Introduction

Meta-analysis (MA) is a set of statistical techniques to combine results from two or more separate studies (Higgins et al., 2011), while a systematic review consists of a clearly formulated question and explicit methods to identify, select, and critically appraise relevant research. Although the term “meta-analysis” is used interchangeably with “systematic review”, strictly speaking, an MA is an (optional) component of a systematic review. Ideally articles summing evidence by means of MA should always be conducted within the framework of a proper systematic review, in order to identify and minimize the many sources of bias found, like reporting biases. However, this is not always the case (Chavalarias and Ioannidis, 2010; Polychronopoulou et al., 2010).

The evaluation of the reporting quality of published MAs is very useful, as it is directly related to the study’s methodology and conclusions (Huwiler-Müntener et al., 2002; Moher et al., 1998). Attempts have been made over the years to assess the quality of systematic reviews and MAs by creating several relevant instruments for reporting and appraising them with varying efficacy (Sampson et al., 2008a). The QUOROM (Quality of Reporting of Meta-analyses; Moher et al., 1999) and subsequently the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) statement (Liberati et al., 2009) incorporate items deemed important for the transparent reporting of an MA. In that direction, the AMSTAR (Assessment of Multiple Systematic Reviews; Shea et al., 2007b) tool was introduced, designed to critically appraise the methodological quality of systematic reviews, and has proved to be reliable and valid (Shea et al., 2007a, 2009).

Existing biases in systematic reviews and MAs may be examined in complex systematic reviews (Whitlock et al., 2008) or “meta-epidemiological” studies, in which the influence of specific characteristics on treatment effect estimates is explored (Sterne et al., 2002). The characteristics of orthodontic systematic reviews (without meta-analytical synthesis) as well as their shortcomings have been previously highlighted by the authors of the current article (Papageorgiou et al., 2011). These may originate from the suboptimal reporting quality of the reviews, from characteristics of the orthodontic literature itself associated with surrogate endpoints (Richards, 2005), or from studies of low power and reliability (Tulloch et al., 1989; Rinchuse et al., 2008). Indeed, MAs in orthodontics
seem to present low reporting quality and include few high-quality primary studies (Papadopoulos and Gkiaouris, 2007). The methodology and reporting of orthodontic MAs have not been studied extensively so far.

The aims of this study were (a) to evaluate the methodology and quality characteristics of MAs (systematic review articles with meta-analytical data synthesis) related to orthodontics, (b) to assess whether their reporting quality has been improved over the years, and (c) to evaluate the appropriateness of the statistical methods being implemented in these MAs.

Materials and methods

Study sample

Electronic search strategies were developed and executed to identify MAs relevant to orthodontics published in journals, dissertations, or conference proceedings (Supplementary Tables S1a and S1b). No restrictions were made concerning year, language, or publication status. The reference lists of acquired articles were also searched for relevant articles. Databases of research registers were searched to identify ongoing or unpublished reviews. All databases were searched on 4 March 2011 and were manually rechecked on July 2011. Relevant authors identified previously to have published systematic reviews were contacted for additional articles (Papageorgiou et al., 2011). Two of the authors (S.N.P. and M.A.P.) screened independently the titles and abstracts of the retrieved citations to exclude non-eligible articles. A copy of the full text was obtained for the remaining articles. The same authors read each full-text article in order to determine whether it met the inclusion criteria. Additional material that was included as appendix in the original articles was acquired, when needed.

In this paper, the term “meta-analysis” was used for articles summing evidence from two or more separate sources in a single estimate. This must be distinguished from the statistical procedure of MA, since an article may report the results of more than one MA procedures. An article was considered as eligible for inclusion if it reviewed the literature appropriately in order to find relevant articles and combined data from two or more separate studies into a single estimate (number). These articles could be labelled as “meta-analyses” or “systematic reviews”, although the latter were included only if they included a combined estimate. All other types of studies, such as case reports, case series without a control, case-control, cohort studies, other observational studies, randomized controlled clinical trials (RCTs), narrative reviews, as well as systematic reviews without meta-analytic procedures were excluded.

Data extraction

Data to be collected were defined a priori from pilot searching of the literature and discussion among the authors, based on a previous report (Papageorgiou et al., 2011). A number of general and specific reporting items (characteristics) for each review were assessed. Two of the reviewers (S.N.P. and M.A.P.) completed in duplicate data extraction in a pre-designed collection form. The reviewers were not blinded to journal and author names, since masked assessment is inconsistent (Jadad et al., 1996). Any disagreement was resolved by consulting the last author (A.E.A.) until a final consensus was achieved. Inter-reviewer agreement on study selection and data extraction was assessed by Cohen’s kappa coefficient.

Epidemiological characteristics were based on the first author of each article in order to extract the country’s data, continent’s data, and the “academic source” of each article (i.e. whether the first author originated from an orthodontic academic department, a non-orthodontic academic department, or not from an academic department). The Cochrane Database of Systematic Reviews was classified as a general journal. Journals having published less than five orthodontic MAs and the Cochrane Database of Systematic Reviews were unified in a journal category named “Other”. The same was followed for quality assessment tools of primary studies that were used in less than five MAs. Journal impact factors (IFs) were acquired from ISI Journal Citation Reports (2010 Journal Citation Reports® Science Edition, 2011). H-indices were acquired from SCImago Journal & Country Rank (SCImago, 2007). Separate citation counts were acquired from Web of Science, Scopus, and Google Scholar, and their average value was used for analyses. All citation counts and journal metrics were acquired during the third week of August 2011.

Quality assessment

Articles were evaluated using the 11-item AMSTAR tool (Shea et al., 2007b). Each item was assessed using a four-point scale: “Yes”, “Can’t tell”, “No”, and “Not applicable”. A criterion was defined as “Can’t tell” if it was half met. For example, the fifth criterion, “A list of included and excluded studies should be provided”, was scored as “Can’t tell” if either the included or the excluded studies were listed. Non-applicable items were excluded from the maximum scoring capability of each MA. Summary scores were extracted by giving one point for each “Yes” and half a point for each “Can’t tell” (instead of just giving only one point for each “Yes”) in an attempt to maximize data output. Summary scores are reported as percentages. The “acceptance-to-publication time” was calculated as publication date minus acceptance date. Since the exact dates of acceptance and publication were not always available for all MAs, only the corresponding months were used for calculating the “acceptance-to-publication time”.

