Review Article

pISSN 2234-7518 • eISSN 2005-372X
https://doi.org/10.4041/kjod.2017.47.6.401

THE KOREAN JOURNAL of ORTHODONTICS
KJO

Outcomes of comprehensive fixed appliance orthodontic treatment: A systematic review with meta-analysis and methodological overview

Spyridon N. Papageorgiou
Damian Höchli
Theodore Eliades

Clinic of Orthodontics and Pediatric Dentistry, Center of Dental Medicine, University of Zurich, Zurich, Switzerland

Objective: The aim of this systematic review was to assess the occlusal outcome and duration of fixed orthodontic therapy from clinical trials in humans with the Objective Grading System (OGS) proposed by the American Board of Orthodontics. **Methods:** Nine databases were searched up to October 2016 for prospective/retrospective clinical trials assessing the outcomes of orthodontic therapy with fixed appliances. After duplicate study selection, data extraction, and risk of bias assessment according to the Cochrane guidelines, random-effects meta-analyses of the mean OGS score and treatment duration were performed and 95% confidence intervals (CIs) were calculated. **Results:** A total of 34 relevant clinical trials including 6,207 patients (40% male, 60% female; average age, 18.4 years) were identified. The average OGS score after treatment was 27.9 points (95% CI, 25.3–30.6 points), while the average treatment duration was 24.9 months (95% CI, 24.6–25.1 months). There was no significant association between occlusal outcome and treatment duration, while considerable heterogeneity was identified. In addition, orthodontic treatment involving extraction of four premolars appeared to have an important effect on both outcomes and duration of treatment. Finally, only 10 (39%) of the identified studies matched compared groups by initial malocclusion severity, although meta-epidemiological evidence suggested that matching may have significantly influenced their results. **Conclusions:** The findings from this systematic review suggest that the occlusal outcomes of fixed appliance treatment vary considerably, with no significant association between treatment outcomes and duration. Prospective matched clinical studies that use the OGS tool are needed to compare the effectiveness of orthodontic appliances.
[Korean J Orthod 2017;47(6):401-413]

Key words: Orthodontics, Treatment outcome, Treatment duration, Meta-analysis

Received January 10, 2017; Revised February 21, 2017; Accepted March 29, 2017.

Corresponding author: Spyridon N. Papageorgiou.
Senior Teaching and Research Assistant, Clinic of Orthodontics and Pediatric Dentistry, Center of Dental Medicine, University of Zurich, Plattenstrasse 11, Zurich 8032, Switzerland.
Tel +41-44-634-32-87 **e-mail** snpapage@gmail.com

The authors report no commercial, proprietary, or financial interest in the products or companies described in this article.

INTRODUCTION

Fixed appliances have become an integral part of comprehensive orthodontic treatment as versatile tools that enable three-dimensional control of tooth movement. Through the years, considerable effort has been invested in the optimization of orthodontic appliances to increase their treatment efficiency,[1-5] with the primary goals of developing interventions that aim to enhance the therapeutic effects of fixed appliances or interventions that aim to reduce the duration of orthodontic treatment.

Assessment of the success of orthodontic treatment generally involves evaluations of the patient's post-treatment records. However, without a valid and reliable evaluation method, treatment outcome assessments are difficult and often subjective. The American Board of Orthodontics (ABO) developed the Objective Grading System (OGS) for the precise evaluation of orthodontic treatment outcomes using the final dental casts and panoramic radiographs of patients.[6] The OGS rates eight criteria that contribute to ideal intercuspation and function. Best occlusion and alignment receive a score of 0 points, while deviations from ideal are given penalty points. Consequently, a high percentage of accordance can be achieved in both interexaminer and intraexaminer assessments, as reported in the orthodontic literature.[7] In addition to functioning as an objective clinical examination tool, the OGS is also used for the assessment of treatment progress and final outcomes with increased reliability, validity, and precision.[8] The ABO also developed the discrepancy index (DI) as a pretreatment scoring system, which has become an accepted and reliable index for the quantification of treatment complexity on the basis of orthodontic diagnostic records.[9]

A systematic evaluation of the range of typical treatment outcomes is crucial for the development of a standard of care[10] that can be used to judge the quality of orthodontic treatment.[11] To the best of our knowledge, no objective quality assessment using the ABO OGS has been performed in the field of orthodontics. Although previous systematic reviews have investigated the typical duration of orthodontic treatment,[12,13] they have not assessed the possible association between treatment duration and outcome, nor between treatment duration and initial discrepancy.

