviews that review a body of data and in the case of a meta-analysis perform a statistical analysis to combine the results of primary studies. If conducted well they can allow a more objective appraisal of research evidence and may explain the heterogeneity between the results of different studies[1]. Systematic reviews and meta-analyses are considered to provide the highest level of evidence because they combine the results of randomised control trials, the 'gold standard' method of research for comparing the relative effectiveness of competing interventions. Consequently the sample size is increased which in turn increases the statistical power and the validity of the results obtained[2]. They have been classified by the Centre for Evidence Based Medicine as level 1a evidence.[3]

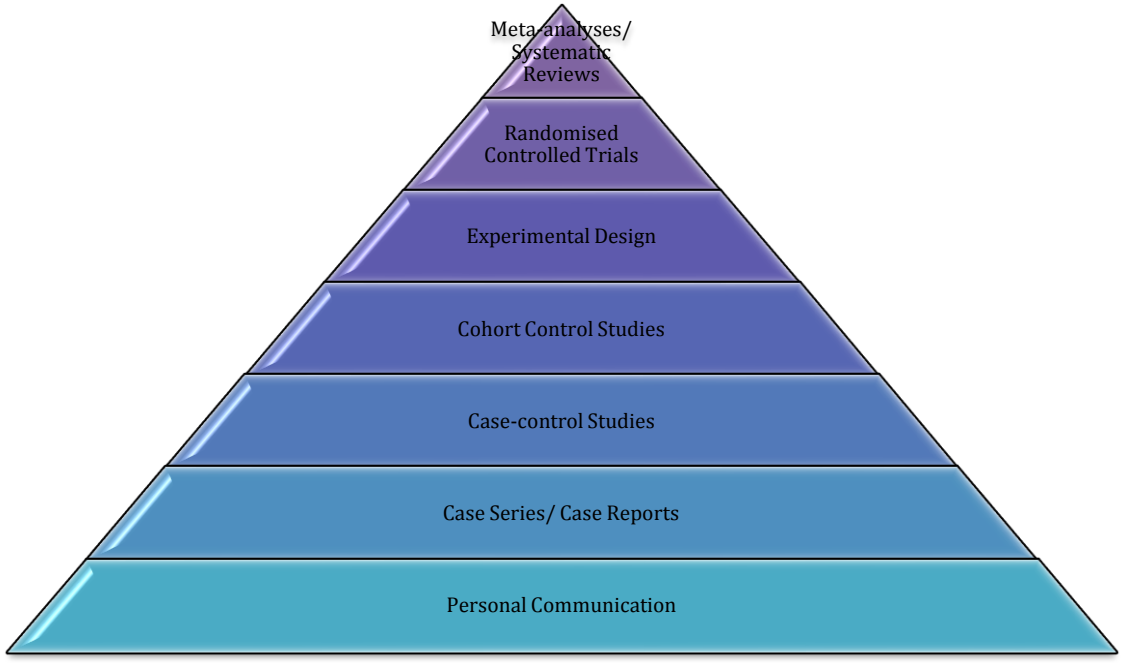


Figure 1 Hierarchy of evidence (modified from 4)

However, as with all types of research, systematic reviews and meta-analyses are not without their shortcomings. It has been found that a deficiency in their methodology can produce inflated or incorrect conclusions and hence an attempt was made in 1999 to improve the quality of reporting systematic reviews and meta-analyses by the instigation of the QUOROM statement.[5] The QUOROM statement consists of a set of guidelines that comprise all the factors that should be included in a high-quality systematic review or meta-analytic paper[5]. Since its publication there have been a number of articles in the medical literature that have looked at the compliance of systematic reviews with the QUOROM guidelines.[6-16] However at present there are no studies in the dental literature that have compared the quality of papers before and after the publication of the QUOROM statement.

Systematic Reviews and Meta-analyses: Are they synonymous?

While the two terms are often used interchangeably, it is important to distinguish between meta-analyses and systematic reviews.

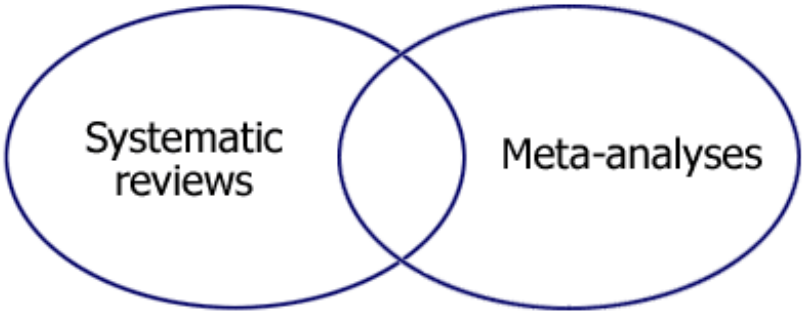


Figure 2 Systematic reviews and meta-analyses

A systematic review is "an overview of primary studies which contains an explicit statement of objectives, materials and methods and have been conducted according to explicit and reproducible methodology" [17] in order to limit bias. The term 'systematic review' refers to the whole process of data collection and analysis and all available evidence is included. A meta-analysis may be included in the results section of a systematic review and it is usually the final step in conducting such a review but it is not always necessary or appropriate to do so. There are several advantages of systematic reviews and these include minimum bias in identifying and rejecting studies as well as rigorous methodologies, which ensure that more accurate conclusions can be drawn. Pooling the results of studies allows the combining of vast amounts of information and this may enable the early identification of potentially effective therapeutic measures and thus their rapid application.

On the other hand a meta-analysis is a mathematical or statistical analysis that "combines or

integrates the results of several independent clinical trials considered by the analyst to be combinable".[18] As little as two primary clinical trials may be used to conduct a meta-analysis and it is fundamental that all studies included have addressed the same hypothesis using similar techniques. In other words a meta-analysis is possible without conducting a systematic review beforehand, although this is generally not good practice and an unbiased systematic review is a good way to start. A meta-analysis is conducted in two stages. The first stage consists of data extraction and the second stage constitutes evaluating the appropriateness of combining these data to obtain a point estimate of treatment effect. It is not valid to combine quantitative results from different studies, a weighted average is calculated after looking at the results of each study. Meta-analyses potentially offer several advantages such as the provision of a systematic method for synthesising evidence and the provision of quantitative overall results from individual studies. They also reduce the need for continued studies and allow questions to be addressed and answered that would not have been possible in individual studies due to their small sample size.

The Cochrane Collaboration

The Cochrane Collaboration is an international organisation that is made up of health care professionals, physicians, researchers and consumers. Its main aim is to "prepare, maintain and promote the accessibility of systematic reviews in all areas of health care". [19,20] Initially, a study was conducted in 1974, based on the writings of British Epidemiologist Archie Cochrane who had stressed the importance of having access to all of the available evidence in health care. [21]

Later on the British National Health Service provided funding which facilitated the establishment of the first "Cochrane Centre" in 1992. This led to the formation of six more Cochrane Centres

by 1994 and in 1995 the Cochrane Collaboration was officially registered as a company and charity. At present, 52 Collaborative Review Groups are responsible for most of the work done by the Cochrane Collaboration which is mainly the preparation and maintenance of Cochrane reviews. The methods used to prepare these Cochrane reviews are under constant scrutiny and various Methods Groups undertake the task of ensuring that rigorous and systematic methodologies are used to provide a high standard of reporting them.

The output of the Cochrane Collaboration can be found online in the Cochrane Library,[21] which is updated quarterly. The Cochrane Library contains several databases including the Cochrane Database of Systematic Reviews (CDSR), the Cochrane Central Register of Controlled Trials (CENTRAL) and the Cochrane Review Methodology Database (CRMD). The CDSR is the primary output of the Cochrane Collaboration and contains systematic reviews produced by the collaboration itself. The CENTRAL contains a bibiliography of trials that have been downloaded from other databases such as MEDLINE, EMBASE and LILACS in combination with results of hand searching journals and conference reports. The CRMD, on the other hand, contains a bibliography of books and articles that deal with evaluating effects of health care.

Methodologies of Systematic Reviews and Meta-Analyses

Systematic reviews and meta-analyses require a lot of intensive labour to produce them. It is important for researchers to demonstrate expertise not only in the subject they are reviewing but also in the methodology required to conduct such reviews. The rules of evidence-based medicine must be followed and it is especially necessary that methodology being used is systematic.

With the recent rapid increase in the number of systematic reviews and meta-analyses in the literature it has become increasingly apparent that not all of these reviews are truly systematic in nature and many of them lack rigour in their methodology, which may result in bias leading to inaccurate conclusions and consequently incorrect application of the results obtained. One of the main reasons for potential bias is the fact that although systematic reviews and meta-analyses are considered to provide the highest level of evidence, they remain studies which are both retrospective and observational. It is because of these disadvantages that it is crucial for these reviews to be constantly monitored and evaluated in order to guarantee a high standard of reporting so that bias is limited and results can be correctly applied to clinical practice.

Assessing Systematic Reviews and Meta-Analyses

Once the importance of evaluating systematic reviews and meta-analyses became recognised, researchers set out to find ways in which these studies could be assessed. Several papers were published that subjectively evaluated reviews while others produced generic checklists and guidelines that allowed a more systematic and objective appraisal. [22-27]

Oxman and Guyatt[22] developed a set of guidelines in 1991 to evaluate review articles in the literature. They named their guidelines the Overview Quality Assessment Questionnaire (OQAQ), which consists of a checklist that included nine items that can be scored as 'done', 'partially done/cannot tell' or 'not done' and a tenth item requiring a summary evaluation. The authors used the questionnaire to assess the quality of 36 published review articles. The sensitivity of OQAQ was evaluated by nine assessors who were faculty members at the Department of Clinical Epidemiology and Biostatistics and it was found that reviews that scored higher on the OQAQ had employed more rigorous methodologies. It was therefore concluded

that the questionnaire was a valid tool for assessing reviews.

In 1993 a program was established known as the Critical Appraisal Skills Program (CASP) which aimed to enable individuals to develop the skills necessary to make sense of and critically evaluate research evidence and use that knowledge in practice.[23] Based on a paper published in 1994 entitled "User's guide to the medical literature: How to use an overview" [24] the CASP produced a set of guidelines for systematic reviews and meta-analyses that would enable the reader to critique a review and get a better idea about its quality. The CASP guidelines concentrate on three main issues which are the validity of the study, the accuracy of the results and finally their applicability.[23] They consist of a series of ten questions where eight of the questions may be answered with 'yes', 'no', or 'can't tell' and the remaining two questions require more detailed responses.

Del Mar and Glasziou wrote a Cochrane review about "Antibiotics for Sore Throat" in 1999 and devised the Glasgow Appraisal Tool.[25] This tool is also a checklist with ten questions that assesses the validity of both systematic reviews and meta-analyses. The questions mainly look at the inclusion criteria for studies, the appropriateness of combining results, their overall precision and whether all the important outcomes are considered. [26]

Another tool that is available for the assessment of reviews is provided by the Aggressive Research Intelligence Facility (ARIF). ARIF is a specialist unit that is based in Birmingham and it aims to improve the incorporation of research evidence into healthcare decisions.[27] The tool produced by the unit begins with three screening questions which allow the appraiser to assess the review quickly, followed by another thirteen questions that allow the evaluation of the

review in more depth.

Other checklists include those constructed by Blettner, Cook, Geller, Goldschmidt, Meinert, Nony, Pogue and Sacks and these checklists vary in length and complexity ranging from a few with ten items to one containing 101 items.[21] Despite the construction of many quality assessment tools, there is a large variation in the quality of the instruments themselves and each one of them has different advantages and disadvantages depending on the criteria used to develop it.

Table 1 Critical Appraisal Tools for Systematic Reviews

Name	Type	Quality of:	Number of Items	Reference
OQAQ	Checklist	Report and method	9	Oxman and Guyatt (11)
CASP	Checklist	Report	10	Oxman and Guyatt (13)
Glasgow	Checklist	-	10	Glasziou (14)
ARIF	Checklist	Report and method	12	Birmingham

The QUOROM Statement

In 1996 the CONSORT statement was published in an effort to improve the standard of reporting randomised controlled trials after it was suggested that the standard of reporting them was poor.[28] The guidelines included in this statement provided a very meticulous framework that helped authors to publish high quality reports of RCTs. Even though several guidelines already existed that allowed the evaluation of systematic reviews and meta-analyses, they lacked the thoroughness of the CONSORT statement and so a conference was held in an attempt to devise

a similar tool that would comprehensively help researchers to improve the quality of reporting reviews. It was attended by clinical epidemiologists, clinicians, statisticians and researchers, who conducted meta-analyses, as well as editors who had an interest in systematic reviews. The conference led to the publication of the QUOROM statement in 1999.[5] The QUOROM statement stands for the **Qu**ality **O**f **R**eporting **O**f **M**eta-analyses and consists of a checklist and a flow diagram. The checklist is made up of eighteen items, in table format, that comprise all the factors that should be included in a high-quality meta-analysis or systematic review. The items are mainly related to the abstract, methods and results sections of a systematic review of randomised controlled trials and eight of these items are evidence based. Adhering to these guidelines ensures that the authors pay attention to detail and adequately report relevant information regarding the search strategy, paper selection, validity assessment, data abstraction, study characteristics and quantitative data synthesis.[5] The flow diagram details the selection process by which randomised controlled trials are initially selected and finally either discarded or included in the systematic review.

The QUOROM statement is the 'gold standard' in evaluating the standard of reporting systematic reviews due to a large portion of it being evidence based, meaning that the inclusion of many items in the checklist was based on research evidence. This implies that any systematic review or meta-analysis that has failed to comply with a certain item from the checklist will potentially have biased results. The guidelines were formed after a systematic review of systematic reviews was conducted and also after an expert panel consensus conference that employed a modified Delphi approach to analyse systematic reviews. The QUOROM statement was also pre-tested and modified afterwards to ensure its validity as a tool to assess the quality of systematic reviews and meta-analyses.[21]

When compared to the QUOROM statement, other scales and checklists have been found to focus on the methods section of a systematic review while neglecting other aspects. Out of all of the available instruments, only 5% (one checklist) looked at the title of a systematic review, 10% evaluated the abstract, 62% critiqued the introduction, 57% looked at data abstraction and 52% stressed the importance of including a description of the primary studies included in the review.[21]

Quality Assessment of Systematic Reviews and Meta-analyses: A Review of the Literature

As mentioned previously, even before the conference that led to the publication of the QUOROM statement, many tools for critical appraisal of systematic reviews and meta-analyses were being used to assess the standard of reporting them. One paper looked at meta-analyses in particular and recognised the importance of study design in providing accurate results. [29] The authors found that many of the previous meta-analyses lacked consistency and there was some heterogeneity of study outcomes, which led to inappropriate pooling of results and ultimately incorrect conclusions. They devised a set of guidelines to assess meta-analyses based on their discussion of the design and statistical issues. Other studies primarily looked at the extent to which heterogeneity was evaluated in meta-analyses and they showed that only 45% to 68% of meta-analytic procedures did tests for heterogeneity. [21]

In 1987 a study was conducted to assess the quality of 86 published meta-analyses.[30] The articles were assessed based on 23 items from six content areas that were considered to be necessary in meta-analytic procedures and it was found that only 28% of these meta-analyses included all six content areas. This study was subsequently updated in 1996 and included meta-

analyses published since 1987 and it was concluded that there had been very little improvement in the standard of reporting of meta-analyses.[31]

In 2005, meta-analyses, in the general surgical literature, were critically evaluated using OQAQ.[21] Papers, published between 1997 and 2002, were assessed and 487 potentially relevant papers were identified from MEDLINE. Out of these 51 met the inclusion criteria and they were subsequently critically appraised using the questionnaire. It was discovered that most studies exhibited deficiencies in their methodologies and it was postulated that this could be due to a lack of external collaboration and a lack of experience in conducting meta-analyses.