An MA was considered as of “low quality” if the AMSTAR score was between 0 and 4 points, of “moderate quality” if the AMSTAR score was between 4.5 and 8 points, and of “high quality” if the AMSTAR score was between 8.5 and 11 points.
Statistical analysis

Modified AMSTAR score was tested for its normal distribution with visual histograms and the Kolmogorov–Smirnov test ($P = 0.005$). Square root transformations were applied to achieve a distribution as close as possible to normal ($P = 0.079$). The data are presented as medians, interquartile range (IQR), and 95% confidence intervals (CIs) in natural units. The Student’s $t$-test and one-way ANOVA were used on the normalized data to identify differences.

To evaluate the association between the modified AMSTAR score and specific methodological characteristics, generalized linear models (GZLMs) were used. The natural modified AMSTAR score was used and the models were specified as having a normal distribution and a 0.5 power link. For bivariate comparisons, we constructed a separate model for each characteristic. A multivariate model was constructed to account for confounding factors, including independent variables with significant results from the bivariate models. Due to the large variables/studies ratio, instead of the usually used $P \leq 0.10$ a value of $P \leq 0.05$ was utilized for selecting variables for the multivariate model, and a post hoc Sidak correction for multiple comparisons was applied to the comparisons with reference categories.

Risk ratios (RRs) with 95% CIs were used as summary statistics to compare quality and reporting between specific time points: 2001 versus 2006, as well as 2006 versus 2011. The years 2001, 2006, and 2011, indicating the start, middle, and end of the last decade, were selected to investigate possible improvements (Hopewell et al., 2010) in reporting of key items associated with quality of orthodontic MAs. The appropriateness of the statistical methods used was evaluated with items based on a relevant report of Cochrane reviews (Riley et al., 2011).

In reporting the data, an alpha level of 0.05 was used as the criterion for the statistical significance of the estimated effects. All statistical analyses were performed with the IBM SPSS software package (version 19.0; SPSS, Chicago, IL, USA) except for the forest plots, which were drawn with the RevMan software version 5.1 (Review Manager, 2011).

Results

Literature search

The electronic search yielded 9070 initial citations (Supplementary Table S1b), while two additional papers were identified through communication with authors of systematic reviews. From the 5698 records that remained after duplicate exclusion, a total of 5618 records were excluded with reasons, and thus 80 MAs were deemed eligible for data extraction (Figure 1). The kappa scores before reconciliation for the selection and data extraction procedures were 0.910 and 0.867, respectively, thus indicating almost perfect agreement.

Publishing characteristics

The 80 MAs included 1086 primary studies, 62 214 251 patients, and were published between 1991 and 2011 in 39 distinct journals (Supplementary Table S2). Figure 2 presents the time evolution of MAs published from 1996 to 2011. No orthodontic MAs were published during 1992–95, while one MA published in 1991 is omitted in the graph.
The MAs referred to a wide set of areas in the field of orthodontics (Table 1). The evaluation of the MAs according to AMSTAR is presented in Supplementary Table S3, while their general characteristics and the corresponding modified median AMSTAR scores are presented in Supplementary Table S4.

Although a citation analysis is not formally displayed in the tables, the following information was obtained. Out of the 80 MAs, 66 (82.5%) were cited at least once in one of the three databases searched. For 78 MAs (97.5%), the citations from Google Scholar were equal or higher to those from Scopus or Web of Science. Also, the six most extreme citation counts (over 100 citations) for an MA were identified via Google Scholar. The average citations per MA from the three databases had a median of 1.7 citations (95% CI: 2.0–4.1, IQR = 3.8) per year. Chronologically, the highest citation counts per MA were found for those published in 1997 (61.0 citations), followed by 1999 (58.5 citations), and 1998 (42.0 citations), respectively. MAs from North America received more citations per MA than from other continents (26.3 citations), while MAs from the Netherlands received more citations per MA than other countries (38.3 citations). At journal level, the highest citation count per MA belonged to the Cleft Palate-Craniofacial Journal (28.3 citations), followed by the Angle Orthodontist (5.7 citations) and the American Journal of Orthodontics and Dentofacial Orthopedics (3.7 citations). MAs originating from university department (based on the first author) received on average more citations than non-academic ones (14.2 versus 2.0 citations, respectively). Through Web of Science, a total of 986 citations could be tracked for 45 MAs (19 non-indexed MAs and 16 non-cited MAs). At country level, the US contributed the greatest to the citing of orthodontic MAs (21.7%), followed by Canada (7.1%) and the UK (7.0%). At continent level, Europe contributed the greatest (41.8%), followed by North America (29.3%) and Asia (16.6%).

### Methodological quality
Reporting quality varied among reviews ranging from 13.6% to 100.0%, with a median of 50.0% and an IQR of 40.9% (median = 5.5 and IQR = 4.0 AMSTAR points). Twenty-five MAs (31.3%) were of low quality, 37 (46.3%) of moderate quality, and 18 (22.5%) of high quality. Supplementary Table S3 provides in summary the evaluation of the 80 MAs according to the AMSTAR tool. Twenty-two reviews (27.5%) clearly reported only the review question or only the inclusion criteria, while 25 MAs (31.3%) conducted in duplicate only study selection but not data extraction. Grey literature was not at all scanned for relevant articles in 55 reviews (68.8%). A list of excluded studies was not provided in 52 reviews (65.0%), while two reviews (2.5%) did not provide included or excluded studies in a list or a table at all.

### Factors associated with reporting quality
Table 2 provides the results from univariate and multivariate regression analyses used in order to explore MAs’ characteristics possibly related to an increase of the modified AMSTAR score. Comparisons were made between baseline (reference category) and each of the remaining groups per characteristic.