Therefore, the aim of this systematic review was to assess the occlusal outcomes and duration of fixed appliance orthodontic therapy from clinical trials in humans with the OGS of the ABO.

MATERIALS AND METHODS

Protocol and registration

The protocol for this systematic review was prepared *a priori* and registered in PROSPERO (CRD42016049203), and all *post hoc* changes were appropriately noted. This systematic review was conducted and reported in accordance with the Cochrane Handbook[14] and PRISMA statement,[15] respectively.

Eligibility criteria

We initially aimed to assess the comparative effectiveness of various orthodontic fixed appliances in terms of occlusal outcomes using parallel randomized and prospective nonrandomized trials in human patients. However, the pilot search indicated that very limited material was available (only two prospective trials); therefore, the review protocol was based on the inclusion of prospective or retrospective cohort studies assessing fixed appliance orthodontic treatment to provide an explorative overview of treatment outcomes (Appendix A). Studies where the OGS was not used or improperly used, nonclinical studies, and animal studies were excluded. Studies regarding novel orthodontic appliances with an unclear evidence base were excluded from the clinical part of the review but included in the explorative methodological overview.

Information sources and literature search

Nine electronic databases were systematically searched, without any limitations, from inception up to October 7, 2016 (Appendix B). Two additional sources, namely Google Scholar and the ISRCTN registry, and the reference/citation lists of included studies and relevant reviews were manually searched for additional studies or protocols. There were no limitations concerning language, publication year, or publication status.

Study selection and data collection

Titles identified from the search were screened by one author (SNP), and the corresponding abstracts/full texts were subjected to subsequent duplicate, independent checking using the eligibility criteria by a second author (DH), while conflicts were resolved by a third author (TE).

The characteristics of included studies and numerical data were extracted in duplicate by two authors (SNP, DH) using predetermined and piloted extraction forms. Missing or unclear information was requested from the authors of the studies.

Risk of bias in individual studies

The risk of bias in the included nonrandomized studies was assessed using the Downs and Black checklist[16] after initial calibration. Because the primary aim of this review

https://doi.org/10.4041/kjod.2017.47.6.401

was to provide an overview of possible OGS scores after orthodontic treatment, a main risk of bias assessment was included using the Downs and Black checklist for cohort studies. In a separate methodological overview of comparative cohort studies with two or more experimental groups, we also assessed whether confounding due to baseline differences in malocclusion severity measured using the DI between compared groups was appropriately addressed by matching or covariate adjustment.

Data synthesis: cohort studies

The outcome of fixed appliance treatment is bound to be affected by patient- and appliance-related characteristics.[3-5] Accordingly, a random-effects model proposed by Paule-Mandel[17] was deemed appropriate to incorporate this variability[18] because it outperforms the older DerSimonian and Laird estimator.[17] A weighted mean with the corresponding 95% confidence interval (CI) was calculated across studies for the primary and secondary outcome as a primary analysis. The produced forest plots were augmented with contours denoting the magnitude of the observed effects.[19]

Data synthesis: comparative cohort studies with at least two groups

The mean difference (MD) was used to pool the influence of reported treatment-related characteristics across included case–control studies. The effect of matching by initial discrepancy on the results of the meta-analyses was assessed by calculating the difference in MDs (ΔMD) between matched and nonmatched groups through random-effects meta-regression. Then, the absolute ΔMDs were pooled across comparisons using random-effects meta-analysis.

Heterogeneity

Absolute and relative between study heterogeneity were quantified using tau^2 and I^2 statistics, respectively. Relative heterogeneity was defined as the proportion of total variability in the results as explained by heterogeneity, not by chance. To quantify our uncertainty, 95% CIs were calculated for the heterogeneity statistics. Furthermore, 95% predictive intervals (95% PrI), which incorporate existing heterogeneity and provide a range of possible effects for a future clinical setting, were calculated for the meta-analyses of three or more studies.[20]

Risk of bias across studies and additional analyses

Indications for reporting biases (including small-study effects) were assessed using Egger's linear regression tests in meta-analyses of at least 10 studies. In cases of bias, robustness of the results was checked using subgroup sensitivity analyses according to precision.