In another attempt to assess the quality of systematic reviews, four different systematic reviews, that had been published on the effect of Vitamin E on the cardiovascular system, were reviewed.[33] While the question posed in each of these four reviews was the same, the methodologies used in the identification and selection of studies were different and this led to very contradictory results. Two reviews concluded that Vitamin E did not benefit the cardiovascular system, one review found no correlation between Vitamin E and the cardiovascular system and the final review claimed that the vitamin was in fact harmful in high doses.

A further study on Cochrane reviews, published in 2001,[34] evaluated their quality. Although this was following the emergence of the QUOROM guidelines, the authors looked at Cochrane reviews from 1998, before the guidelines were published and used a fairly subjective evaluation process without any systematic method of assessing the reviews. Reviews were rated as having or not having major problems and the reasons for this were documented in detail. The study

concluded that Cochrane reviews only have minor problems if any and these problems tend to over-estimate the benefit of the therapeutic measure being looked at. [34]

Meta-analyses in occupational epidemiology have also been assessed in order to identify the major issues that may affect their quality.[35] Relevant articles were searched using several electronic databases and ultimately 60 articles were selected for analysis. More than half of the meta-analyses investigated the heterogeneity of studies, however eight of these studies used meta-analysis models even though they had significant heterogeneity. Most of the meta-analyses combined the results of all the primary studies they selected even though these studies varied immensely in the amount of information on exposure. The authors encouraged the proper exploration of heterogeneity so that the standard of reporting in occupational epidemiology could be improved.

The strengths and limitations of meta-analyses, based on aggregate data, have also been investigated.[36] All meta-analyses identified from MEDLINE, relating to cancer, were classified as either utilising individual patient data or summary or aggregate data. The vast majority of studies were found to be based on aggregate data and the authors saw this as a limiting factor. They concluded that whenever possible, individual patient data should be used due to their numerous advantages. The authors also concluded that aggregate patient data continues to be a part of most of the systematic reviews produced by the Cochrane Collaboration, the U.S Preventive Services Task Force and several other professional societies.

The standard of reporting Cochrane reviews with the standard of systematic reviews and meta-analyses published in paper-based journals has been explored.[37] The Cochrane reviews were

identified from the Cochrane Library and the paper-based journals were selected from MEDLINE. Only papers published in 1995 were identified for inclusion in the study and all 36 Cochrane reviews were included whereas 39 articles were selected from those found in MEDLINE. When this study was conducted, the QUOROM statement had not yet been published so the authors looked at the factors they thought were critical in a high quality review, including the number of authors, trials and patients included in the reviews as well as the trial sources, inclusion and exclusion criteria, language restrictions, primary outcome, trial quality assessment, heterogeneity testing and effect estimates. The authors found that while the papers selected from MEDLINE had more authors, trials and patients, the Cochrane reviews more thoroughly described the inclusion and exclusion criteria, assessed trial quality, avoided language restrictions, and updated their articles. It was concluded that Cochrane reviews employed superior methodologies and updated their reviews more often than systematic reviews and meta-analyses published in paper-based journals.

Quality Assessment using the QUOROM Guidelines

Recently studies have emerged in the literature, which have evaluated the quality of reporting as well as the compliance of systematic reviews and meta-analyses with the QUOROM statement[6-16] and in general it seems that much improvement has yet to be achieved. In 2005, one study tested the compliance of meta-analyses in the critical care literature with the QUOROM statement.[6] Studies published before and after the QUOROM statement were compared. 139 meta-analyses were included in the study and it was found that their overall quality was poor as only 30% had minimal flaws. Problems were found in the key parts of conducting a meta-analysis such as performing a literature review, limiting bias in the selection process and correctly referring to the validity of the chosen papers. When the papers written

before the QUOROM guidelines and the ones written after it were compared, an improvement was found.

Another study[7] showed that systematic reviews of traditional Chinese medicine were of low quality, as they did not follow guidelines outlined in the QUOROM statement rendering their results inconclusive and unreliable. On the other hand, a study looking at the quality of systematic reviews in the Cochrane Neonatal Review Group found that there was, in fact, a significant improvement in the quality of systematic reviews after the publication of the QUOROM statement.[8]

Hou *et al*., undertook a quality appraisal of systematic reviews and meta-analyses of pneumonia in China.[30] Two reviewers who were blinded to the authors analysed 326 reviews using the QUOROM statement as a guideline. Out of a possible 18 items that could be addressed, seven articles addressed ten items, five articles addressed two items and four papers addressed two items. It was concluded that the quality of reviews in the Chinese literature relating to pneumonia was weak in some aspects such as defining data sources, selection, searching, validity assessment, review methods and study characteristics.

The quality of systematic reviews used in oncology practice was evaluated in another study.[10] The authors applied the QUOROM guidelines to all reviews related to breast and colon cancer prevention and therapy. Eighty reviews were selected and assessed and the results showed that 29% of the reviews did not even match the definition of a systematic review. Twenty-one percent of the reviews did not adequately describe the searching methods employed and 70% were not systematic and were of very low quality. The authors warned oncologists that they

were in fact relying on poorly written documents and improvement was necessary.

A review of meta-analyses dealing with pharmacotherapy of post-traumatic stress disorder also assessed their standard using the QUOROM guidelines.[11] It was shown that the quality of meta-analyses was acceptable in the PTSD literature.

Bereza *et al*., evaluated the compliance of meta-analyses of anxiety disorders with the QUOROM guidelines.[12] They identified 136 papers from several electronic databases that were published between 1995 and 2007 but only 16 of them met the inclusion criteria. Results of this study showed that the QUOROM quality score was approximately 62% (SD19%). The lowest scores were obtained for the results sections of the meta-analyses, whereas the introduction and discussion sections scored the highest. The overall quality was 58% (SD 28%). It was concluded that the standard of reporting was below average and significant improvement was required.

Another study looked at the compliance of Health Technology Assessments (HTA) with the QUOROM guidelines.[13] All systematic reviews of therapeutic interventions in HTA that had been published between 2001 and 2005 were included in the study, resulting in a total of 87 papers. The results showed that 49% of all systematic reviews used a study selection flow diagram. It is interesting to note that when only systematic reviews, containing a meta-analysis were analysed, compliance with the QUOROM guidelines was found to be 32%.[13] Only 23% of all systematic reviews included a diagram that expressed the relationship between citations and studies. [14]

A random sampling study of meta-analyses in the medical literature was conducted in 2008

where a random sample of 161 papers was identified from MEDLINE and assessed according to the QUOROM guidelines.[14] The mean QUOROM score was 12.3 out of a possible 18 and this mean increased significantly from 10.5 in 2000 to 13.0 in 2005, indicating an improvement in the compliance with the QUOROM statement over time. The mean score of Cochrane reviews was 14.2 whereas the score for paper-based articles was 11.7. This indicates that Cochrane reviews exhibited a higher standard of reporting when compared with published papers. However, when looked at in detail, it became evident that Cochrane reviews scored higher mainly in the abstract section but they obtained lower scores in their trial flow. The authors encouraged an improvement in the standard of reporting articles published in the future.

Shea *et al*., compared the standard of reporting of original systematic reviews with updated Cochrane reviews in order to determine whether the updating process led to an improvement in their quality.[15] Fifty-three Cochrane reviews that were published in 2002 were included in the study and the updated and original versions were assessed using the QUOROM guidelines. It was concluded that the quality of reporting Cochrane reviews improved in some areas after they were updated. However, even though the standard of reporting increased on some individual items there was no overall improvement seen in the updated articles and the methodologies were for the most part consistent.

Shea *et al* also compared Cochrane reviews and systematic reviews published in paper-based journals.[16] Cochrane reviews were found to have a higher standard of reporting in some aspects whereas paper-based reviews had a higher standard in reporting others. The overall quality however was found to be low.

Quality Assessment in the Orthodontic Literature

While there has been a substantial increase in the number of systematic reviews and meta-analyses published in the literature, there remains a paucity of such studies conducted in orthodontics and it was found that between 1966 and 2002, while the medical literature had 8418 published meta-analyses, the orthodontic literature had only 13.[38] This may be because it is not possible to perform a meta-analysis due to a lack of appropriate primary studies that can be included for data synthesis.

As of yet there is one paper that has critically evaluated the quality of reporting meta-analyses in the orthodontic literature. The study was conducted by Papadopoulos and Gkiauris and they searched several electronic databases and hand searched journals in order to identify all meta-analyses in Orthodontics.[39] They retrieved 98 papers initially but only 16 papers met the inclusion criteria and so only those meta-analyses were included in the evaluation. It was found that, in general, some of the published articles incorporated appropriate methodologies and adequate quantitative data synthesis whereas others lacked rigorous methodologies such as biases in the selection process of the studies, lack of sufficient information to permit repeatability by other researchers and a lack of high-quality research leading to small sample sizes. Out of the 16 papers assessed the ones that provided the best level of evidence were those discussing maxillary protraction treatment,[40,41] the treatment of posterior crossbites,[42] the repeatability of lateral cephalometric measurements,[43] the relationship between anterior tooth injuries and size of the overjet,[44] correlation of external apical root resorption with treatment-related factors[45] and the prevalence of tooth agenesis.[46] Each paper was appraised individually according to where the authors thought their deficiencies lay but there was no assessment of the papers using specific guidelines such as those found in the PRISMA or

QUOROM statements.

The PRISMA Statement

In 2009, the QUOROM statement was updated to the PRISMA statement (**P**referred **R**eporting **I**tems of **S**ystematic reviews and **M**eta-Analyses).[48] This was done in order to address the advances in the science of systematic reviews. The PRISMA statement consists of a 27-item checklist and a flow diagram, similar to the one found in the QUOROM statement.

Because the guidelines have only changed recently, only one paper so far has looked at the compliance of systematic reviews and meta-analyses with the PRISMA statement.[49]

Aim

- The overall aim of this study was to assess the quality of systematic reviews and meta-analyses in the dental literature.

Objectives

Primary Objectives

- To assess the compliance of systematic reviews and meta-analyses in the orthodontic, periodontal, preventive dentistry and endodontic literature with the PRISMA guidelines
- To assess whether there has been an improvement in the quality of systematic reviews and meta-analyses over time

Secondary Objectives

- To determine whether or not there is a difference in the perception of quality of the review when scoring with the QUOROM guidelines or the PRISMA guidelines
- To determine if there is a difference between the compliance of reviews with the PRISMA statement amongst the specialities
- To determine if there is a difference between the standard of Cochrane reviews and non-Cochrane systematic reviews and meta-analyses

Null Hypotheses

Primary hypothesis

- There is no correlation between the compliance of reviews with the PRISMA guidelines and the year of publication of those reviews

Secondary Hypotheses

- There is no difference in the compliance of systematic reviews and meta-analyses with the QUOROM statement and the PRISMA statement.
- There is no difference in the compliance of Cochrane reviews and non-Cochrane systematic reviews with the guidelines of the PRISMA statement.
- There is no difference between the specialities in their compliance with the PRISMA guidelines.

Material and Methods

Study Design

This was a retrospective observational study.

Sample

Sample Size Calculation

The sample size calculation was originally based on expected differences between QUOROM scores before and after the publication of these guidelines and was derived from the raw data from a systematic evaluation of the quality of meta-analyses in the critical care literature.[6] The calculation was performed using Altman's nomogram and it was found that to achieve 80% power, with a 0.05 two-sided significance level, 165 papers were needed in the present study. As a preliminary search of systematic reviews in the orthodontic literature failed to return a sufficient number of papers the decision was made to include three other specialties in the study, namely periodontics, preventive dentistry and endodontics.

Papers

The following papers were included in the study:

- Systematic reviews and meta-analyses pertaining to any of the specialities of orthodontics, periodontics, preventive dentistry and endodontics
- All original Cochrane reviews in the Cochrane library that were found to be relevant to these four specialities
- Updated versions of these Cochrane reviews, if available

Time Periods

Originally the intent of this study was to include the five years preceding the publication of the QUOROM guidelines and the ten years following it, but a sparsity of papers published before the year 1999 necessitated the omission of any time period restrictions and thus all systematic reviews and meta-analyses were included irrespective of publication date.

Search Methods for Identification of Studies

Electronic Searches

An advanced computerised search was conducted of the Cochrane library. Four separate searches were performed, one for each speciality. To identify orthodontic papers the term 'orthodontics' was entered as a MeSH descriptor and all trees were exploded. The search results displayed papers from five different databases. All papers identified within the Cochrane Database of Systematic Reviews and the Database of Abstracts of Reviews of Effects were included in the study. The Cochrane Database of Systematic Reviews displays the most up-to-date version of Cochrane reviews so each Cochrane review was further inspected for any earlier versions and these were included as well. The same search strategy was repeated for the other three specialities with the terms 'periodontics,' 'preventive dentistry' and 'endodontics' entered separately for each search. This review did not have any language restrictions. The date of the last search was January 31st, 2011.

Searching other resources

No other search strategies were employed.

Data Collection and Analysis

Data were collected on each paper for both the QUOROM and the PRISMA guidelines. A checklist was compiled for each, based on the published statement.

For the QUOROM statement the guidelines were further divided into a checklist of 63 items. For each criterion except for the ones pertaining to the inclusion and exclusion criteria in the methods section, a score of one was awarded where the criterion was met and score of zero where it was not. Items 28-31 relate to the inclusion criteria and where at least three out of four criteria were met one point was awarded. If two or less criteria were met, the score for the inclusion criteria was zero. Items 32-35 relate to the exclusion criteria and again a score of one was only awarded if at least three out of these four criteria were met. The maximum score for each paper was therefore 57. In some instances a specific criterion was not applicable to the study and in this case, instead of a score the term N/A (or non-applicable) was appointed for that item in the score sheet. To allow direct comparisons between the reviews the final score for each paper was calculated as a percentage.

$$\text{Final score} = \frac{\text{Total points achieved} \times 100}{(57 - \text{number of N/As})}$$

The maximum attainable score for each study was 100% and this indicated that the reporting of the systematic review or meta-analysis abided by all the criteria recommended by the QUORUM statement.