After adjusting for potential confounding factors through the multivariate analysis, several characteristics remained significantly associated with quality score (Table 2). More specifically, the following statistically significant observations were made. Every passing year was associated with a 0.5% (95% CI: 0.1–0.9%) increase in AMSTAR score. Internal financial support was associated with increased quality score by 8.0% (95% CI: 2.1–13.9%) compared to no financial support. Use of a protocol seemed to increase AMSTAR score by 6.4% (95% CI: −1.7 to

### Table 1 Subjects investigated by the orthodontic MAs.

| Thematic category                                         | n   |
|-----------------------------------------------------------|-----|
| Diagnosis and treatment planning                          | 5   |
| Cephalometry                                              | 2   |
| Moyer’s method/mixed dentition analysis                   | 2   |
| Crown height                                              | 1   |
| Treatment modalities                                      | 31  |
| Anchorage management and reinforcement                    | 3   |
| Chincup treatment                                         | 1   |
| Functional appliances                                     | 5   |
| Incisor intrusion                                         | 1   |
| Maxillary expansion                                       | 3   |
| Maxillary protraction                                     | 2   |
| Non-compliance appliances                                 | 2   |
| Premolar extraction                                       | 1   |
| Self-ligation                                             | 3   |
| Temporomandibular joint disorders/bruxism                 | 7   |
| Treatment of avulsed teeth                                | 1   |
| Treatment of posterior clefts                             | 1   |
| Treatment of prominent front teeth                        | 1   |
| Clinical entities                                         | 28  |
| Cleft lip and palate: etiology, effects, and treatment    | 20  |
| Obstructive sleep apnea                                   | 5   |
| Tooth agenesis                                            | 1   |
| Tooth transposition                                       | 2   |
| Fluorization and caries prevention                        | 2   |
| Periodontal considerations                                | 4   |
| Injury due to increased overjet                           | 1   |
| Treatment efficiency                                      | 2   |
| Intercanine width                                         | 1   |
| Stability                                                 | 1   |
| Adverse effects                                           | 4   |
| Bond strength                                             | 1   |
| External apical root resorption                           | 1   |
| Nickel hypersensitivity                                   | 1   |
| Pain management                                           | 1   |
| Tobacco cessation program                                 | 1   |
| Orthodontic education                                     | 1   |
| Orthodontic literature                                    | 1   |
Table 2  Regression analyses for predictors of modified AMSTAR score among orthodontic MAs.

| General characteristics of the MAs | Univariate | Multivariate |
|-----------------------------------|------------|--------------|
|                                   | $\beta$    | 95% CI       | $P$-value  | $\beta$    | 95% CI       | $P$-value  |
| Journal type ($N = 80$)           | General dentistry | Reference 0.2 | $-7.3$ to $7.6$ | 0.967 | NT          |            |
|                                   | Specialty (orthodontics) | Reference |            |            |            |            |
| Published in the Cochrane Library ($N = 80$) | Yes | 25.7 | 18.8 to 32.7 | 0.000 | $-1.9$ | -10.1 to 6.3 | 0.648 |
| Journal ($N = 80$)                | No         | Reference $-5.0$ | $-19.1$ to $9.1$ | 0.483 | NT          |            |
|                                   | AO         | Reference $-8.0$ | $-25.2$ to $9.3$ | 0.352 | NT          |            |
|                                   | CPC        | Reference $-4.7$ | $-14.0$ to $4.6$ | 0.333 | NT          |            |
| Number of affiliations ($N = 80$) | General dentistry | Reference 1.1 | $-0.6$ to $2.8$ | 0.211 | NT          |            |
|                                   | Specialty (orthodontics) | Reference |            |            |            |            |
| Number of authors ($N = 80$)      | Yes        | Reference $-3.7$ | $-11.2$ to $3.8$ | 0.334 | NT          |            |
| Bioskistatistician / epidemiologist involved ($N = 80$) | Yes | 5.5 | $-1.7$ to $12.7$ | 0.138 | NT          |            |
| Authored multiple MAs ($N = 80$)  | Yes        | Reference $-6.5$ | $-22.1$ to $9.0$ | 0.411 | NT          |            |
|                                   | No         | Reference $-8.7$ | $-24.3$ to $7.0$ | 0.284 | NT          |            |
|                                   |          | Reference $-6.5$ | $-16.8$ to $3.8$ | 0.214 | NT          |            |
| Academic source ($N = 80$)        | Orthodontic department | Reference $-2.4$ | $-13.0$ to $8.1$ | 0.652 | NT          |            |
|                                   | Other      | Reference $12.5$ | $0.4$ to $24.6$ | 0.044 | NT          |            |
| Professor in authors ($N = 38$)   | Yes        | Reference $-9.5$ | $-16.8$ to $-2.2$ | 0.011 | NT          |            |
|                                   | No         | Reference $-8.7$ | $-24.3$ to $7.0$ | 0.284 | NT          |            |
| Financial support ($N = 36$)      | External   | Reference $12.5$ | $0.4$ to $24.6$ | 0.044 | NT          |            |
|                                   | Internal   | Reference $-9.5$ | $-16.8$ to $-2.2$ | 0.011 | NT          |            |
| “Systematic review/meta-Analysis” stated in title ($N = 80$) | Yes | 8.0 | $2.1$ to $13.9$ | 0.017 | NT          |            |
| Language ($N = 80$)               | English bilingual | Reference $-23.4$ | $-68.4$ to $21.6$ | 0.307 | NT          |            |
|                                   | Other      | Reference $-18.3$ | $-39.0$ to $3.5$ | 0.014 | NT          |            |
| Update of review ($N = 80$)       | Yes        | Reference $21.6$ | $11.3$ to $31.9$ | 0.000 | NT          |            |
| Protocol ($N = 80$)               | No         | Reference $22.0$ | $16.4$ to $27.6$ | 0.000 | NT          |            |
| Dates of search ($N = 77$)        | Yes        | Reference $8.7$ | $-6.4$ to $23.9$ | 0.259 | NT          |            |
|                                   | No         | Reference $10.2$ | $-1.6$ to $22.0$ | 0.089 | NT          |            |
| Boolean strategy ($N = 76$)       | Yes        | Reference $14.9$ | $8.3$ to $21.5$ | 0.000 | NT          |            |
|                                   | No         | Reference $1.5$ | $0.7$ to $2.2$ | 0.000 | NT          |            |
| Number of databases ($N = 76$)    | Yes        | Reference $6.3$ | $-1.0$ to $13.5$ | 0.091 | NT          |            |
| Participant flow diagram ($N = 80$) | Yes | 19.9 | $11.7$ to $28.2$ | 0.000 | NT          |            |
| Forest plot ($N = 80$)            |            | Reference $19.9$ | $11.7$ to $28.2$ | 0.000 | NT          |            |
| Included studies type ($N = 78$)  | RCT        | Reference $11.8$ | $-19.2$ to $-4.5$ | 0.002 | NT          |            |
|                                   | RCT and Other | Reference $-22.5$ | $-29.8$ to $-15.2$ | 0.000 | NT          |            |
|                                   | Other      | Reference $-0.4$ | $-0.8$ to $0.001$ | 0.051 | NT          |            |
| Number of included studies ($N = 80$) | Yes | 0.0 | $0.0$ to $0.0$ | 0.084 | NT          |            |
| Number of participants ($N = 72$) | Yes        | Reference $0.0$ | $0.0$ to $0.0$ | 0.211 | NT          |            |
| Participants per included study ($N = 72$) | Yes | 0.0 | $0.0$ to $0.0$ | 0.211 | NT          |            |
| Assessed quality of included studies ($N = 80$) | Yes | 25.9 | $20.9$ to $30.9$ | 0.000 | NT          |            |
| Assessment tool ($N = 40$)        | No         | Reference $10.5$ | $5.1$ to $15.9$ | 0.000 | NT          |            |
|                                   | Cochrane   | Reference $-6.5$ | $-13.7$ to $0.8$ | 0.080 | NT          |            |
|                                   | Custom     | Reference $-0.5$ | $-10.3$ to $9.2$ | 0.914 | NT          |            |
|                                   | Jadad      | Reference $-11.7$ | $-20.1$ to $-3.3$ | 0.006 | NT          |            |
| Conclusions ($N = 80$)            | Significant | Reference $-8.0$ | $-16.8$ to $0.9$ | 0.079 | NT          |            |
|                                   | Can’t tell | Reference $-8.0$ | $-16.8$ to $0.9$ | 0.079 | NT          |            |