We planned to seek possible sources of heterogeneity through prespecified random-effects meta-regressions with the Knapp and Hartung adjustment at the study level. These were based on the patient age, sex (% male patients), extraction rate, and mean baseline DI. In addition, a possible interrelation between the mean OGS score and treatment duration was investigated.

Sensitivity analyses were performed by dividing included cohort studies into (a) those that explicitly reported the use of only one-phase fixed-appliance treatment and (b) those that reported the use of two-phase treatment or those that did not provide clear reports. If considerable differences were identified between these subsamples, the subsample with clear reporting of one-phase fixed appliance treatment was used, because direct comparison between one- and two-phase treatment was neither possible nor within the scope of this study. All statistical analyses were performed using Stata SE 14.2 (Stata Corp, College Station, TX, USA) by one author (SNP). A two-tailed *p*-value of 0.05 was considered significant for hypothesis testing, although for heterogeneity testing and reporting bias testing, a value of 0.10 was considered significant because of low power.[21]

RESULTS

Study selection

A total of 480 and 23 papers were identified through electronic (Appendix B) and manual searches, respectively (Figure 1). After the removal of duplicates and

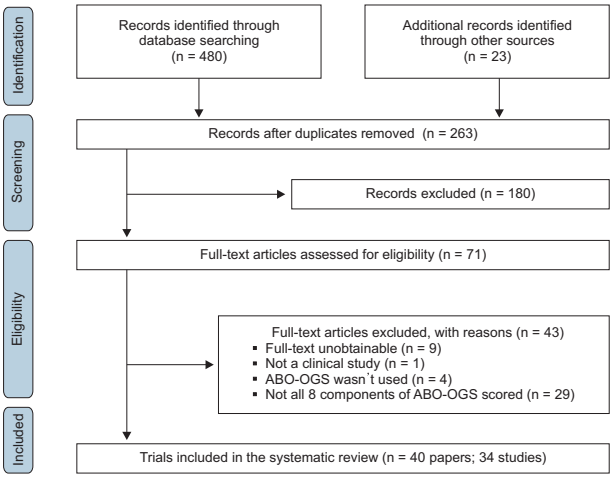

Figure 1. Study flowchart showing the identification and selection of eligible studies.
ABO-OGS, Objective Grading System (OGS) proposed by the American Board of Orthodontics.

 https://doi.org/10.4041/kjod.2017.47.6.401

Table 1. Characteristics of the studies included in our systematic review assessing the occlusal outcomes and duration of orthodontic fixed appliance treatment

No.	Study	Setting	Patient (n)	Sex, male (%)	Mean age (yr)	Extraction (%)	Malocclusion	DI	OGS	Tx time	Only FA	Factors
1	Junqueira, 2012; Mendes, 2012	BRA; uni Sao Paulo	68	56	NR	71	Cl. II/div1	No	Yes	No	No	Ex
2	Marques, 2012	BRA; various practices	60	37	17.4	NR	NR	No	Yes	Yes	No	Clinician
3	Li, 2015*	CHI; uni	76	36	32.2	Ex	Cl. I	Yes	Yes	Yes	Yes	Int*
4	Carvajal-Flórez, 2016	COL; uni Antioquia 10–11	34	44	20.9	NR	NR	No	Yes	Yes	No	Finishing†
5	Barbosa Lis, 2014	COL; uni Antioquia 11–12	39	48	NR	Ex/nonEx	NR	No	Yes	Yes	No	-
6	Rodríguez, 2014	COL; uni Manizales 11–12	31	NR	NR	NR	NR	No	Yes	No	Yes	-
7	Anthopoulou, 2014; Mislik, 2016	GRE; uni & 5 practices	55	35	14.9	45	Cl. I Ex/ nonEx	No	Yes	No	Yes	Clinician; Ex
8	Jain, 2013	IND; college Manipal	40	NR	16.6	Ex	NR	Yes	Yes	Yes	Yes	Prescription†
9	Soltani, 2012	IRA; 2 practices	60	23	NR	NonEx	Cl. I	No	Yes	No	Yes	Prescription†
10	Farhadian, 2005	IRA; uni	60	23	17.9	Ex/nonEx	Cl. I	No	Yes	Yes	Yes	Ex
11	Deguchi, 2011	JAP; uni 02–05	30	0	24.3	Ex	Open bite	Yes	Yes	No	Yes	Anchorage†
12	Deguchi, 2015	JAP; uni Okayama	25	20	24.2	Ex	Cl. II	Yes	Yes	Yes	Yes	Int*
13	Viwattanatipa, 2016	THA; TBO	200	NR	NR	NR	NR	Yes	Yes	No	Yes	-
14	Akinci Cansunar, 2014; Cansunar, 2014; Cansunar, 2016‡	TUR; 9 unis	1,098	40	16.3	49	Cl. II	Yes	Yes	Yes	Yes	1- or 2-phase; Ex
15	Brown, 2015	USA; practice Bandeen	64	50	13.8	NonEx	NR	Yes	Yes	Yes	Yes	Int*
16	Kuncio, 2007*	USA; practice Kuncio	11	9	26.8	NonEx	NR	No	Yes	Yes	Yes	-
17	Djeu, 2005*	USA; practice Shelton	48	NR	23.7	NonEx	NR	Yes	Yes	Yes	Yes	-
18	Alford, 2011*	USA; practice Snyder	63	51	17.8	NonEx	NR	Yes	Yes	Yes	No	-
19	Aszkler, 2014	USA; practice Wick	30	NR	NR	Ex/NonEx	NR	No	Yes	No	Yes	-
20	Chalabi, 2015	USA; uni Buffalo	50	38	NR	NR	Various	No	Yes	No	No	Gender
21	Brown, 2011	USA; uni Detroit Mercy 03–07	714	NR	NR	NR	NR	Yes	Yes	No	No	-
22	Yang-Powers, 2002	USA; uni Illinois	124	44	14.5	37	Various	No	Yes	Yes	No	Clinician
23	Knierim, 2006	USA; uni Indiana 01–03	437	42	17.8	28	Various	No	Yes	Yes	No	Early debond; Ex; malocclusion