For the PRISMA statement the guidelines were again used to construct a checklist of 63 items. The rules for scoring were similar to the QUOROM checklist but no items were grouped so for every item a score of one was possible where the criterion was met. This enabled each paper to

achieve a maximum score of 63. Again, the final score was calculated as a percentage.

$$\text{Final score} = \frac{\text{Total points achieved} \times 100}{(63 - \text{number of N/As})}$$

Pilot Study

A pilot study was conducted on a ten percent, randomly selected sample of the identified papers to determine inter-examiner reliability for the PRISMA checklist. A list of papers was prepared by JEH using a random number generator and these papers were scored separately by the supervisor, JEH, and the examiner, NR. Inter-examiner agreement was determined by calculating a percentage agreement kappa value for each item on the checklist and where a criterion scored less than 0.80, out of a possible maximum score of 1.00, the description of the item was modified, after discussion with the supervisor, to clarify the assessment criteria. The item was then rescored until an acceptable level of reliability was obtained with a kappa score of 0.80 or above.

Data Collection

The data were entered into two Microsoft Excel® spreadsheets (Microsoft Office 2003, Microsoft Corporation, Redmond, WA98052-7329, USA), one for the QUOROM checklist and one for the PRISMA checklist. For the QUOROM checklist the spreadsheet used was made up of 57 columns representing the criteria being assessed and each row represented the systematic reviews and meta-analyses included in the study. For the PRISMA checklist the spreadsheet used consisted of 63 columns for each criterion and a row for each paper assessed. Due to the time taken to assess a paper a maximum of three papers were scored at a time to prevent examiner fatigue and hence reduce scoring errors.

Reliability

To attain an acceptable level of intra-examiner agreement for both the QUOROM checklist and the PRISMA checklist a reliability study was conducted in addition to the main study. A list of ten percent of the papers was prepared using a random number generator and data on these papers were collected on both checklists by NR six months after the completion of data collection for the entire sample. NR was blinded to the initial results whilst re-examining the papers. A kappa score was calculated for each item on the two checklists, with the highest possible score being 1.00, indicating perfect agreement.

Statistical Analysis

SPSS and RevMan

Comparison of checklists

The QUOROM and PRISMA checklists were compared using the Bland and Altman test to determine whether or not the guidelines were interchangeable. Further statistical analyses were conducted on the PRISMA checklist scores as these guidelines are more recent.

Descriptive statistics

Descriptive statistics were initially used to summarise the data obtained for the PRISMA guidelines. The mean percentage score, standard deviation, 95% confidence interval, median, range and interquartile range were calculated for the following data sets:

- The data for the total sample
- The total sample of Cochrane reviews
- The total sample of non-Cochrane reviews

- The total sample of orthodontics reviews
- The orthodontic Cochrane reviews
- The orthodontic non-Cochrane reviews
- The total sample of periodontic reviews
- The periodontic Cochrane reviews
- The periodontic non-Cochrane reviews
- The total sample of preventive dentistry reviews
- The preventive dentistry Cochrane reviews
- The preventive dentistry non-Cochrane reviews
- The total sample of endodontic reviews
- The endodontic Cochrane reviews
- The endodontic non-Cochrane reviews

Comparisons of scores

Normality of the data was determined using visual tests such as the Q-Q plot and the P-P plot in addition to Shapiro-Wilks' test for normality and the Kolmogorov-Smirnov test. These tests indicated that the data were not normality distributed so non-parametric tests were used to analyse the data. Data were considered to be statistically significant when $p<0.05$.

Weighted mean differences were used to make several single comparisons. Each set of data were weighted according to sample size. The following single comparisons were made:

1. All Cochrane reviews vs. all non-Cochrane reviews
2. All orthodontic Cochrane review papers vs all orthodontic non-Cochrane reviews
3. All periodontal Cochrane reviews vs. all periodontal non-Cochrane reviews
4. All preventive dentistry Cochrane reviews vs. all preventive dentistry non-Cochrane reviews

5. All endodontic Cochrane reviews vs. all endodontic non-Cochrane reviews

Kruskall-Wallis tests were used to compare the four specialities i.e: orthodontics vs periodontics vs preventive dentistry vs endodontics

It was not feasible to compare reviews published before and after the publication of the PRISMA statement, which were only published recently, as there was an insufficient number of papers published after the statement was issued. Similarly, a relatively small number of papers was published before the QUOROM statement in 1999 so it was not possible to compare reviews published before and after this statement was issued. Therefore, in order to determine whether or not there had been an improvement in the quality of reporting systematic reviews over time, each paper was first assigned an age at 31st January 2010 according to the year of publication. For example, a paper published in 2009 was appointed an age of one year, a paper published in 2008 was two years old and so forth. The ages of the papers were then plotted against their PRISMA scores and a line of best fit was drawn. A Pearson correlation coefficient was calculated and this was used to establish if there was a relationship between the age of a paper and the score attained. The papers were then separated into Cochrane and non-Cochrane reviews, after which each group was plotted separately against age in order to determine if there was a difference in correlation coefficients between the two groups, which would indicate that one group was improving with time while the other was not.

Finally, multiple linear regression was used to model the extent to which three predictor variables determined the final PRISMA score of a paper. These variables were:

1. The type of paper (Cochrane or non-Cochrane).
2. The speciality the paper was related to
3. Age of paper

Results

Sample Characteristics

The search strategy identified 181 potentially relevant systematic reviews and meta-analyses across the four specialities. (Figure 2.) There were no systematic reviews or meta-analyses published before 1996 and the most recent systematic review included in the study was published in 2009. The earliest systematic review published in the literature was in the speciality of periodontics.

Table 2 Year of publication of earliest systematic review by speciality

Speciality	Review	Year
Orthodontics	• Orthodontic treatment for posterior crossbites	1999
	• Occlusal treatments in temporomandibular disorders: a qualitative systematic review of randomised controlled trials	1999
	• Therapeutic outcome assessment in permanent temporomandibular joint disc replacement	1999
Periodontics	• Microbiological response to mechanical treatment in combination with adjunctive therapy: a review of the literature	1996
Preventive Dentistry	• A systematic review of the effectiveness of health promotion aimed at improving oral health	1998
Endodontics	• A systematic review of in vivo retrograde obturation materials	2003

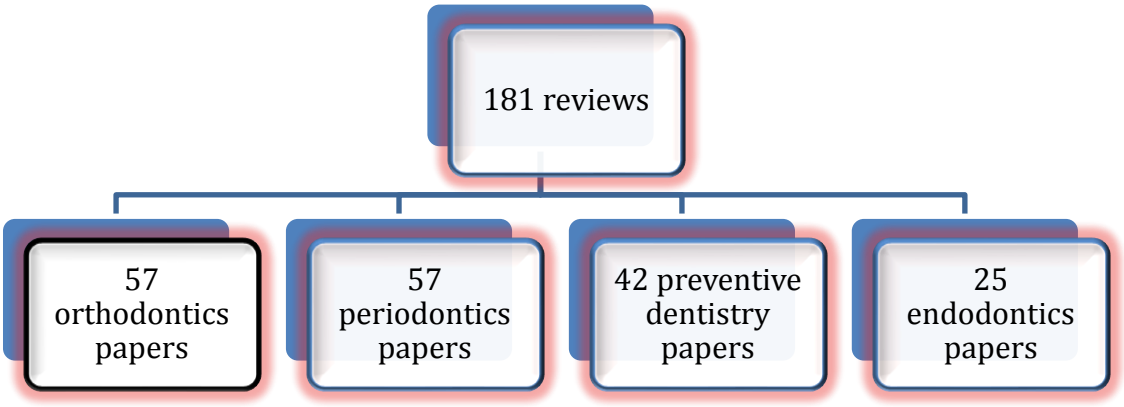


Figure 3 Flowchart of distribution of papers amongst the specialities

Table 3 Number of Cochrane and non-Cochrane reviews in each speciality

Speciality	Cochrane Reviews	Non-cochrane reviews
Orthodontics	20	37
Periodontics	9	48
Preventive dentistry	13	29
Endodontics	10	15

Pilot Studies

Pilot to assess the validity of the search

A ten percent sample of the papers identified by the search was analysed to establish if they were indeed systematic reviews or meta-analyses. The papers were numbered from one to 181 and a random number generator was used by supervisor JEH to select 19 papers. The papers were assessed by examiner NR in accordance with the Cochrane Handsearchers' Handbook. All 19 papers met the criteria thus it was concluded that the search was accurate and all 181 papers were included in the study.

Pilot to assess inter-examiner agreement

A ten percent sample of the articles was selected using a random number generator by supervisor JEH resulting in a total of 19 papers. The 63-item PRISMA checklist was used to score these papers by JEH and NR, independently and in duplicate. For each item there were three possible scores: one, zero, or non-applicable. Inter-examiner agreement was assessed by tabulating the number of agreements and disagreements for each item and subsequently calculating a kappa value. Hence 63 kappa values were calculated and evaluated. An acceptable level of agreement was said to be attained when a kappa value of 0.80 or above was obtained, indicating at least 80% agreement.

Out of the 63 items, six had relatively low kappa scores:

- Item 7: mention of the participants in the abstract (kappa score 0.53)
- Item 20: specification of the comparisons in the methods section (kappa score: 0.72)
- Item 21: specification of the outcomes in the methods section (kappa score: 0.72)
- Item 28: mention of the date last searched in the methods section (kappa score: 0.47)

- Item 29: presentation of the full electronic strategy for at least one database, including any limits used, so that it could be repeated (kappa score: 0.72)
- Item 39: presentation of any assessment of risk of bias across studies (kappa score: 0.53)

Differences in scoring of these items were resolved by discussion between JEH and NR and comments were added to definitions of these six items to increase the accuracy of scoring.

Table 4 Items with low Kappa scores

Item	Descriptions	Comments
6	Abstract: mention of the participants	Actual population looked at in each trial eg adolescents, II/1
20	Methods: specification of the comparisons	Compare to other group or control
21	Methods: specification of the outcomes	Detailed outcome measures, including measuring units
28	Methods: mention of the date last searched	Explicitly state date of last search, not just the time period included
29	Methods: presentation of the full electronic strategy for at least one database, including any limits used	Search terms not enough, table required with detailed strategy
56	Discussion: presentation of any assessment of risk of bias across studies	E.g. publication bias

Following discussion these items were rescored and new kappa values were calculated which demonstrated very good inter-examiner agreement, so no further modification of the checklist was required.

Pilot to assess intra-examiner agreement

A ten percent sample of the papers was assessed for intra-examiner agreement. 19 papers were selected by the supervisor, JEH, using a random number generator and were rescored by NR six months after the completion of data collection for the main study. Both the QUOROM checklist and the PRISMA checklist were used to rescore the papers. Kappa values were calculated for each item on each of the checklists and a very high level of intra-examiner agreement was achieved indicating excellent reliability of the measurement tools.

Main study

Assessing normality of the data

Both the scores obtained for the QUOROM checklist and the PRISMA checklist were tested for normality. A variety of investigations were employed including diagrammatic representations of the distribution of the data, and two tests for normality, namely the Shapiro-Wilk and the Kolmogorov-Smirnov tests. (See Figures 3-6 and Table 5). The majority of the investigations indicated that the data were not normally distributed therefore non-parametric tests were used to analyse this data.

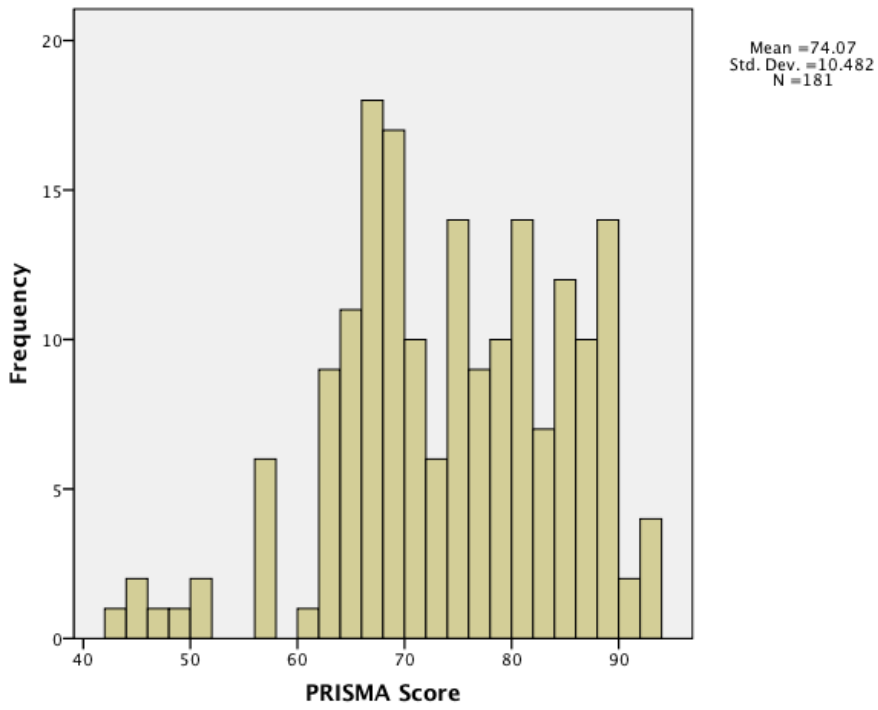


Figure 4 Histogram of the frequency of each PRISMA score

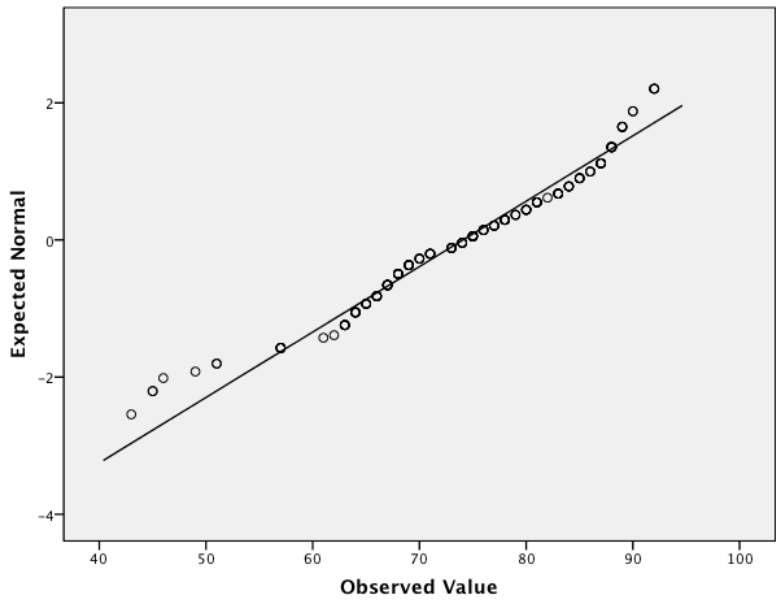


Figure 5 Normal Q-Q Plot of PRISMA score

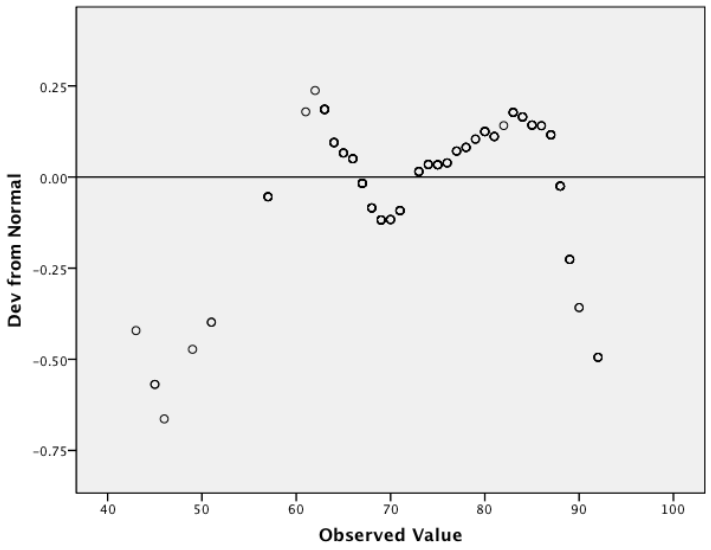


Figure 6 Detrended normal Q-Q plot of PRISMA score

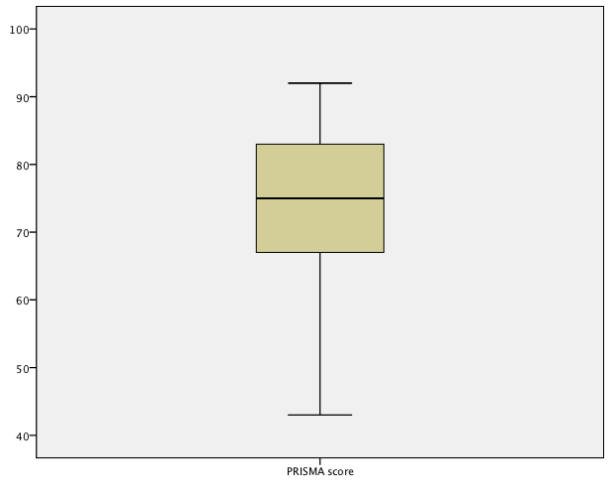


Figure 7 Boxplot of PRISMA score

Table 5 Tests for normality

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
PRISMA Score	0.068	181	0.40	0.965	181	P<0.001

Comparison of the checklists

The Bland and Altman test was used to compare the scores of the QUOROM and the PRISMA checklists. The results of this test showed that there was a bias of -3.3, which was statistically significant ($p<0.001$). A visual analysis of the difference plot indicated that approximately 5% of the points lay outside the 95% limits of agreement.