(Continued)
11.0%). Presentation of the complete Boolean search strategy was associated with increased quality score by 5.9% (95% CI: 2.4–9.4%). The assessment of the reporting quality of included primary studies was associated with a net score increase of 13.0% (95% CI: 8.8–17.3%). Finally, the corresponding AMSTAR score was associated with a 1.5% (95% CI: 0.8–2.3%) increase for every additional IF point of the published journal.

Comparisons of methodological items reporting during the years

With regard to the comparison in reporting of the methodological items of MAs between the years 2001, 2006, and 2011, no definitive conclusions can be drawn. The small number of MAs per years 2001 (n = 4), 2006 (n = 6), and 2011 (n = 12) is possibly contributing to the non-significance level of the results and the wide 95% CIs observed. Concerning the comparison between 2006 and 2001, tendencies were found for improvement in three items and deterioration in five items (Figure 3A), while between the years 2011 and 2006, tendencies were found for improvement in five items and deterioration in three items (Figure 3B).

Statistical methods used in the MAs

Number of MAs

The statistical methods used in the included studies are presented in Table 3. The 80 MAs included in current investigation presented a large number of meta-analytic procedures for data combination. A median number of two forest plots per review (min = 0; max = 22) was given, while each plot presented one or more summary estimates. Such large numbers of forest plots per review can be interpreted by the variety of questions, outcomes, interventions, and subgroups tested. It also brings the concern of outcome hierarchy and multiple testing, which increase the risk of false significant results. Indeed, only 16 (20.0%) MAs did clearly define primary and secondary outcomes.

### Table 2

Continued

| General characteristics of the MAs | Univariate | Multivariate |
|-----------------------------------|------------|--------------|
|                                   | β          | 95% CI       | P-value | β          | 95% CI       | P-value |
| Country (N = 80)                  |            |              |         |            |              |         |
| US                                | Reference  | Reference    |          |            |              |         |
| UK                                | 20.1       | 11.3 to 29.0 | 0.000   | −9.1       | −16.3 to −1.8| 0.199   |
| Canada                            | 0.3        | −10.5 to 11.1| 0.954   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| China                             | 4.2        | −6.2 to 14.6 | 0.427   | 0.000      | 4.1          | 0.95 to 7.2| 0.000   |
| Greece                            | −0.5       | −11.4 to 10.4| 0.929   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| Italy                             | −2.1       | −16.7 to 12.5| 0.782   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| Switzerland                       | 2.9        | −10.8 to 16.7| 0.676   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| Australia                         | 3.0        | −15.6 to 21.6| 0.751   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| Brazil                            | 20.3       | 4.9 to 35.7  | 0.010   | −1.5       | −8.8 to 5.7  | 1.000   |
| Germany                           | −17.8      | −53.2 to 17.7| 0.326   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| Korea                             | −13.6      | −46.5 to 19.3| 0.418   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| Saudi Arabia                      | −9.7       | −40.5 to 21.1| 0.538   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| Scotland                          | −9.7       | −40.5 to 21.1| 0.538   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| Singapore                         | 30.0       | 10.9 to 49.2 | 0.002   | −8.0       | −19.1 to 3.1 | 0.929   |
| South Africa                      | −33.1      | −82.9 to 16.7| 0.193   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| Spain                             | −2.6       | −30.3 to 25.1| 0.856   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| Uganda                            | −27.3      | −70.6 to 15.9| 0.215   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| Continent (N = 80)                |            |              |         |            |              |         |
| Europe                            | Reference  | Reference    |          |            |              |         |
| North America                     | −8.9       | −16.5 to −1.2| 0.023   | 1.6        | −1.7 to 4.9 | 0.878   |
| Asia                              | −3.3       | −13.8 to 7.2 | 0.540   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| Africa                            | −31.3      | −58.7 to −3.8| 0.026   | −11.1      | −22.0 to −0.2| 0.161   |
| South America                     | −5.9       | −28.1 to 16.2| 0.599   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| Oceania                           | 11.4       | −6.7 to 29.5 | 0.217   | 0.000      | 10.4         | 1.1 to 19.7| 0.510   |
| Mean citations (N = 80)           |            |              |         |            |              |         |
| Annual citation rate (N = 80)     |            |              |         |            |              |         |
| Journal’s impact factor (N = 73)  |            |              |         |            |              |         |
| Journal’s Hirsch index (h-index)  |            |              |         |            |              |         |
| Acceptance-to-publication time     |            |              |         |            |              |         |
| (days) (N = 52)                   |            |              |         |            |              |         |

MAs, meta-analyses; NT, not tested; CI, confidence interval; AJODO, American Journal of Orthodontics and Dentofacial Orthopedics; AO, Angle Orthodontist; CPC, Cleft Palate-Craniofacial Journal; RCT, randomized controlled trial; β, regression coefficient.