Table 1. Continued

No.	Study	Setting	Patient (n)	Sex, male (%)	Mean age (yr)	Extraction (%)	Malocclusion	DI	OGS	Tx time	Only FA	Factors
24	Vu, 2008	USA; uni Indiana 04–06	455	39	16.3	28	Various	Yes	Yes	Yes	No	Canine impaction; compliance; DI; early debond; Ex; FFA; gender; HG; malocclusion; missed appointments; oral hygiene; orthognathic surgery; RME
25	Detterline, 2010	USA; uni Indiana 05–08	828	38	16.3	NonEx	NR	Yes	Yes	Yes	Yes	Slot size†
26	Park, 2008	USA; uni Indiana 97–07	200	31	15.2	25	Various	Yes	Yes	Yes	Yes	Finishing†
27	Pinskaya, 2004; Hsieh, 2005	USA; uni Indiana 98–00	521	42	16.0	31	Various	No	Yes	Yes	No	Ex
28	Campbell, 2007	USA; uni Indiana 98–03	382	45	15.3	NR	Various	Yes	Yes	Yes	No	Early debond; malocclusion
29	Deguchi, 2005	JAP; uni Okayama 02	72	26	18.7	58	Various	Yes	Yes	Yes	No	-
30	Schabel, 2008	USA; uni Michigan	48	NR	NR	NR	NR	No	Yes	No	No	-
31	Hoybjerg, 2013	USA; uni Oklahoma 02–10	90	46	15.2	50	Various	No	Yes	Yes	Yes	Ex
32	Santiago, 2012	USA; uni Puerto Rico 07–08	64	42	14.1	NR	NR	Yes	Yes	Yes	No	Compliance; gender
33	Ferguson, 2016*	USA; uni Saint Louis	28	NR	NR	NonEx	Cl. I	No	Yes	No	Yes	-
34	Sohrabi, 2016	USA; uni Washington 12–13	102	NR	16.4	17	Cl. I/II/III	No	Yes	Yes	No	-