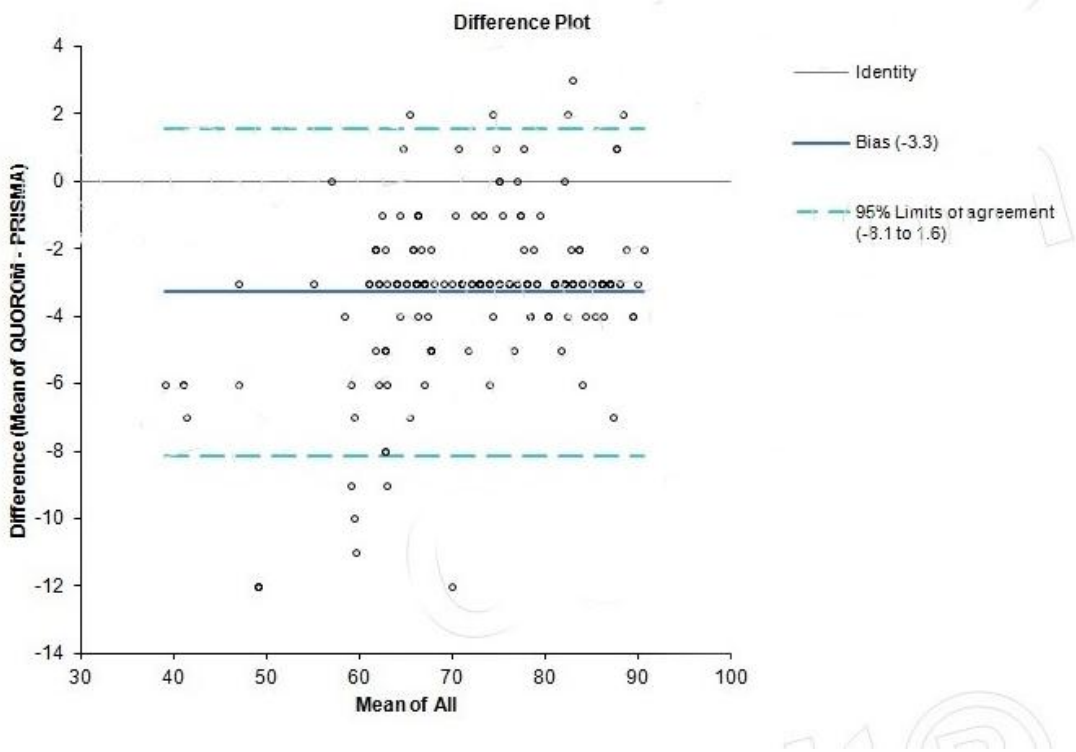


Figure 8 Bland and Altman plot for interexaminer agreement

Although there was bias between the two measurement tools which was statistically significant, save for a few outliers the QUOROM scores were consistently lower than the PRISMA scores by approximately 3 percent per score.

Descriptive Statistics for the QUOROM and PRISMA guidelines

In addition to calculating means, standard deviations and 95% confidence intervals for the data, medians, ranges and interquartile ranges were calculated as well because the data were not normally distributed. The mean overall score for the papers when assessed using the QUOROM checklist was 70.86% and the median was 72.00%. For the PRISMA checklist the mean overall score was 74.07% and the median was 75.00%. (Tables 6 and 7).

Table 6 Descriptive statistics for the QUOROM guidelines

Statistic	Value (%)
Mean	70.86
Standard Deviation	11.36
95% Confidence Interval	69.20-70.86
Median	72.00
Minimum	37
Maximum	89
Range	52
Interquartile Range	17

Table 7 Descriptive statistics for the PRISMA guidelines

Statistic	Value (%)
Mean	74.07
Standard Deviation	10.48
95% Confidence Interval	72.53-75.61
Median	75.00
Minimum	43
Maximum	92
Range	49
Interquartile Range	16

Highest and lowest scoring papers

The highest score achieved by any paper was 92%. This was achieved by three orthodontic Cochrane reviews and one Cochrane review in periodontics.

Orthodontics:

1. Occlusal interventions for periodontitis in adults
2. Occlusal adjustment for treating and preventing temporomandibular joint disorders (the updated version)
3. Fluorides for the prevention of white spots on teeth during fixed brace treatment (the updated version)

Periodontics:

1. Guided tissue regeneration for periodontal infra-bony defects

The lowest score achieved by any paper was in the speciality of preventive dentistry, entitled: "Reviews of evidence on interventions to prevent dental caries, oral and pharyngeal cancers, and sports-related craniofacial injuries**."**

Comparison of Cochrane and Non-Cochrane Reviews

Out of the 181 papers, 52 papers were Cochrane reviews and 129 were other systematic reviews or meta-analyses. The mean overall score for Cochrane reviews, assessed using the PRISMA checklist was 85.19% (SD 5.03). On the other hand the median score was 86.00%. The mean overall score for non-Cochrane reviews was 69.59% (SD 8.60) and the median score was 69.00%. (See table 8).

Table 8 Descriptive statistics for Cochrane and non-Cochrane reviews

Statistic	Cochrane Reviews	Non-Cochrane Reviews
Mean	85.19	69.59
Standard Deviation	5.03	8.60
95% Confidence Interval	83.79-86.59	68.09-71.09
Median	86.00	69.00
Minimum	65	43.00
Maximum	92	90.00
Range	27	47.00
Inter-quartile range	5	10.50

When the mean PRISMA score for Cochrane reviews was compared to the mean PRISMA score for non-Cochrane reviews using weighted mean difference, the difference in means was found to be statistically significant ($p<0.00001$, WMD 15.60, 95% CI 13.58, 17.62). (See figure 8).

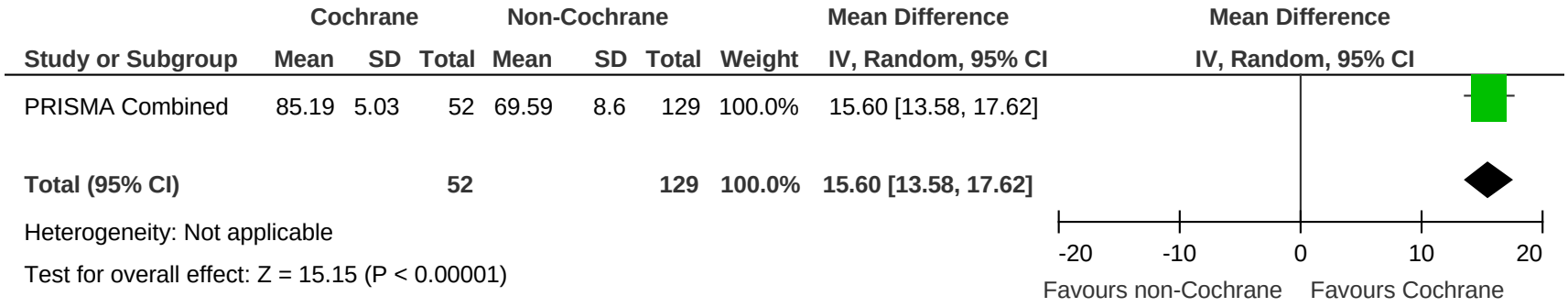


Figure 8 Mean difference for mean PRISMA score of all Cochrane reviews and the mean PRISMA score of all non-Cochrane reviews

Differences amongst the specialities

The specialities of orthodontics and periodontics had the highest total number of reviews published, while endodontics had the smallest number of published reviews. The greatest number of Cochrane reviews has been published in the orthodontic literature.

The mean score for each speciality is as follows:

1. Orthodontics mean score: 75.07% (SD 10.36)
2. Periodontics mean score: 74.91% (SD 7.96)
3. Preventive dentistry mean score: 71.50% (SD 13.73)
4. Endodontics mean score: 74.20 (SD 9.37)

Table 9 Descriptive statistics for PRISMA scores by speciality

	Orthodontics	Periodontics	Preventive Dentistry	Endodontics
Mean	75.07	74.91	71.50	74.20
Standard Deviation	10.36	7.96	13.73	9.37
95% Confidence Interval	72.32-77.82	72.80-77.03	67.22-75.78	70.33-78.07
Median	73	75	74	80
Minimum	57	57	43	57
Maximum	92	92	89	87
Range	35	35	46	30
Inter-quartile Range	19	10.50	17.50	16.50

Kruskall-Wallis Tests

As can be seen above, when comparing the mean PRISMA score for each speciality it is evident that there were differences between them. The speciality of orthodontics scored the highest

with a mean score of 75.07%. Periodontics scored the second highest with a score of 74.91%. The lowest score belonged to preventive dentistry and was 71.50%. The mean score for endodontics was 74.20%, althought interestingly, endodontics had the highest median score. Nonetheless, the Kruskall-Wallis test showed that there were no statistically significant differences between the four specialities (p= 0.851, chi-square= 0.796, df=3).

However, the percentage of Cochrane reviews in relation to the total sample, which had already been shown to score higher on average than non-Cochrane reviews, was different for each speciality. In orthodontics 35% of the papers were Cochrane reviews. In periodontics Cochrane reviews made up only 19% of all the papers, while in preventive dentistry they made up 31% and in endodontics 40% of the papers were Cochrane reviews. (Table 9). Due to this unequal distribution of Cochrane and non-Cochrane systematic reviews amongst the specialities these two groups were compared separately.

Table 10 Distribution of Cochrane and non-Cochrane reviews amongst the specialities

Speciality	Cochrane Reviews	Non-Cochrane Reviews	Total
Orthodontics	20	37	57
Periodontics	9	48	57
Preventive dentistry	13	29	32
Endodontics	10	15	25

Accordingly the Kruskall-Wallis test was used to first compare Cochrane reviews amongst the specialities. It was shown that there was a statistically significant difference between the mean scores of orthodontics, periodontics, preventive dentistry and endodontics (chi-square 9.88, df

3, p=0.0196). The highest mean score for Cochrane reviews was 87.22% and was found in the speciality of periodontics. Orthodontics came in second with a mean score of 85.35%. This was closely followed by preventive dentistry that had a mean score of 85.23% and the lowest mean score was in the speciality of endodontics and was 83.00%.

Non-cochrane reviews were also compared using the test and once more there was a statistically significant difference between the specialities (chi-square 9.48, df=3, p=0.0235). Again, periodontics had the highest mean score which was 72.6%, followed by orthodontics (mean= 69.51%). The speciality of endodontics achieved a mean score of 68.33 and this time it was preventive dentistry that scored the lowest with a mean of 65.34%.

Table 11 Mean scores for Cochrane and non-Cochrane reviews in each speciality

Speciality	Cochrane Reviews	Non-Cochrane Reviews
Orthodontics	85.35 (SD 7.15)	69.51 (SD 7.05)
Periodontics	87.22 (SD 2.68)	72.6 (SD 6.31)
Preventive Dentistry	85.23 (SD 3.09)	65.34 (SD 12.05)
Endodontics	83 (SD 2.58)	68.33 (SD 7.35)

Orthodontics

The mean score for orthodontic Cochrane reviews was 85.35% (SD 7.15). For other orthodontic reviews and meta-analyses the mean score was 69.51% (SD 7.05). (Table 12). Comparison of the means using the weighted mean difference showed a statistically significant difference between them (p< 0.00001, WMD 15.84, 95% CI 11.97, 19.71). (Figure 9).

Table 12 Descriptive statistics for Cochrane and non-Cochrane reviews in orthodontics

Statistic	Cochrane Reviews	Non-Cochrane Reviews
Mean	85.35	69.51
Standard Deviation	7.15	7.05
95% Confidence Interval	82.01-88.70	67.16-71.86
Median	87.50	68
Minimum	65	57
Maximum	92	87
Range	27	30
Inter-quartile range	3.75	11

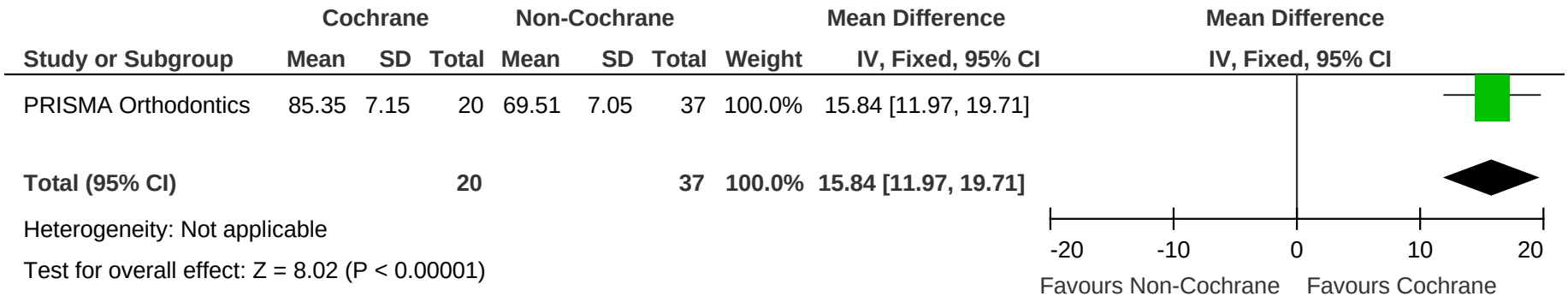


Figure 9 Weighted mean difference for mean PRISMA score of Cochrane reviews and non-Cochrane in orthodontics

Periodontics

Cochrane reviews in periodontics had a mean score of 87.22% (SD 2.68). Non-Cochrane reviews and meta-analyses had a mean score of 72.6% (SD 6.31). (Table 13). When these means were compared using the weighted mean difference it was shown that there was a statistically significant difference between the two (p< 0.00001, WMD 14.62, 95% CI 12.12, 17.12). (Figure 10).