Downloaded from https://academic.oup.com/ejo/article-abstract/36/1/74/715723 by guest on 15 March 2020
**Table 3** Findings from the assessment of included MAs.

| Category/item (N = 80) | n (%) |
|------------------------|-------|
| Number/presentation of MA procedures |       |
| Forest plot number      |       |
| None                    | 25 (31.3%) |
| One                     | 10 (12.5%) |
| Two or more             | 45 (56.3%) |
| Defined primary/secondary outcomes | 16 (20.0%) |
| Heterogeneity           |       |
| Explained how heterogeneity was assessed | 55 (68.8%) |
| Examined causes for heterogeneity | 40 (50.0%) |
| Subgroups/sensitivity analysis | 39 (48.8%) |
| Predefined subgroups    | 18 (22.5%) |
| F for deciding between FE/RE | 20 (25.0%) |
| Actually did            | 18 (22.5%) |
| Reported between-study variance “tau-squared” | 8 (10.0%) |
| (For RE) Provided clinical justification/interpretation | 8 (10.0%) |
| (For FE) Moderate or large heterogeneity present | 8 (10.0%) |
| Large heterogeneity without explanation | 17 (21.3%) |
| Meta-regression         | 4 (5.0%) |
| Publication bias        |       |
| Explained how publication bias was assessed | 26 (32.5%) |
| No assessment/explanation | 11 (13.8%) |
| Issue in discussion     | 15 (18.8%) |
| Funnel plot             | 21 (26.3%) |
| Statistical test for publication bias | 12 (15.0%) |
| Evidence of heterogeneity | 6 (7.5%) |

MA, meta-analysis; FE, fixed-effects model; RE, random-effects model.

*Maximum number is less than 80, since they were not applicable to all studies; *Out of n = 49 studies with at least one RE estimate and of n = 26 studies with at least one FE estimate, respectively.

**Discussion**

This study provides a comprehensive assessment of the design and reporting characteristics of the largest cohort of orthodontic MAs until the collection of the present data and follows a previous evaluation of orthodontic systematic reviews without meta-analytical synthesis (Papageorgiou et al., 2011). The number of these reviews has increased over time with variability in the reporting quality. The examined MAs predominantly addressed questions about the effectiveness of therapeutic interventions and rare clinical entities (e.g. cleft lip and palate or obstructive sleep apnoea).

Although MAs are considered by many clinicians as the top of the “evidence pyramid” and attract the highest relative citation counts (Patsopoulos et al., 2005), many of the identified reports did not report methods and bias assessment in sufficient detail. Certain characteristics of transparent design were not reported by a large number of reviews, including comprehensive literature search, validity of selection/extraction procedures, or methodological assessment of included studies, all of which are important for the replication and evaluation of the review (Sutton et al., 1998). Moreover, a suboptimal reporting of potential conflicts of interest was identified, despite the increasing concern that funding agencies influence the outcomes of biomedical research (Smith, 2005; Götzsche et al., 2009). Noteworthy,
articles disclosing sources of funding have been shown to be significantly more likely to be published than those without any disclosure (Lee et al., 2006), which indicates that transparent MAs could also be published easier.

About 39% of the 80 MAs were published in orthodontic journals. Although North America is regarded as the most prolific continent regarding orthodontic literature (Kanavakis et al., 2006), the majority of orthodontic MAs were produced in Europe (42.5%) with North America coming second (36.3%). At the journal level, the American Journal of Orthodontics and Dentofacial Orthopedics and the Angle Orthodontist received MAs from three continents, and the Cleft Palate-Craniofacial Journal from two continents. The majority of MAs published by each of the three journals originated from Europe. The number of orthodontic journals has increased during the last years and the quality of the MAs accepted for publication by them was found higher than general biomedical journals, although not significantly. The IF characteristics of orthodontic journals have been previously discussed (Eliades and Athanasiou, 2001). In this study, scientific impact was measured both by the journal’s IF and the h-index equivalent for journals (Braun et al., 2006), which seems to be quite robust (Vanclay, 2007). In this analysis, higher AMSTAR score was associated in some with both the journal’s h-index (Supplementary Table S4) and IF (Table 2), showing a possible preference of submitting high-quality MAs to journals with major impact. However, MAs published in prestigious medical journals (such as the ones presenting higher IFs) have also been found to demonstrate exaggerated results compared with trials in other journals (Siontis et al., 2011).

The impact of orthodontic MAs was also assessed with the average of the citation counts from three databases. Citation counts differed among Google Scholar, Web of Science, and Scopus, possibly reflecting the quantitatively and qualitatively different coverage of each database (Kulkarni et al., 2009) and the small overlap among them (Mehta et al., 2007). No association was observed between AMSTAR score and either average citations, or individual citation counts of each database. The same observation was made for systematic reviews in orthodontics (Papageorgiou et al., 2011) or original research articles in the field of psychiatrics (Nieminen et al., 2006). In the present study, self-citations were not excluded. However, citation counts were not given overly much attention, as a citation does not guarantee the respect of the article referenced, but only that it is active in the scientific debate (Patsopoulos et al., 2005).

In this study, the reporting of orthodontic MAs was assessed with the AMSTAR tool, which is the most recent evidence-based appraisal instrument. AMSTAR has been validated to some extent (Shea et al., 2007a, 2009; Sampson et al., 2008a) and proposed by the Canadian Agency for Drugs and Technologies in Health and the World Health Organization as the best tool to critically appraise MAs (Proposed Evaluation Tools for COMPUS, 2005; Oxman et al., 2006; Pantoja and Campbell, 2009). The modifications made to the AMSTAR scoring system that were also used in a previous report (Papageorgiou et al., 2011) may possibly help to overcome assessment difficulties reported for it (Faggion et al., 2012). Nevertheless, like every instrument, AMSTAR does present weaknesses. For example, there is no recommendation on how the scientific quality of studies should be assessed. Other critics include the applicability of AMSTAR to reviews of non-randomized studies (Fedorowicz et al., 2011) or to mixed-methods reviews (Bouchard et al., 2011), which are reviews evaluating studies that employ qualitative, quantitative, and mixed methodology.