Data modifications according to the eligibility of the included reports was as follows.
(i) Pulfer 2009 was excluded from the descriptives because it drew upon the data of Hsieh 2005 and Knierim 2006 to pool them together.
(ii) Junqueira 2012 and Mendes 2012 were judged to have mostly overlapping patients; only data from Mendes 2012 are reported, which were the more extensive of the two.
(iii) Anthopoulou 2014 and Mislik 2016 had overlapping patient populations where different factors were assessed. The demographics of Anthopoulou 2014 are reported here.
(iv) Akinci Cansunar 2014, Cansunar 2014, and Cansunar 2016 were judged to have mostly overlapping patients in their report. Data from Akinci Cansunar 2014 are reported here.
(v) Pinskaya 2004 and Hsieh 2005 were omitted as they included both labial and lingual appliances.
(vi) Only a subgroup of patients originating from the Okayama University was included from the Deguchi 2005 study, because the cohort from Indiana University was described in multiple other reports.
*Patient groups pertaining to treatment alternatives noneligible for this review (aligners, lingual appliances, computer- or corticotomy-assisted orthodontics) were excluded.
†Intervention groups were pooled and not separately assessed because of the retrospective nature of the included studies.
‡Some reported in different reports on the same cohort.
Ex, Extraction; DI, discrepancy index; OGS, Objective Grading System; Tx, treatment; FA, fixed appliances; uni, University; NR, not reported; Cl., class; div, division; Int, intervention; Ex, extraction treatment; Non-Ex, nonextraction treatment; FFA, fixed functional appliance; TBO, Thai Board of Orthodontics; HG, headgear; RME, rapid maxillary expansion.

initial screening, 71 papers were assessed using the eligibility criteria and 40 were included in our systematic review (Figure 1; Appendix C). In four instances, multiple publications pertaining to the same or overlapping patient cohorts were grouped together. Thus, a total of 34 studies were finally included in our systematic review.

Study characteristics

The characteristics of the included studies can be seen in Table 1. The 34 included studies originated from private practices or educational institutions from 10 different countries and included a total of 6,207 patients (median, 64 patients/study). There were 1966 (39.6%) male patients and 3,000 (60.4%) female patients with an average age of 18.4 years. Among the 34 included studies, 25 (73.5%) reported information about the inclusion or exclusion of tooth extractions; four included extraction patients, seven included non-extraction patients, and the remaining eleven studies had reported an average extraction rate of 40%, and three did not report the percentage of extractions. The treated malocclusions were often unspecified, and the DI was used to gauge the severity of the initial malocclusion in only 16 (47.1%) studies. In 18 (52.9%) studies, the authors explicitly stated that only one-phase treatment with fixed appliances was performed, while in the remaining 16 (47.1%) studies, two-phase treatment was performed for some of the included patients. All of the included studies measured the post-treatment OGS score, which was the primary outcome, while 23 (67.6%) studies also measured the treatment duration, which was the secondary outcome.

Risk of bias within studies

The risk of bias assessment for the 34 included studies is shown in Figure 2 and Appendix D–E. A high risk of bias for at least one domain was found in 31 studies (91.2%). The most problematic domains included the study design (where 85% studies were retrospective) and blinding (79% studies did not use blinding).

Data synthesis and additional analyses: cohort studies

A total of 29 (85.3%) of the 34 included studies could be used in the meta-analyses for the primary outcome (ABO OGS); the remaining either reported on overlapping patient populations or had missing data. The results of the random-effects meta-analysis indicated that the overall OGS score after treatment was 27.9 points (95% CI, 25.3–30.6 points) with high heterogeneity and no considerable differences between the subsample of studies that included strictly one-phase fixed appliance treatment (27.5 points; 95% CI, 24.5–30.5 points) and the subsample of studies reporting two-phase/unclear treatment (28.3 points; 95% CI 24.5–32.1 points; p for difference between subsamples > 0.1) (Table 2, Figure 3).

The meta-analysis of the 18 included studies reporting the secondary outcome of treatment duration indicated that the mean treatment duration among all studies was 24.9 months (95% CI, 24.6–25.1 months) with high heterogeneity (Figure 4). The average treatment duration differed significantly (p = 0.004) between the subsample of studies reporting one-phase fixed appliance treatment (24.8 months; 95% CI, 21.4–28.3 months) and the subsample of studies reporting two-phase/unclear treatment (31.6 months; 95% CI, 30.8–32.3 months). The difference in the mean duration between the