Table 13 Descriptive statistics for Cochrane and non-Cochrane reviews in periodontics

Statistic	Cochrane Reviews	Non-Cochrane Reviews
Mean	87.22	72.60
Standard Deviation	2.68	6.31
95% Confidence Interval	85.16-89.28	70.77-74.44
Median	87	73.50
Minimum	83	57
Maximum	92	90
Range	9	33
Inter-quartile range	4	8.75

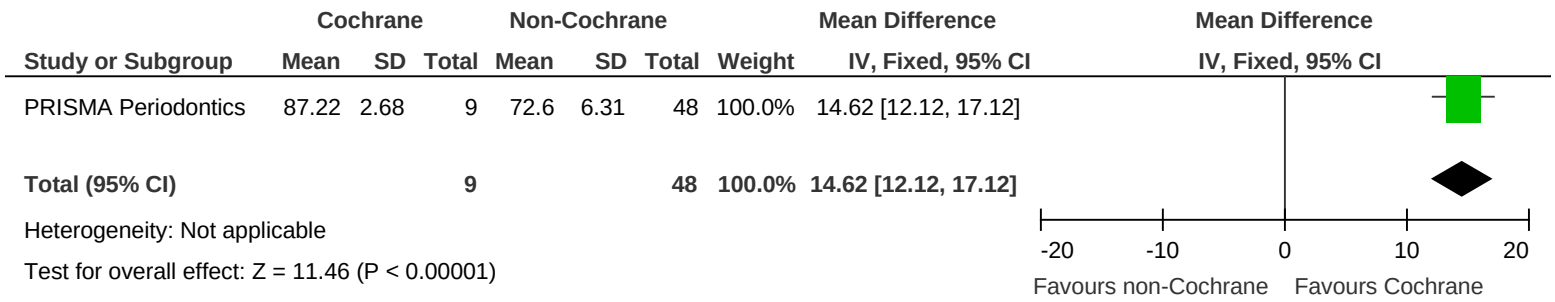


Figure 10 Weighted mean difference for mean PRISMA score of Cochrane reviews and non-Cochrane reviews in periodontics

Preventive Dentistry

In preventive dentistry, Cochrane reviews had a mean score of 85.23% (SD 3.09) whereas other reviews had a mean PRISMA score of 65.34% (SD 12.05). (Table14). Comparison of the two means using weighted mean difference showed that there was a statistically significant difference between Cochrane and non-Cochrane reviews in this speciality ($p<0.00001$, WMD 19.89, 95% CI 15.19, 24.59). (Figure 11).

Table 14 Descriptive statistics for Cochrane and non-Cochrane reviews in preventive dentistry

Statistic	Cochrane Reviews	Non-Cochrane Reviews
Mean	85.23	65.34
Standard Deviation	3.09	12.05
95% Confidence Interval	83.37-87.10	60.76-69.93
Median	86	67
Minimum	81	43
Maximum	89	86
Range	8	43
Inter-quartile range	6	20.50

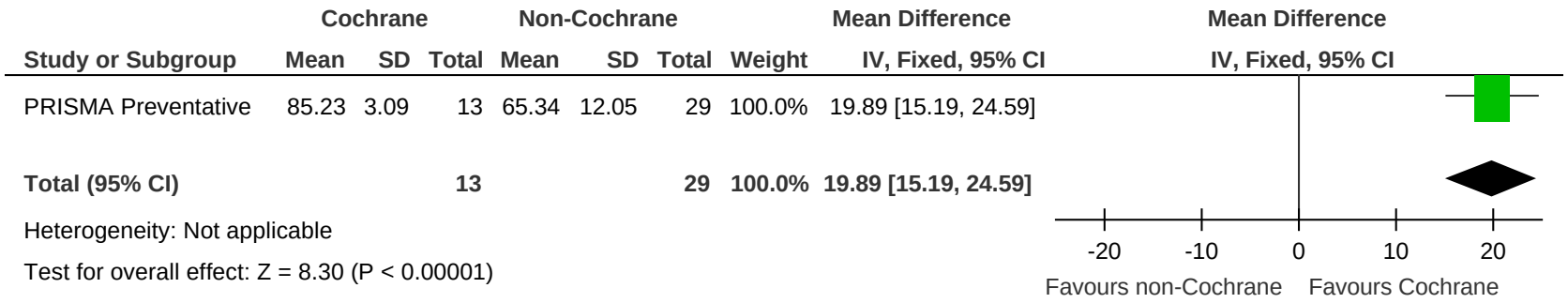


Figure 11 Weighted mean difference for mean PRISMA score of Cochrane reviews and non-Cochrane reviews in preventive dentistry

Endodontics

As with the other three specialities, in endodontics Cochrane and non-Cochrane reviews had different mean scores with Cochrane reviews scoring higher than other types of review articles. The mean score for Cochrane reviews was 83% (SD 2.58) and for other reviews the mean was 68.33% (SD 7.35). (Table 14). Weighted mean difference was again used to test for significant differences between the means and it was found that there was indeed a statistically significant difference between them ($p<0.00001$, WMD 14.67, 95% CI 10.62, 18.72). (Figure 12).

Table 15 Descriptive statistics for Cochrane and non-Cochrane reviews in endodontics

Statistic	Cochrane Reviews	Non-Cochrane Reviews
Mean	83.00	68.33
Standard Deviation	2.58	7.35
95% Confidence Interval	81.15-84.85	64.26-72.41
Median	82.94	67
Minimum	80	57
Maximum	87	82
Range	7	25
Inter-quartile range	4	6

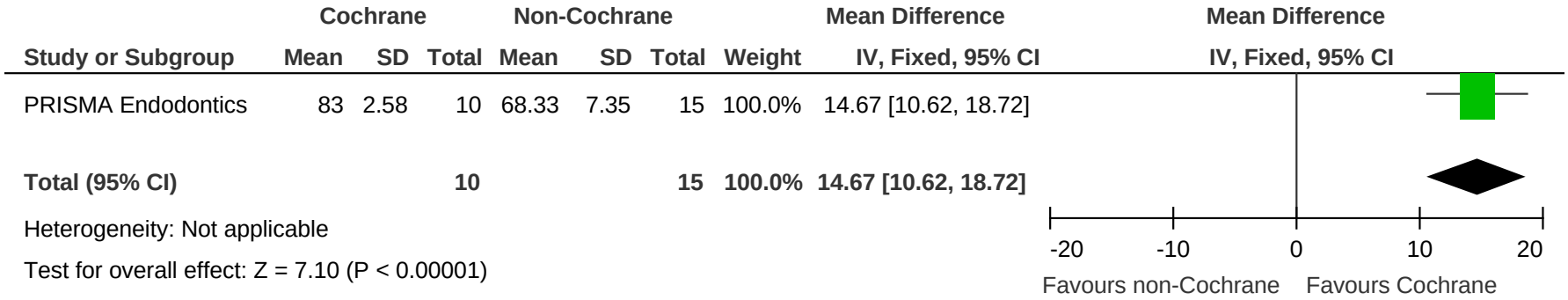


Figure 92 Weighted mean difference for PRISMA score of Cochrane reviews and non-Cochrane reviews in endodontics

Meta-analysis of mean scores for Cochrane and non-Cochrane reviews

The results for the mean Cochrane and the mean non-Cochrane scores for each speciality were combined and a meta-analysis was performed to assess the overall difference in scores between the two types of reviews, while taking into account the different sample sizes for each speciality. There was no significant heterogeneity within the sample ($p=0.26$) so a random effects model was not required when comparing Cochrane and non-Cochrane reviews. When the results of each speciality were combined there was a statistically significant difference between the two mean scores ($p<0.00001$, WMD 15.59, 95% CI 13.86, 17.32). (Figure 13).

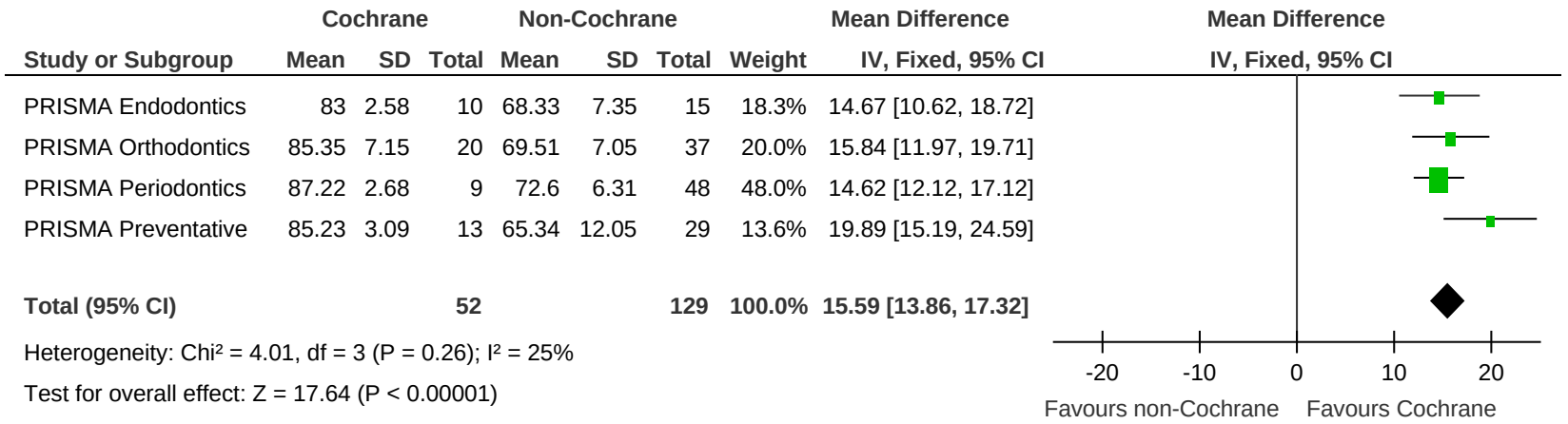


Figure 13 Meta-analysis of mean scores of Cochrane and non-Cochrane reviews

Changes in PRISMA scores with time

As mentioned previously, each paper was assigned an age according to publication date, with the minimum possible age being zero years if the paper was published in 2010 and included in the Cochrane Library before the date of last search (January 31st, 2010). The most recent papers included in this review were published in 2009 and the oldest paper was 14 years old. The age of each paper was plotted against the PRISMA score it achieved in order to establish if there was any correlation between the two variables.

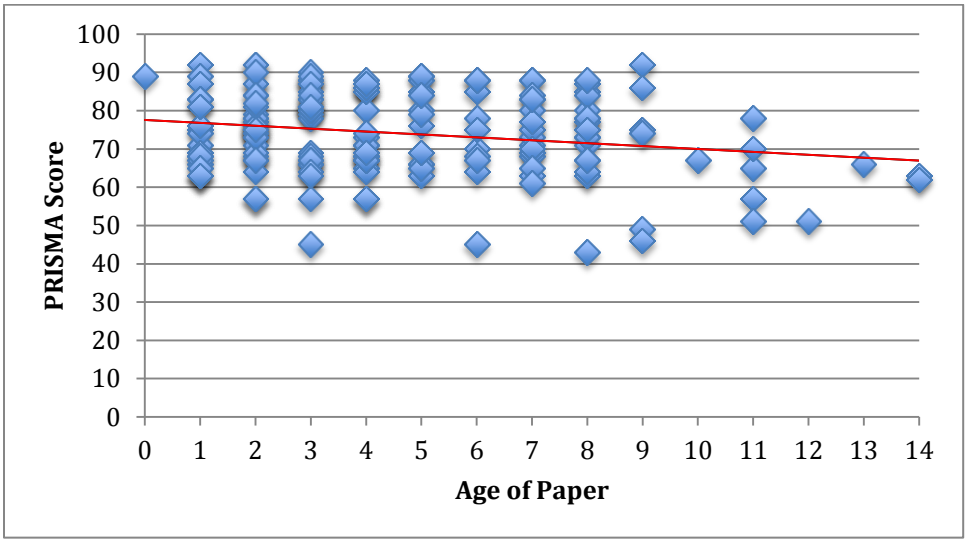


Figure 10 Scatter plot of age of paper against PRISMA score

The correlation coefficient was calculated at -0.153, which indicated that there was a weak negative linear relationship between age of paper and PRISMA score. The negative value would seem to indicate that a younger (i.e. newer) paper was more likely to score higher than an older paper.

This relationship was significantly different at p=0.040. However, giving the low value of the correlation coefficient itself, any improvement in score over time is likely to be very minor. (Table 16).

Table 16 Correlations for PRISMA score vs age of paper

	PRISMA Score	Age of Paper
PRISMA Score		
Pearson Correlation	1	-0.153
Sig. (2-tailed)		0.040
N	181	181
Age of Paper		
Pearson Correlation	-0.153	1
Sig. (2-tailed)	0.040	
N	181	181

Cochrane and non-Cochrane reviews were also assessed separately in order to determine if one of them exhibited greater improvements in reporting quality over time.

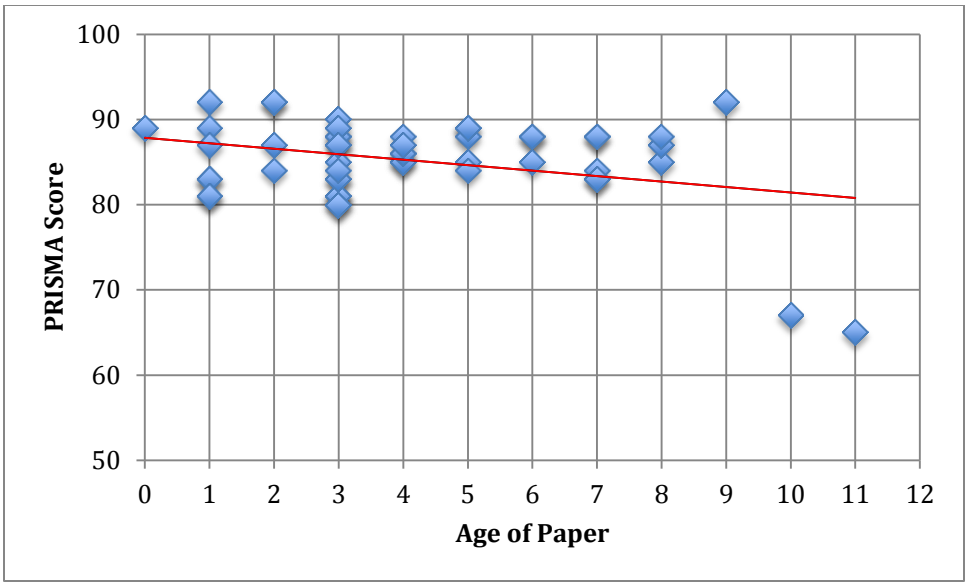


Figure 11 Scatter plot of age of cochrane review against PRISMA score

In this case the correlation coefficient was -0.324, which again indicated that there was a weak negative linear relationship between age of Cochrane paper and PRISMA score.