Improved quality was significantly related to certain MA characteristics. Financial support from the originating institution was associated with higher quality score. Also, MAs based on a protocol for guidance scored higher than others, something also observed for orthodontic systematic reviews (Papageorgiou et al., 2011).

Full provision of the Boolean search strategy was accompanied by an increase in AMSTAR score and is crucial for the reproducibility of the literature search (Maggio et al., 2011). Other reports have pointed out that the suboptimal reporting in orthodontics during the search and selection procedures needs improvement (Flores-Mir et al., 2006). Evaluation of electronic searches in dental systematic reviews has yielded similar results. A survey of systematic reviews’ authors reported a lack of comprehensive literature searches (Major et al., 2009), which was also found in this study. Indeed, only 45% of the included MAs reported undertaking an extensive literature search according to the AMSTAR tool. A comprehensive assessment of systematic reviews in dentistry by Glenny et al. (2003) found that 8 out of 15 proposed key items were often not assessed, with literature search having the most problems. Better search and selection methodologies have been reported for certain dental specialities compared to others, although these specialities were also the most prolific ones in terms of publications (Major et al., 2007). Lately, as no consensus on search reporting methods exists (Sampson et al., 2008a), an evidence-based guideline for the peer review of electronic search strategies was developed (Sampson et al., 2009) in order to assure the quality of electronic searches and increase their effectiveness.

Methodological assessment of included studies was clearly associated with an increase in AMSTAR score. Regrettfully, only half of the MAs formally assessed possible problems in the primary studies, while the most frequently used tool was the Cochrane risk of bias tool (Higgins et al., 2011). Evaluating the quality, or preferably the “risk of bias”, as described in the PRISMA statement (Liberati et al., 2009) and the Cochrane Handbook (Higgins and Green, 2011) is a crucial component of a systematic review. Poor primary study quality has been possibly associated with effect overestimation or underestimation (Hopewell et al., 2007), although
its role is to a large extent uncertain (Balk et al., 2002). Under existing evidence, it may not be necessary for risk of bias assessments in a systematic review to be conducted under blinded conditions (Morissette et al., 2011). A wide variety of checklists and scales exist (Moher et al., 1995; Sanderson et al., 2007; Higgins et al., 2011), although the use of scales has been cautioned (Jüni et al., 1999; Greenland and O’Rourke, 2001). Systematic reviews should always evaluate and take into account the internal validity of included trials, but also their applicability and generalizability or external validity (Dekkers et al., 2010). However, adjustment of the MA on the basis of quality scores is considered inappropriate and should be avoided (Herbison et al., 2006).

Some additional issues should be also discussed, although their influence was non-significant in the present analysis. Firstly, the superior reporting of Cochrane reviews has been noted previously (Moher et al., 2007a; Papageorgiou et al., 2011) and may be aided by the guidelines of the Cochrane Collaboration, although criticisms for these reviews have also been issued (Lang et al., 2007). In addition, the electronic publishing allows more freely the provision of details. Cochrane reviews provide more details concerning the inclusion and exclusion criteria and are updated more often (Jadad et al., 1998; Shea et al., 2007a,b). Indeed, once the specific characteristics of a proper systematic review are taken into account, Cochrane reviews were not associated with higher AMSTAR score (Table 2). Secondly, systematic reviews have to be up to date in order to be valid, but only six of the identified MAs were updates of previous ones. The low update rate of non-Cochrane MAs, which account for 80% of all systematic reviews (Moher et al., 2007a), may relate to the few methods or strategies currently existing for the actual updating (Moher et al., 2007b). The rapid dissemination of the informative value of a systematic review can deteriorate due to publication lag, which may account for up to 20% of an MA’s life span (Sampson et al., 2008b). Even worst, systematic reviews without MA may be given even lower priority by editors. Thirdly, the quality improvement associated with the participation of statisticians or epidemiologists among authors, reported for controlled clinical trials (Delgado-Rodriguez et al., 2001) or orthodontic systematic reviews (Papageorgiou et al., 2011), was not statistically significant in this study. However, the use of statistical expertise is associated with higher publication acceptance rates (Altman et al., 2002) and advocated due to the moderate statistical skills frequently found among clinical medical researchers (Perneger et al., 2004) or orthodontic postgraduate students (Polychronopoulou et al., 2011).

Although an annual minimal increase of AMSTAR score (0.5%) was identified, reporting seems not to have significantly improved in some areas of methodological importance. Any tendencies found by comparing 2001 and 2006 are not discussed further and should be interpreted with caution, due to the non-significance and the wide 95% CIs. However, between the years 2006 and 2011, a tendency for improvement was found regarding Boolean strategy reporting, flow diagram reporting, and the quality assessment of included studies. A decrease was noted in the number of MAs that provided complete search dates or included a statistician/epidemiologist.

The appropriateness of statistical methods for the pooling of data is essential for the validity and consistency of the MAs and is often criticized (Feinstein, 1995; Sharpe, 1997). Although numerous articles assessing reports of MAs exist, only a few of them incorporate terms of actual statistical methods (Minelli et al., 2009; Jude-Eze, 2011; Korevaar et al., 2011; Riley et al., 2011; Melchior et al., 2012). Regarding the distinction between primary and secondary outcomes, which was seldom found in this analysis (Table 3), it should be mentioned that this is best made in the Methods section. It is often impossible for the reader to distinguish primary from secondary outcomes from the data provided in the Results section or the forest plots.

Many issues were raised analysing the heterogeneity assessments of the MAs, which was often inadequate. A similar analysis of MAs of preclinical studies showed that the percentage of studies reporting on heterogeneity increased after 2005 (Korevaar et al., 2011). First of all, the authors should predefine how the choice between a fixed-effects and a random-effects model for the MA was made. Selection of one from these two methods is often based on either a “large” $I^2$ value or on the $I^2$-value for a chi-square test for heterogeneity. The size and impact of the between-study heterogeneity are statistically properly measured by the between-study variance estimate (“tau-squared”) and $I^2$, respectively. It is recommended, however, that both statistical and clinical reasoning are considered for this decision (Higgins and Green, 2011). Also, the pooled result from the two MA models is not the same and should be interpreted accordingly. The result of a random-effects model produces the average of the intervention effects across studies, while the result of a fixed-effects model produces the best estimate of a common intervention effect across studies (Riley et al., 2011). Sensitivity analyses, subgroup analyses, meta-regressions can be prespecified to identify causes for heterogeneity, but potential pitfalls of these approaches should be reported (Brookes et al., 2004; Lambert et al., 2002). Finally, for a random-effects MA, the 95% prediction interval for the underlying intervention effect and the “tau-squared” provide invaluable information to the readers.

Although publication bias is known to endanger the validity of MA (Dwan et al., 2008) and lead to further emergence of biasing evidence (Ioannidis and Lau, 2001; Trikalinos et al., 2004), few MAs investigated their presence. A similar evaluation in the field of preclinical studies showed an improvement in reporting of publication bias (Korevaar et al., 2011). Firstly, publication bias (or file-drawer problem) must be considered during both the planning and interpretation of a review article. It should be noted that bias assessments are not always feasible (due to a small number...
of studies) or reliable (due to large heterogeneity), and that funnel plot asymmetry may exist due to reasons other than publication bias (Sterne et al., 2011). As suggested in the Cochrane Handbook (Higgins and Green, 2011), for each MA with at least 10 primary studies, authors should produce a funnel plot and visually assess plot asymmetry or “small study effects” (Harbord et al., 2006). This is the tendency that intervention effects estimated in smaller studies differ from those estimated in larger studies. Secondly, authors should report the results of the predefined statistical tests for publication bias (Begg and Mazumdar, 1994; Egger et al., 1997; Harbord et al., 2006; Peters et al., 2006). Bias identified in any way should be investigated by sensitivity analysis and placed in context of the MA results and conclusions.

Among the limitations of the present study is the fact that there is still a possibility of missing existing MAs, despite the extensive search performed. In addition, it is also possible that some reporting details and characteristics of the MAs have been discarded during the peer-review process. Finally, it is feasible that a research project may undergo substantial changes between different stages (e.g. protocol, execution, article publication).

Provision of reporting guidelines is a validated mean of improving the quality of published material, as the reporting quality of RCTs and MAs has clearly improved (Plint et al., 2006; Al Faleh and Al-Omran, 2009) following the introduction of the CONSORT (Consolidated Standards of Reporting Trials) and QUOROM statements (Moher et al., 1999, 2001). A recent masked randomized trial identified that additional journal peer review based on reporting guidelines resulted in a moderate improvement in manuscript quality, although authors still had difficulties in adhering to high standards of reporting during the writing phase (Cobo et al., 2011). In addition, guidance from the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) approach (Guyatt et al., 2008) or other actions (Tricco et al., 2011) may improve the interpretation of systematic reviews. Regarding the evidence basis of the identified MAs, it should be emphasized that many lacked certain procedures of an extensive systematic review (PLoS Medicine Editors, 2007).

Conclusion

The critical appraisal of MAs in the field of orthodontics suggests that the average quality of MAs could be characterized as low to medium. Although a minimal trend for improvement characterized the last decade, significant flaws were found. Without complete and transparent reporting and appropriate use of statistical methods, it is difficult for readers to assess the validity of MAs or identify MAs with misleading conclusions. The endorsement (or even enforcement) of the PRISMA statement (Liberati et al., 2009) would hopefully improve the conduct and reporting of systematic reviews and consequently of the MAs in orthodontics.

Supplementary material

Additional data is available as supplement at European Journal of Orthodontics online.

Acknowledgements

The authors are grateful to Dr E. G. Kaklamanos, Aristotle University of Thessaloniki, and Dr C. Flores-Mir, School of Dentistry, University of Alberta, for providing their full-text papers prior to official publication. The authors declare to have no conflict of interest. M. A. P., A. E. A. and S. N. P. had authored included MAs, but the first two had no involvement in the data extraction or evaluation of the corresponding studies. The data regarding the one MA that was co-authored by S. N. P. were also checked by I. C. Konstantinidis, Mount Sinai School of Medicine.

References

2010 Journal Citation Reports® Science Edition 2011, Thomson Reuters (http://thomsonreuters.com) Accessed on August 24 2011

Al Faleh K, Al-Omran M 2009 Reporting and methodologic quality of Cochrane Neonatal review group systematic reviews. BMC Pediatrics 9: 38

Altman D G, Goodman S N, Schrotter S. 2002 How statistical expertise is used in medical research. Journal of the American Medical Association 287: 2817–2820

Balk E M et al. 2002 Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. Journal of the American Medical Association 287: 2973–2982

Begg C B, Mazumdar M 1994 Operating characteristics of a rank correlation test for publication bias. Biometrics 50: 1088–1101

Bouchard K, Dubuisson W, Simard J, Dorval M 2011 Systematic mixed methods reviews are not ready to be assessed with the available tools. Journal of Clinical Epidemiology 64: 926–928

Braun T, Glänzel W, Schubert A 2006 A hirsch-type index for journals. Scientometrics 69: 169–173

Brookes S T, Whiteley E, Egger M, Smith G D, Mulheran P A, Peters T J 2004 Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. Journal of Clinical Epidemiology 57: 229–236

Chavalarias D, Ioannidis J P 2010 Science mapping analysis characterizes 235 biases in biomedical research. Journal of Clinical Epidemiology 63: 1205–1215

Cobo E et al. 2011 Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical journal: masked randomised trial. British Medical Journal 343: d6783

Dekkers O M, van Elm E, Algra A, Romijn J A, Vandenhouwelle J P 2010 How to assess the external validity of therapeutic trials: a conceptual approach. International Journal of Epidemiology 39: 89–94

Delgado-Rodriguez M, Ruiz-Canela M, De Irala-Estevez J, Llorca J, Martinez-Gonzalez A 2001 Participation of epidemiologists and/or biostatisticians and methodological quality of published controlled clinical trials. Journal of Epidemiology and Community Health 55: 569–572

Dwan K et al. 2008 Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PloS One 3: e3081
Egger M, Davey Smith G, Schneider M, Minder C 1997 Bias in meta-analysis detected by a simple, graphical test. British Medical Journal 315: 629–634

Eliades T, Athanasiou A E 2001 Impact factor. A review with specific relevance to orthodontic journals. Journal of Orofacial Orthopedics 62: 74–83

Faggion C M Jr, Listl S, Giannakopoulos N N 2012 The methodological quality of systematic reviews of animal studies in dentistry. Veterinary Journal 192: 140–147