This relationship was statistically significant at p=0.019. (Table 17).

Table 17 Correlations for Cochrane reviews

	Cochrane Review Score	Age of Paper
Cochrane Review Score		
Pearson Correlation	1	-0.324
Sig. (2-tailed)		0.019
N	52	52
Age of Paper		
Pearson Correlation	-0.324	1
Sig. (2-tailed)	0.19	
N	52	52

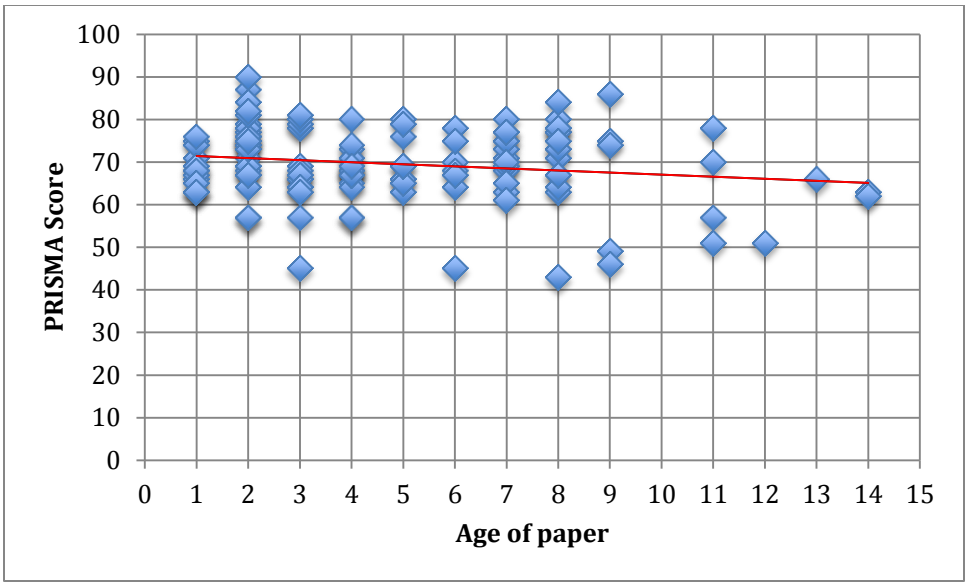


Figure 12 Scatter plot of age of paper against non-Cochrane PRISMA scores

The correlation coefficient for the above scatter plot was -0.071. This correlation however was not statistically significant, with p=0.422. (Table 18).

Table 18 Correlations for Non-Cochrane reviews

	Non-Coch. Review Score	Age of Paper
Non-Coch. Review Score		
Pearson Correlation	1	-0.071
Sig. (2-tailed)		0.422
N	129	129
Age of Paper		
Pearson Correlation	-0.071	1
Sig. (2-tailed)	0.422	
N	129	129

Multiple Linear Regression

A multiple linear regression model was used to determine the contribution of three predictors to the final PRISMA score.

These three predictors are:

1. The type of review i.e. Cochrane vs non-Cochrane
2. The speciality the paper is ascribed to
3. The age of the review

The regression analysis demonstrated that the three predictors contributed to just under 50% (0.465) of the variability in the final PRISMA score. (Table 15). The analysis also showed that the first two predictors displayed a statistically significant contribution to the final PRISMA score

whereas the final predictor, age of paper, did not (p=0.129). The p values for the type of review and the speciality of the paper were both p<0.0001. (Table 17).

Table 19 Regression model summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.689a	.474	.465	7.666

a. Predictors: (Constant), 1,2 and 3

Table 20 ANOVAb

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	9377.500	3	3125.833	53.196	p<0.0001
	Residual	10400.567	177	58.760		
	Total	19778.066	180			

a. Predictors: (Constant), 1,2 and 3

b. Dependent Variable: PRISMA score

Table 21 Coefficients a

Model	Unstandardised Coefficients		Standardised Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	104.183	2.644		39.408	0.000
Predictor1	-15.471	1.268	0.670	-12.202	0.000
Predictor2	-1.116	0.554	0.110	-2.014	0.046
Predictor3	-0.243	0.159	0.084	-1.527	0.129

a. Dependent Variable: PRISMA score

Discussion

Summary of the main findings

The mean overall score for the systematic reviews, when assessed using the QUOROM checklist, was 70.86% (95%CI 69.20%, 70.86%). This means that on average, a systematic review complied with approximately 45 out of the 63 items on the checklist.

The mean overall score for the PRISMA checklist was 74.07% (95%CI 72.53%, 75.61%), indicating that each systematic review complied with 47 out of a possible 63 items, on average. Therefore, systematic reviews tended to score higher with the PRISMA checklist, by complying with two extra items.

The highest score was 92% which was achieved by three orthodontic Cochrane reviews and one Cochrane review in periodontics. The lowest score achieved by any paper was for a non-Cochrane review in the speciality of preventive dentistry.

Cochrane reviews made up 29% of the assessed papers. The mean PRISMA score for these reviews was 85.19% (95%CI 83.79%, 86.59%), as opposed to 69.59% (95%CI 68.09%, 71.09%) for other systematic reviews and meta-analyses. This indicates that on average, Cochrane reviews scored 15.6% higher than other reviews and this difference was statistically significant ($p<0.00001$; mean difference 15.60; 95%CI 13.58, 17.62)..

Amongst the specialities, there were also differences in individual scores. The mean PRISMA score was 75.07% (95%CI 72.32%, 77.82%) for orthodontics, 74.91% (95%CI 72.80%, 77.03%) for periodontics, 71.50% (95%CI 67.22%, 75.78%) for preventive dentistry and 74.20% (95%CI 70.33%, 78.07) for endodontics. Therefore, the orthodontic literature scored the highest, when compared to the other three specialities however, this difference was not statistically significant

(p= 0.85). It could have been related to the fact that orthodontics had the highest number and proportion of Cochrane reviews, which were shown to score higher.

When the Cochrane reviews were assessed within each speciality, it was found that the speciality of periodontics scored the highest, with periodontic Cochrane reviews complying with 87.22% (95%CI 85.16%, 89.28%)of the PRISMA guidelines. Orthodontic Cochrane reviews complied with 85.35% (95% CI 82.01, 88.70%) of the guidelines, preventive dentistry with 85.23% (95% CI 83.37%, -87.10%) and endodontics with 83.00% (95% CI 81.15%, 84.85%) of the guidelines. Hence, periodontic papers on average obtained a score of 55 out of 63 items, orthodontic papers scored 53.7/63, preventive dentistry papers scored 53.6/63 and endodontic papers, which scored the lowest, scored on average 52/63. These differences were statistically significant (p value= 0.0196).

Non-Cochrane reviews were also compared between the specialities and once again periodontics scored the highest, with their systematic reviews complying with 72.6% (95% CI 70.77%, 74.44%) of the PRISMA guidelines. Orthodontics scored the second highest at 69.51% (95% CI 67.16%, 71.86%) , followed by endodontics at 68.33% (95% CI 64.26%, 72.41%) and preventive dentistry at 65.34% (95% CI 60.76%, 69.93%). In other words, periodontic systematic reviews complied with 45/63 items on average, orthodontic papers with 44/63 items, endodontic papers with 43/63 items and preventive dentistry, with 41/63 items on average. These differences were also statistically significant (p-value= 0.0235).

The compliance of papers over time was assessed using correlation coefficients, which compared the age of each paper with its PRISMA score. The correlation coefficient was -0.153, which indicated that there was a weak negative linear relationship between the compliance of reviews with the PRISMA guidelines and the year of publication of those reviews. This was statistically significant at p=0.040. The correlation coefficients were also calculated for Cochrane

reviews and non-Cochrane reviews separately. There was again a weak negative linear relationship between the age of a Cochrane review and its PRISMA score, with a correlation coefficient of -0.324 and this was statistically significant (p=0.019). . For non-Cochrane reviews however, the correlation coefficient was -0.071 which was not statistically significant (p=0.422). This indicates that only Cochrane reviews were improving over time.

A multiple linear regression model was used to determine the contribution of three predictors to the final PRISMA score. This showed that the type of review, the speciality of the paper and the age of the paper contributed to 50% of the variability of the final PRISMA score. The analysis also showed that the first two predictors displayed a statistically significant contribution to the final PRISMA score whereas the final predictor, age of paper, did not (p=0.129). The p values for the type of review and the speciality of the paper were both p=0.000.

Interpretation of the main findings

Comparison of checklists

The Bland and Altman test for agreement between clinical measurements was used to compare the modified PRISMA and the QUOROM checklists. Another option would have been to calculate a correlation coefficient but in this case, the Bland and Altman test was more suitable, because a correlation coefficient would not be able to measure bias between two tests, but instead could actually show a high level of agreement where two tools obtained consistently different results. The Bland and Altman test showed that there was statistically significant bias between the two tests, meaning that that the two measurement tools did not agree to a sufficient degree. However, visual inspection of the figure (figure 7) indicated that the bias, although significant, was for the most part consistent, with each paper obtaining a lower score when assessed using the QUOROM checklist and except for a few outliers, most of the QUOROM scores were

approximately 3-5% lower than their corresponding PRISMA scores. Accordingly, the two checklists could be used interchangeably with adjustments being made to allow for the lower scores obtained by the QUOROM checklist.

Main differences between the checklists

The original QUOROM guidelines constituted 13 criteria, which were modified for the purpose of this study to give a checklist of 63 items with a maximum possible score of 57. Meanwhile the PRISMA guidelines were more detailed and consisted of 27 criteria, which were further divided into 63 items to form the checklist constructed for this study. The main reason for breaking down the guidelines into a larger number of criteria was to enable a more accurate evaluation of each paper. Some of the original criteria were quite detailed with several requirements in each, which could potentially increase the likelihood of measurement error, so making sure each criterion looked at one specific factor within a paper ensured a more methodological and detailed appraisal and potentially allowed areas that were consistently under-reported to be identified.

As mentioned above, the first and main difference between the checklists is the number of items in each list. If the same paper was scored with each set of guidelines and the same number of points were awarded, the final score as a percentage would still be different as each set of points would be divided by a different denominator.

Another difference between the checklists was the actual criteria themselves. Some criteria in the QUOROM guidelines were omitted from the PRISMA statement, while some new criteria were introduced into the new statement that had not been included before. Also, the distribution of items between the checklists was different. For example, with regards to the

abstract, the QUOROM checklist looked at 20 criteria, whereas the PRISMA only looked at 11. This resulted in lower QUOROM scores for this section, as there was a higher chance of not meeting a criterion. The PRISMA guidelines had new criteria not mentioned in the QUOROM statement, such as systematic review registration number, protocol, registration and registration information. These generally did not achieve positive scores in most of the papers examined, which lowered the PRISMA score in that aspect.

Items that stood out in the QUOROM statement

The wording of several of the QUOROM guidelines did not permit the use of the option of "non-applicable" on the scoring sheet and hence resulted in a score of "zero" in many of the papers. This includes items such as item 20 (inclusion of subgroup analyses), item 44 (utilization of the principle measures of effect) and item 45 (method of combining results). If the wording of the items had included the phrase "where applicable" at the end, scoring could have potentially been more fair, for example where combining results or undergoing subgroup analyses was not feasible.

In the introduction section of the QUOROM checklist, item 15 required the description of the characteristics of the RCTs excluded in the results section. Although most papers met the criteria to score a "one" in item 14 (characteristic of RCTs included in the results), 63% of papers scored a "zero" on item 15. No similar item was found in the PRISMA checklist. Items 28 to 35 looked at the inclusion and exclusion criteria. Whereas 89% of the papers scored a "one" for the inclusion criteria, only 60% of the papers described the exclusion criteria in sufficient detail to score a point.

Finally, criterion 62, which required the discussion of potential biases in the review process, was rarely met, with only 33% of papers adequately discussing this point.

Items that stood out in the PRISMA statement

As mentioned previously, the PRISMA guidelines had 12 items related to the title and the abstract, compared with 21 items for the QUOROM guidelines. The more succinct division of items in the PRISMA statement allowed papers to score higher in the initial items than they did in the title and abstract section of the QUOROM guidelines. The only item that scored significantly lower in this section was item 12 (systematic review registration number), with almost no papers achieving a positive score.

The development of the PRISMA guidelines

The QUOROM guidelines were published in 1999[5] but since then a lot has changed. First, there has been a great increase in the knowledge about the conduct and reporting of systematic reviews. For example, the Cochrane Library's Methodology Register increased to over 11,000 entries as of March, 2009.[50] Secondly, authors have increasingly used systematic reviews to summarise the available evidence from randomised trials. However, despite these advances, the quality of reporting of systematic reviews has been below the required standard. This led to a panel concluding that the QUOROM statement was flawed and needed to be updated and expanded, so they developed the PRISMA statement.

During the process of data collection in the present study, it was found that the PRISMA statement was more easy to follow, as the guidelines were stated more explicitly, while the phrasing of some of the QUOROM guidelines was more ambiguous. This made scoring of the papers with the PRISMA checklist more straightforward and was preferred by the author.

Cochrane versus non-Cochrane reviews

Cochrane reviews scored significantly higher than non-Cochrane reviews and their mean score was also found to improve with time, whereas non-Cochrane reviews did not. This is likely to be related to the robust editorial process involved in writing a Cochrane review.

Before any Cochrane review is written, potential reviewers must first propose a title, which is negotiated with the Assistant Managing Editor.[51] This is followed by submitting a registration form. Once the title has been accepted, the authors are then encouraged to attend a workshop on protocol development, and will receive a protocol template. The final protocol draft is then checked by the Cochrane Collaboration's Information and Management System in RevMan format, after careful proof-reading. The protocol itself then undergoes both an internal and external refereeing process. It is checked by the Editorial Base, which then sends suggestions for improvement back to the authors if required. The revised draft is checked after the implementations have been done and then is transferred to the group's editors and external referees. They then make comments and once these have been satisfactorily addressed, the final protocol is proof read and approved by the Co-ordinating Editor and submitted for inclusion in *The Cochrane Library* to the publisher.

After the protocol has been accepted, the authors receive a review template and proceed with preparing the actual review. This template ensures that all reviews are standardised and all-inclusive. The final review draft is proof-read and sent to the Editorial Base. The Cochrane review then undergoes a similar refereeing process to the one conducted on the protocol. The final review is then proof read, copy-edited and approved by the Co-ordinating Editor and submitted for inclusion in *The Cochrane Library* to the publisher. There's also the Cochrane

Handbook and style guidelines that set out how each stage should be carried out and there are strict quality control guidelines.

Cochrane reviews are also regularly updated, where authors are responsible for searching through the literature from the date of the last search to date to identify any recently published trials related to their review and to update their review accordingly. This ensures that any new trials, that might alter the results of the existing Cochrane review, are included to provide the most up-to-date evidence. For these reasons, Cochrane reviews generally of a more superior quality to other reviews, which are unlikely to undergo such a thorough process of preparation.