Fedorowicz Z, Sequeira-Byron P, Jagannath V, Sharif M O 2011 Climate change in endodontics: is it time to recycle “garbage in, garbage out” systematic reviews? Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics 112: 515–517

Feinstein A R 1995 Meta-analysis: statistical alchemy for the 21st century. Journal of Clinical Epidemiology 48: 71–79

Flores-Mir C, Major M P, Major P W 2006 Search and selection methodology of systematic reviews in orthodontics (2000-2004). American Journal of Orthodontics and Dentofacial Orthopedics 130: 214–217

Glenny A M, Esposito M, Coulthard P, Worthington H V 2003 The assessment of systematic reviews in dentistry. European Journal of Oral Sciences 111: 85–92

Gotzsche P C et al. 2009 What should be done to tackle ghostwriting in the medical literature? PLoS Medicine 6: e23

Greenland S, O’Rourke K 2001 Impact factor. A review with specific relevance to orthodontic journals. Journal of Orofacial Orthopedics 62: 74–83

Harbord R M, Egger M, Sterne J A 2006 A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Statistics in Medicine 25: 3443–3457

Herbison P, Hay-Smith J, Gillespie W J 2006 Adjustment of meta-analyses on the basis of quality scores should be abandoned. Journal of Clinical Epidemiology 59: 1249–1256

Higgins J P T, Green S 2011 Cochrane handbook for systematic reviews of interventions 5.1.0 (updated March 2011) (http://www.mrc-bsu.cam.ac.uk/cochrane/handbook.htm) Accessed on October 31 2012

Higgins J P et al.; GRADE Working Group 2008 GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. British Medical Journal 336: 924–926

Harbord R M, Egger M, Sterne J A 2006 A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Statistics in Medicine 25: 3443–3457

Herbison P, Hay-Smith J, Gillespie W J 2006 Adjustment of meta-analyses on the basis of quality scores should be abandoned. Journal of Clinical Epidemiology 59: 1249–1256

Higgins J P T, Green S 2011 Cochrane handbook for systematic reviews of interventions 5.1.0 (updated March 2011) (http://www.mrc-bsu.cam.ac.uk/cochrane/handbook.htm) Accessed on October 31 2012

Higgins J P et al.; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. 2011 The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials. British Medical Journal 343: d5928

Hopewell S, Dutton S, Yu L M, Chan A W, Altman D G 2010 The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. British Medical Journal 340: c723

Hopewell S, McDonald S, Clarke M, Egger M 2007 Grey literature in meta-analyses of randomized trials of health care interventions. Cochrane Database of Systematic Reviews 2: MR000010

Huwiler-Müntener K, Jüni P, Junker C, Egger M 2002 Quality of reporting of randomized trials as a measure of methodologic quality. Journal of the American Medical Association 287: 2801–2804

Ioannidis J, Lau J 2001 Evolution of treatment effects over time: empirical insight from recursive cumulative metaanalyses. Proceedings of the National Academy of Sciences of the United States of America 98: 831–836

Jadad A R et al. 1998 Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. Journal of the American Medical Association 280: 278–280

Jadad A R et al. 1996 Assessing the quality of reports of randomized clinical trials: is blinding necessary? Controlled Clinical Trials 17: 1–12

Jude-Eze E N 2011 Quality of meta-analysis in terms of actual statistical analysis. Master thesis, University of Glasgow

Jüni P, Witschi A, Bloch R, Egger M 1999 The hazards of scoring the quality of clinical trials for meta-analysis. Journal of the American Medical Association 282: 1054–1060

Kanavakis G, Spinos P, Polychronopoulos A, Eliades T, Papadopoulos M A, Athanasiou A E 2006 Orthodontic journals with impact factors in perspective: trends in the types of articles and authorship characteristics. American Journal of Orthodontics and Dentofacial Orthopedics 130: 516–522

Korevaar D A, Hoof L, ter Riet G 2011 Systematic reviews and meta-analyses of preclinical studies: publication bias in laboratory animal experiments. Laboratory Animals 45: 225–230

Kulkarni A V, Aziz B, Shams I, Busse J W 2009 Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. Journal of the American Medical Association 302: 1092–1096

Lambert P C, Sutton A J, Abrams K R, Jones D R 2002 A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. Journal of Clinical Epidemiology 55: 86–94

Lang A, Edwards N, Fleiszer A 2007 Empty systematic reviews: hidden perils and lessons learned. Journal of Clinical Epidemiology 60: 595–597

Lee K P, Boyd E A, Holroyd-Leduc J M, Bacchetti P, Bero L A 2006 Predictors of publication: characteristics of submitted manuscripts associated with acceptance at major biomedical journals. The Medical Journal of Australia 184: 621–626

Liberati A et al. 2009 The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Journal of Clinical Epidemiology 62: e1–34

Maggio L A, Tannery N H, Kanter S L 2011 Reproducibility of literature search reporting in medical education reviews. Journal of the American Society for Information Science & Technology 58: 2105–2125

Melchior A C, Correr C J, Venson R, Pontarolo R 2012 An analysis of quality of systematic reviews on pharmacist health interventions. International Journal of Clinical Pharmacy 34: 32–42

Minelli C, Thompson J R, Abrams K R, Thakkinstian A, Attia J 2009 The quality of meta-analyses of genetic association studies: a review with recommendations. American Journal of Epidemiology 170: 1333–1343

Moher D, Cook D J, Eastwood S, Olkin I, Rennie D, Stroup D F 1999 Improving the quality of reports of meta-analyses of randomized controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 354: 1896–1900

Moher D, Jadad A R, Nichol G, Penman M, Tugwell P, Walsh S 1995 Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Controlled Clinical Trials 16: 62–73

Moher D et al. 1998 Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 352: 609–613

Moher D, Schulz K F, Altman D G; CONSORT GROUP (Consolidated Standards of Reporting Trials) 2001 The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. Annals of Internal Medicine 134: 657–662

Moher D, Tetzlaff J, Tricco A C, Sampson M, Altman D G 2009 Improving the quality of reports of meta-analyses of preclinical studies: publication bias in laboratory animal experiments. Laboratory Animals 43: 175–179

Morissette K, Tricco A C, Horsley T, Chen M H, Moher D 2011 Blinded versus unblinded assessments of risk of bias in studies included in a systematic review. Cochrane Database of Systematic Reviews 9: MR000025