What do the findings mean?

- The compliance of systematic reviews and meta-analyses with the PRISMA guidelines was variable.
- There were no significant differences in the quality of systematic reviews between the four dental specialities.
- Cochrane reviews scored significantly higher than non-Cochrane reviews.
- The standard of reporting of Cochrane reviews has improved over time, whereas this is not the case with non-Cochrane reviews.

Comparisons with other studies

Due to the recent publication of the PRISMA guidelines, there has only been one study assessing the quality of systematic reviews using this checklist[48] and this was in the Chinese medicine literature. This paper however did not calculate an overall PRISMA score for any of the papers so

the results cannot be compared directly. On the other hand, several studies investigated the quality of reviews using the QUOROM checklist, as shown in the literature review.

Systematic reviews of traditional Chinese medicine scored very poorly and did not follow the QUOROM guidelines, rendering their results inconclusive.[28] This finding is in contrast with this study, which showed that although some studies had low scores, others scored very highly, with the highest score being 92%.

Systematic reviews in the Cochrane Neonatal Review Group were shown to have improved after the publication of the QUOROM guidelines.[29] This is similar to the findings in this study, although in this case improvement was assessed over time in general, rather than before and after the QUOROM guidelines, *per se*. No overall score was calculated for the Neonatal Review Group Cochrane reviews so this aspect of the results could not be compared.

Systematic reviews of pneumonia in China were found to score poorly, with the highest score being 10 out of 18 items.[30] As a percentage this can be expressed as 55.56%, which is significantly lower than the 70.86% average found in this study. The difference should be interpreted with caution however, as a modified 63-item checklist was used in this study, versus the 18-item checklist used in the Chinese study.

The oncology literature showed that 70% of reviews were not systematic and 21% of the studies did not adequately describe their search methods.[31] In this study, 83% percent of the 181 papers adequately described the search strategy, which is similar to the findings in the oncology literature in that aspect. However, whereas they found that most papers were of poor quality, this study found that papers scored highly in other items of the QUOROM checklist, which could potentially have made up for other low scores.

A review of meta-analyses dealing with pharmacotherapy of post-traumatic stress disorder[32] concluded that the quality of meta-analyses was acceptable in the PTSD literature, but looking

at the actual average score, papers complied with 59.3% of the QUOROM guidelines, which is 11.56% lower than the results found in this study.

Systematic reviews and meta-analyses of anxiety disorders scored on average 62%, which is 8.86% lower than our findings.[33] The current study found that the lowest scores were in the results section of the papers whereas the highest scores were in the introduction and discussion sections of the reviews. Although these aspects were not compared in the present study, a preliminary assessment of the scoring spreadsheet shows that the 'Abstract' section scored the lowest in most papers and the 'Introduction' scored the highest.

Health Technology Assessments' (HTA) compliance with the QUOROM guidelines was variable.[34] Out of 87 papers, 49% of all systematic reviews used a study selection flow diagram. This is similar to the systematic reviews looked at in the dental literature, with 45% of papers showing a flow diagram in the results section.

One study, looking at both Cochrane and non-Cochrane reviews in the medical literature, demonstrated an increase in score over time, with the score changing from 10.5 out 18 in 2000 to 13.0 in 2005.[35] Again, this is in agreement with the findings of this study, which showed that Cochrane reviews improved with time and that the change was statistically significant. When analysing the Cochrane reviews in detail, the authors found that they only superseded other reviews in the abstract section. However, in the present study, this was not the case.

One study was identified, which assessed the quality of reporting systematic reviews and meta-analyses in the orthodontic literature.[38] Between 1966 and 2002, 13 papers were identified whereas in the present study, 56 papers were identified, because the search was more recent and identified papers written until 2009. Papadopoulos and Gkiauris assessed 16 orthodontic papers that met their inclusion criteria.[39] Each paper was appraised individually according to

where the authors thought their deficiencies lay but there was no assessment of the papers using specific guidelines and as such, their findings cannot directly be compared to the current study.

Limitations of the study

The main limitation of this study is that it was retrospective in nature and hence could potentially be subject to bias, due to some reviews not being included in the sample. In order to avoid this, the sample was obtained from the Cochrane Library database of systematic reviews. The database has been compiled by a combination of electronic searching of several other databases and hand searching relevant journals by trained hand searchers, so theoretically, all true systematic reviews and meta-analyses should have been identified. Reviews not included in the Cochrane Library are unlikely to be systematic, such as literature reviews for example.

Another limitation could be related to the modified checklists used in this study. The original PRISMA and QUOROM checklists consist of lengthy statements detailing different aspects required of reviews. In other words, each statement has more than one criterion within it, so there would have potentially been some ambiguity during scoring if only have a statement was met. In order to overcome this problem and simplify scoring, each statement was subdivided into several items, where each item involved only one criterion. This made the process of scoring the papers easier and more consistent. The wording of the items in the modified checklists remained as close as possible to the original guidelines in order to reduce the likelihood of bias. Good intra-examiner and inter-examiner agreement was achieved with the developed checklists, so again, bias was unlikely.

The wording of items was another limitation of the study. For example, items 16 and 17 in the PRISMA checklist state: protocol information “if a protocol exists”. In view of this, papers that

did not have a protocol scored a “N/A” in these two items, whereas papers with a protocol and no elaboration scored a “zero.” As most non-Cochrane reviews with a protocol did not have sufficient protocol information, the wording of the items resulted in these papers losing more points than papers with no protocol at all. Several other items in the PRISMA checklist were worded as “if applicable,” resulting in many “N/As” when a criterion was not met, whereas this phrase was rarely used in the QUOROM checklist, resulting in more scores of “zero”.

As the scoring was conducted manually, there was also the element of human error, which could have led to bias. One potential for error was examiner fatigue, so in order to reduce this risk, only three papers were scored at any one time. Also, the data were entered directly into a Microsoft excel data spreadsheet to prevent any error arising from transferring data from a sheet of paper to a computer spreadsheet. Entering data into a spreadsheet also reduced the likelihood of error when calculating the scores, as this could be done electronically. Again, there was good intra-examiner and inter-examiner in the pilot studies, so even if there was any bias, it was not significant.

The original intent of the study was to assess systematic reviews in the orthodontic literature. A sample size calculation showed that 165 papers were required to achieve 80% power with a 0.05 two-sided significance level, which is much greater number than the number of systematic reviews in orthodontics. For this reason, three other specialities were included in the study to achieve an adequate sample size. Another original intent was to compare papers published within three time periods: the five years preceding the QUOROM statement, the five years immediately after the statement was published and the last five years. This however proved not to be feasible due to the large discrepancies in the sample sizes between the time periods. For example, in the orthodontic literature, only three systematic reviews were identified in the

Cochrane Library for the first time period, which was not sufficient to make a comparison. For this reason, it was decided to test for a correlation between age of paper and PRISMA score, to see if there had been any improvement with score over time. Calculating a correlation coefficient might not have been the ideal way to test for improvement over time, but with the distribution of the papers (most papers were recent and published in the last three years), this was probably the most suitable method in this particular case.

Finally, the publication of the PRISMA guidelines during data collection for the QUOROM guidelines necessitated the modification of the study aims and methods. A modified PRISMA checklist was compiled and the level of agreement between it and the modified QUOROM checklist that was already in use for data collection, was assessed using the Bland and Altman test. This showed that although there was a significant difference between the two tools, the PRISMA scores were consistently higher than the QUOROM scores by approximately the same amount. As the PRISMA guidelines are more recent, they were subsequently used for the majority of the study. However, when comparing the results of this study with the results of other studies, the data obtained from the QUOROM checklist were used, rather than the PRISMA checklist. This is because as of yet, there has only been paper[48] looking at the compliance of systematic reviews and meta-analyses with the PRISMA guidelines, which was in the Chinese literature, so measuring external validity was difficult.

Implications for practice

The results of this study show that the standard of reporting systematic reviews is largely variable, with Cochrane reviews scoring significantly higher than non-Cochrane reviews.

Although Cochrane reviews have been shown to be improving with time, there is still a large discrepancy between other systematic reviews published within similar time periods. Future

reviews should attempt to follow the PRISMA guidelines more closely in order to improve their reporting quality.

More than 90% of systematic reviews did not include systematic review registration numbers and most of non-Cochrane reviews did not include a protocol, both of which are PRISMA guideline requirements. These should be included in future reviews and Cochrane updates.

Implications for research

Two modified checklists from the QUOROM and PRISMA checklists have been developed for this study which have been shown to be valid and reliable and could be used in further studies.

Only one other study so far has looked at compliance of systematic reviews with the PRISMA guidelines, most likely because the guidelines are relatively new. Future studies can assess the compliance of reviews with the PRISMA guidelines, in other parts of the literature.

This study looked at overall PRISMA scores for papers, but did not specifically look at the percentage of papers achieving a score of “1” on specific items. Further research could look at the items individually, in order to establish which items scored the highest and which aspects of reviews scored the lowest.

Conclusions

- The compliance of systematic reviews and meta-analyses with the PRISMA guidelines was variable with the mean score being 74.07%.
- There was a weak negative linear relationship between the compliance of Cochrane reviews with the PRISMA guidelines and the year of publication of those reviews. This was statistically significant at $p=0.019$ so the null hypothesis can be rejected.
- There was a statistically significant difference in the compliance of systematic reviews and meta-analyses with the QUOROM statement and the PRISMA statement ($p<0.0001$). Scoring with the PRISMA checklist resulted in consistently higher scores than when utilising the QUOROM checklist. Therefore the null hypothesis can be rejected.
- There was a statistically significant difference in the compliance of Cochrane reviews and non-Cochrane systematic reviews with the guidelines of the PRISMA statement ($p<0.00001$). Again, the null hypothesis can be rejected.
- There was no statistically significant difference between the specialities in the compliance with the PRISMA guidelines, so the null hypothesis cannot be rejected at $p>0.05$.
- The three independent variables were found to explain approximately 50% of the variability of the final PRISMA score.

References

1. Eysenck, HJ. Meta-analysis and its problems. Br Med J 1994; 309:789-792.
2. Egger M, Smith GD. Meta-analysis: potentials and promise. Br Med J 1997; 315:1371-1374.
3. www.cebm.net/levels_of_evidence.asp [updated May 2001; accessed 30th October 2009].
4. www.prescribingforbetteroutcomes.org. [updated 2008; accessed 30th October 2009].
5. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999; 354: 1896-1900.
6. Delaney A, Bagshaw SM, Ferland A, Manns B, Laupland KB, Dolg CJ. A systematic evaluation of the quality of meta-analyses in the critical care literature. Critical Care 2005; 9:R575-582.
7. Liu JP, Xia Y. Quality appraisal of systematic reviews or meta-analysis on traditional Chinese medicine published in Chinese Journals. Zhongguo Zhong Xi Yi Jie He Za Zhi 2007; 27:306-311.
8. Al-Faleh K, Al-Omran M. Reporting and methodological quality of Cochrane Neonatal Review Group systematic reviews. BMC Pediatrics 2009; 17:9-38.
9. Hou ZK, Li JS, YuXQ, Li B, Zhou HY, Zhang YX. Quality appraisal of systematic reviews and meta-analysis of pneumonia in China. Zhongguo Wei Zhong Bing Ji Jiu Yi Xue 2009; 21(4): 207-210.
10. Vigna-Taglianti F, Vineis P, Liberati A, Faggiano F. Quality of systematic reviews used in guidelines for oncology practice. Ann Oncol 2006; 17(4): 691-701.
11. Stein DJ, Ipser J, McAnda N. Pharmacotherapy of post-traumatic stress disorder: a

review of meta-analyses and treatment guidelines. CNS Spectr 2009; 14: 25-31.

12. Bereza BG, Machado M, Einarson TR. Assessing the reporting and scientific quality of meta-analyses of randomized controlled trials of treatments for anxiety disorders. Ann Pharmacother 2008; 42(10): 1402-1409.
13. Hind D, Booth A. Do health technology assessments comply with QUOROM diagram guidance? An empirical study. BMC Med Res Methodol 2007; 7: 49.
14. Wen J, Ren Y, Wang L, Li Y, Liu Y, Zhou M, et al. The reporting quality of meta-analyses improves: a random sampling study. J Clin Epidemiol 2002; 61(8): 770-775.
15. Shea B, Boers M, Grimshaw JM, Hamel C, Bouter LM. Does updating improve the methodological and reporting quality of systematic reviews? BMC Med Res Methodol 2006; 6:27.
16. Shea, B, Moher D, Graham I, Pham B, Tugwell P. A comparison of the quality of Cochrane reviews and systematic reviews published in paper-based journals. Eval Health Prof 2002; 25: 116-129.
17. Greenhalgh T. Papers that summarise other papers (systematic reviews and meta-analyses). Br Med J 1997; 315: 672-675.
18. Aronson JK. Meta-analysis. Br J Clin Pharmacol 2005; 60(2): 117-119.
19. Bero L, Rennie D. The Cochrane Collaboration. Preparing, maintaining and disseminating systematic reviews of the effects of health care. JAMA 1995; 274:1935-1938.
20. Clarke M, Langhorne P. Revisiting the Cochrane Collaboration. Br Med J 2001; 323:821.
21. Egger M, Smith GD, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. 2nd ed. London: BMJ Publishing group; 2001.
22. Oxman, AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991; 44(11): 1271-1278.

23. www.phru.nhs.uk/pages/PHD/CASP.htm [updated 2007; accessed 29th October 2009].
24. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. JAMA 1994; 272(17):1367-1371.
25. Del Mar CB, Glasziou PP. Antibiotics for sore throat. [Cochrane review] In: The Cochrane Library Issue, 3, 1999. Oxford: Update Software.
26. ssrc.tums.ac.ir/SystematicReview/Glasgow.asp [updated 2005; accessed 20th October 2009].
27. www.arif.bham.ac.uk/ [updated 2008; accessed 23rd October 2009].
28. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials: The CONSORT statement. JAMA 1996; 276(8):637-639.
29. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. Ann Intern Med 1987; 107(2): 224-233.
30. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med 1987; 316:450–455.
31. Sacks HS, Reitman D, Pagano D, Kupelnick B. Meta-analysis: an update. Mt Sinai J Med 1996; 63:216–224.
32. Dixon E, Hameed M, Sutherland F, Cook DJ, Doig C. Evaluating meta-analyses in the general surgical literature: a critical appraisal. Ann Surg 2005; 241(3): 450-459.
33. Moher D, Tricco AC. Issues related to the conduct of systematic reviews: A focus on the nutrition field. Am J Clin Nutr 2008; 88:1191-1999.
34. Olsen O, Middleton P, Ezzo J, Gotzsche PC, Hadhazy V, Herxhelmer A, et al. Quality of Cochrane reviews: assessment of sample from 1998. Br Med J 2001; 323:829-832.
35. McElvenny DM, Armstrong BG, Jarup L. Higgins JP. Meta-analysis in occupational

epidemiology: a review of practice. Occup Med (Lond) 2004; 54(5): 336-344.

36. Lyman GH, Kuderer NM. The strengths and limitations of meta-analyses based on aggregate data. BMC Med Res Methodol 2005; 5(1): 14.
37. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. JAMA 1998; 280(3): 278-280.
38. Papadopoulos MA. Meta-analysis in evidence-based orthodontics. Orthod Craniofac Res 2003;6:112-26.
39. Papadopoulos MA, Gkiaouris I. A critical evaluation of meta-analyses in orthodontics. Amer J Orthod and Dentofacial Orthop 2007;131: 589-599.
40. Kim JH, Viana MA, Graber TM, Omerza FF, BeGole EA. The effectiveness of protraction face mask therapy: a meta-analysis. Am J Orthod Dentofacial Orthop 1999;115:675-685.
41. Jager A, Braumann B, Kim C, Wahner S. Skeletal and dental effects of maxillary protraction in patients with Angle Class III malocclusion: A meta-analysis. J Orofac Orthop 2001;62: 275-284.
42. Harrison J, Ashby D. Orthodontic treatments for posterior crossbites [Cochrane Review]. The Cochrane Library Issue, 4, 2001. Oxford: Update Software.
43. Trpkova B, Major P, Prasad N, Nebbe B. Cephalometric land- marks identification and reproducibility: a meta analysis. Am J Orthod Dentofacial Orthop 1997;112:165-170.
44. Nguyen QV, Bezemer PD, Habets L, Prahl-Andersen B. A systematic review of the relationship between overjet size and traumatic dental injuries. Eur J Orthod 1999;21:503-515.
45. Segal GR, Schiffman PH, Tuncay OC. Meta-analysis of the treatment-related factors of external apical root resorption. Orthod Craniofac Res 2004;7:71-78.

46. Polder BJ, Van't Hof MA, Van der Linden FP, Kuijpers-Jagtman AM. A meta-analysis of the prevalence of dental agenesis of permanent teeth. Community Dent Oral Epidemiol 2004;32: 217-226.

47. Cohen J. A coefficient of agreement for nominal scales. Educ and Psychol Meas 1960; 20: 37.

48. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. Ann Intern Med 2009, 151(4); 1-4.

49. Li CJ, Lü J, Su NC, Li S, Shi ZD. Preferred reporting items for systematic reviews and meta-analysis for reporting quality of Chinese meta-analysis on stomatology. Chin Journ Ostomat 2011 ;46(5):257-62.

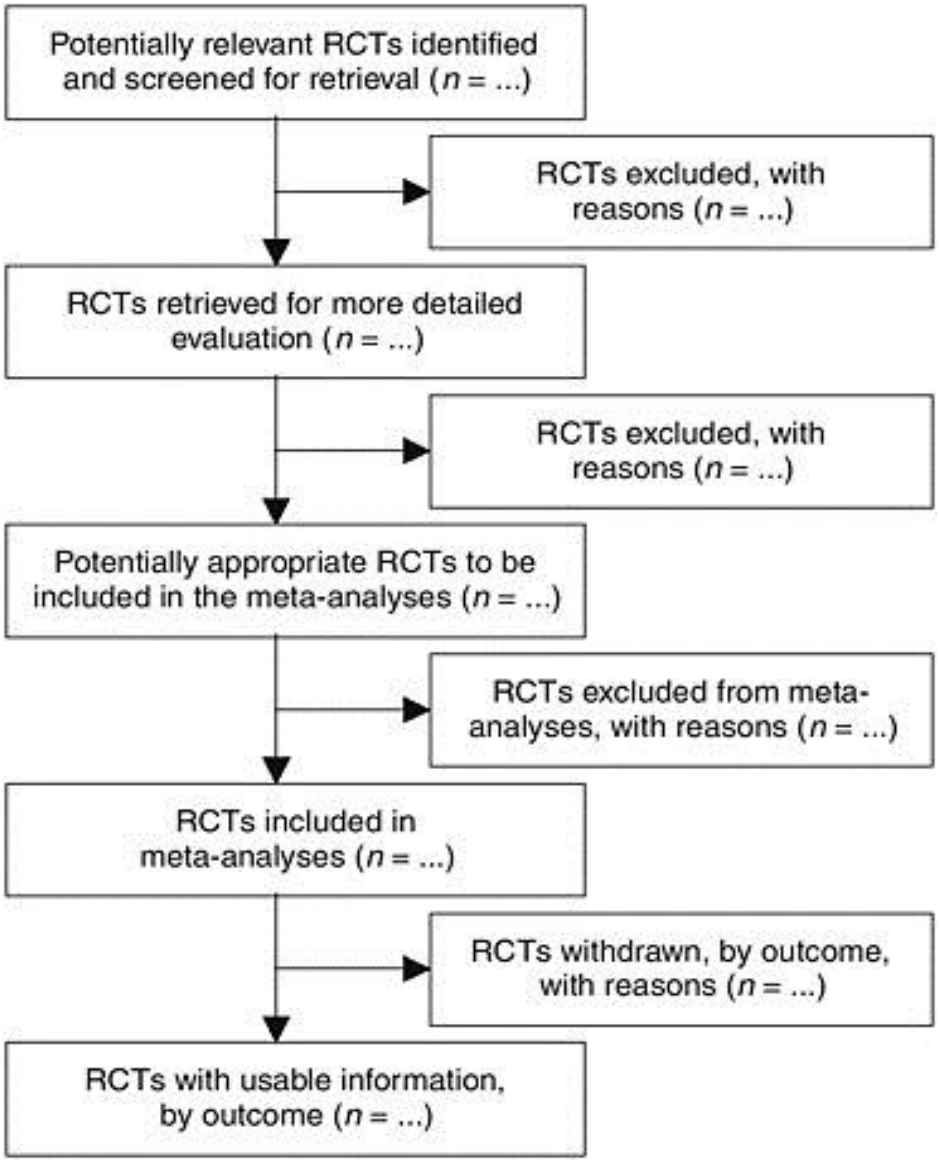
APPENDIX A

Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement checklist

Heading	Subheading	Descriptor	Reported? (Y/N)	Page number
Title		Identify the report as a meta-analysis [or systematic review] of RCTs[24]		
Abstract		Use a structured format[27]		
		Describe		
	Objectives	The clinical question explicitly		
	Data sources	The databases (ie, list) and other information sources		
	Review methods	The selection criteria (ie, population, intervention, outcome, and study design); methods for validity assessment, data abstraction, and study characteristics, and quantitative data synthesis in sufficient detail to permit replication		
	Results	Characteristics of the RCTs included and excluded; qualitative and quantitative findings (ie, point estimates and confidence intervals); and subgroup analyses		
	Conclusion	The main results		
		Describe		
Introduction		The explicit clinical problem, biological rationale for the intervention, and rationale for review		
Methods	Searching	The information sources, in detail[28] (eg, databases, registers, personal files, expert informants, agencies, hand-searching), and any restrictions (years considered, publication status,[29] language of publication[30,31])		
	Selection	The inclusion and exclusion criteria (defining population, intervention, principal outcomes, and study design[32]		
	Validity assessment	The criteria and process used (eg, masked conditions, quality assessment, and their findings[33–35])		
	Data abstraction	The process or processes used (eg, completed independently, in duplicate)[35,36]		
	Study characteristics	The type of study design, participants' characteristics, details of intervention, outcome definitions, &c,[37] and how clinical heterogeneity was assessed		
	Quantitative data synthesis	The principal measures of effect (eg, relative risk), method of combining results (statistical testing and confidence intervals), handling of missing data; how statistical heterogeneity was assessed;[38] a rationale for any a-priori sensitivity and subgroup analyses; and any assessment of publication bias[39]		
Results	Trial flow	Provide a meta-analysis profile summarising trial flow (see figure)		
	Study characteristics	Present descriptive data for each trial (eg, age, sample size, intervention, dose, duration, follow-up period)		
	Quantitative data synthesis	Report agreement on the selection and validity assessment; present simple summary results (for each treatment group in each trial, for each primary outcome); present data needed to calculate effect sizes and confidence intervals in intention-to-treat analyses (eg 2×2 tables of counts, means and SDs, proportions)		
Discussion		Summarise key findings; discuss clinical inferences based on internal and external validity; interpret the results in light of the totality of available evidence; describe potential biases in the review process (eg, publication bias); and suggest a future research agenda		

Quality of reporting of meta-analyses

Flow Diagram



APPENDIX B

63- item QUOROM Checklist

Section	Item No.	Description
Title	1	Identification of the report as a meta-analysis or systematic review of RCT's
Abstract	2	Use of a structural format
	3	Description of the clinical question explicitly in the objectives
	4	Description of the databases used
	5	Description of other information sources used
	6	The selection criteria: population described
	7	The selection criteria: intervention described
	8	The selection criteria: outcome described
	9	The selection criteria:study design described
	10	Description of the methods for validity assessment
	11	Description of methods for data abstraction
	12	Description of the study characteristics
	13	Description of the quantitative data synthesis in sufficient detail to permit replication

	14	Description of the characteristics of the RCTs included in the results section
	15	Description of the characteristics of the RCTs excluded in the results section
	16	Description of the qualitative findings
	17	Description of the quantitative findings
	18	Inclusion of point estimates
	19	Inclusion of confidence intervals
	20	Inclusion of subgroup analyses
	21	Description of the main results in the conclusion
Introduction	22	Description of the explicit clinical problem
	23	Description of the biological rationale
	24	Description of the intervention
	25	Description of the rationale for review
Methods	26	Description of the information sources in detail
	27	Description of any restrictions in the searching process
	28	Description of the inclusion criteria: defining population
	29	Description of the inclusion criteria: intervention
	30	Description of the inclusion criteria: principal outcomes
	31	Description of the inclusion criteria: study design

	32	Description of the exclusion criteria: defining population
	33	Description of the exclusion criteria: intervention
	34	Description of the exclusion criteria: principal outcomes
	35	Description of the exclusion criteria: study design
	36	Description of the criteria used in validity assessment
	37	Description of the processes used in validity assessment
	38	Description of the processes used in data abstraction
	39	Description of the type of study design
	40	Description of the participants' characteristics
	41	Description of the details of intervention
	42	Outcome definitions
	43	Method of assessing clinical heterogeneity
	44	Utilisation of the principal measures of effect
	45	Method of combining results: statistical testing and confidence intervals
	46	Handling of missing data
	47	How statistical heterogeneity was assessed
	48	A rationale for any a-priori sensitivity and subgroup analyses
	49	Assessment of publication bias

Results	50	Provision of a meta-analysis profile summarising trial flow
	51	Provision of descriptive data for each trial: age
	52	Provision of descriptive data for each trial: sample size
	53	Provision of descriptive data for each trial: intervention
	54	Provision of descriptive data for each trial: dose
	55	Provision of descriptive data for each trial: duration
	56	Quantitative data synthesis: Reporting agreement on the selection and validity assessment
	57	Presentation of a simple summary of results
	58	Presentation of data needed to calculate the effect sizes and confidence intervals
Discussion	59	Summary of key findings
	60	Discussion of clinical inferences based on internal and external validity
	61	Interpretation of the results in light of the totality of available evidence
	62	Description of potential biases in the review process
	63	Suggestion of a future research agenda

APPENDIX C

63-item PRISMA checklist with guidance notes

Section	Item	Description	Comments
Title	1	Identify the report as a systematic review, meta-analysis, or both.	
Abstract	2	Provision of a structured summary	
	3	Background	
	4	Objectives	
	5	Data sources	
	6	Study eligibility criteria	
	7	Participants	
	8	Interventions	
	9	Study appraisal and synthesis methods	
	10	Results and limitations	
	11	Conclusions and implications of key findings	
	12	Systematic review registration number	

Introduction			
Rationale	13	Description of the rationale for the review in the context of what is already known	
Objectives	14	Provision of an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3 out of 5 for PICOS
Methods			
Protocol and	15	Indication if a review protocol exists	
Registration	16	If a protocol does exist, indication if and where it can be accessed	
	17	Provision of registration information including registration number if a protocol is available	
Eligibility criteria	18-26	Specification of study characteristics: participants Specification of study characteristics: interventions Specification of study characteristics: comparisons Specification of study characteristics: outcomes Specification of study characteristics: study design Specification of study characteristics: length of follow-up Specification of report characteristics: years considered Specification of report characteristics: language Specification of report characteristics: publication status	
Information sources	27	Description all information sources	
	28	Mention of date last searched	
Search	29	Presentation of full electronic search strategy for at least one database, including any limits used, such that it could be repeated	

Study selection	30	Statement of the process for selecting studies (i.e., screening, eligibility, included in systematic reviews	
	31	If applicable, statement of the processes for selecting studies included in the meta-analysis	
Data collection	32	Description of the method of data extraction from reports (e.g., piloted forms, independently, in duplicate	
	33	Description of any processes for obtaining and confirming data from investigators	
Data items	34	Listing and definition of all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made	At least three of PICOS must be mentioned
Risk of bias in individual studies	35	Description of methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level)	
	36	Description of how this information is to be used in any data synthesis	
Summary measures	37	Statement of the principal summary measures (e.g., risk ratio, difference in means)	
Synthesis of results	38	Description of the methods of handling data and combining results of studies, if done, including measures of consistency for each meta-analysis	
Risk of bias across studies	39	Specification of any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies)	
Additional analyses	40	Description of methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression) if done	
Results			
Study	41	Giving numbers of studies screened	
selection	42	Giving numbers of studies assessed for eligibility and included in the review with reasons for exclusions at each stage	

	43	Flow diagram of studies screened and included	
	44	Presentation of characteristics for which data were extracted with regards to study size	
Study characteristics	45	Presentation of characteristics for which data were extracted with regards to participants	
	46	Presentation of characteristics for which data were extracted with regards to interventions	
	47	Presentation of characteristics for which data were extracted with regards to comparisons	
	48	Presentation of characteristics for which data were extracted with regards to outcomes	
	49	Presentation of characteristics for which data were extracted with regards study design	
	50	Presentation of characteristics for which data were extracted with regards to follow up period	
	51	Presentation of characteristics for which data were extracted with regards to citations	
Risk of bias within studies	52	Presentation of data on risk of bias of each study and, if available, any outcome level assessment	
Results of individual studies	53	For all outcomes considered (benefits or harms), presentation, for each study, a simple summary data for each intervention group	
	54	For all outcomes considered (benefits or harms), presentation, for each study, effect estimates and confidence intervals, ideally with a forest plot.	
Synthesis of results	55	Presentation of results of each meta-analysis done, including confidence intervals and measures of consistency	
Risk of bias across studies	56	Presentation of results of any assessment of risk of bias across studies	

Additional analysis	57	Giving the results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression	
Discussion			
Summary of evidence	58	Summary of the main findings including the strength of evidence for each main outcome	
	59	Consideration of their relevance to key groups (e.g., healthcare providers, users, and policy makers)	
Limitations	60	Discussion of limitations at study and outcome level (e.g., risk of bias)	
	61	Discussion of limitations at review-level (e.g., incomplete retrieval of identified research, reporting bias)	
Conclusions	62	Provision of a general interpretation of the results in the context of other evidence	
	63	Implications for future research	
Funding			
	64	Description of the sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review	

Appendix D

PRISMA CHECKLIST

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	

Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	

PRISMA 2009 Flow Diagram