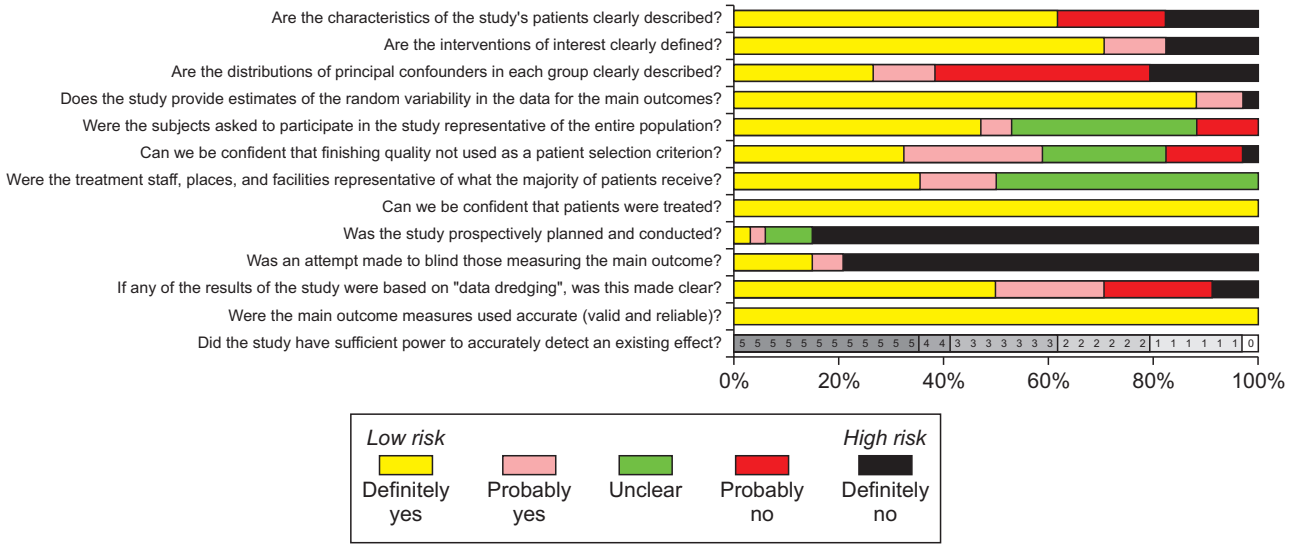

Figure 2. Summary of the risk of bias in the included studies.

https://doi.org/10.4041/kjod.2017.47.6.401

Table 2. Results of the meta-analyses for the primary (OGS score) and secondary (treatment duration) outcomes of orthodontic fixed appliance (FA) treatment

Outcome	Subgroup		Studies, n	Mean	95% CI	Tau² (95% CI)	I² (95% CI), %	95% PrI	Difference (95% CI)	*p*-value
ABO-OGS score	Overall		29	27.93	25.31–30.55	50.00 (34.11–50.00)	99.1 (98.7–99.1)	13.17–42.70		
	Only FA	No	14	28.29	24.53–32.06	50.00 (41.39–50.00)	99.1 (98.9–99.1)	12.33–44.26	−0.70 (−6.28 to 4.88)	0.800
		Yes	15	27.51	24.47–30.54	33.91 (17.90–50.00)	98.2 (96.7–98.8)	14.49–40.53		
Tx duration (mo)	Overall		16	24.86	24.60–25.12	0.00 (50.00–50.00)	0.0 (99.3–99.3)	24.58–25.14		
	Only FA	No	8	31.56	30.79–32.33	0.00 (49.03–50.00)	0.0 (96.7–96.7)	30.60–32.52	−13.22 (−21.60 to −4.84)	0.004
		Yes	8	24.84	21.41–28.27	23.36 (10.81–50.00)	99.2 (98.3–99.6)	12.26–37.42		

OGS, Objective Grading System; CI, confidence interval; PrI, predictive interval; ABO; American Board of Orthodontics; Tx, treatment.

two treatment subsamples was 13.2 months (95% CI, 4.8–21.6 months), although considerable heterogeneity remained even after the separate analysis.

Meta-regressions failed to identify a significant influence of any study-level characteristics on the primary outcome of OGS score or the secondary outcome of treatment duration (Appendix F). However, significant signs of reporting bias (Appendix G) were identified for the secondary outcome of treatment duration through Egger's test ($p = 0.031$), where small/imprecise studies tended to report longer treatment durations compared with the remaining studies (Appendix G). Stratified subgroup analyses according to study precision indicated that bias was mainly concentrated in the subgroup of studies reporting two-phase or unclear treatment (Appendix G), while the subgroup of studies reporting one-phase fixed appliance treatment was relatively robust (Egger's test, $p > 0.05$). Finally, we could not perform sensitivity analyses on the basis of risk of bias in the included studies, because most of them (91%) had a high risk of bias.

Data synthesis and additional analyses: comparative cohort studies with at least two groups

Signs of discordant results (i.e., significant differences between subgroups; Table 2) and reporting bias (Appendix G) were found for the subgroup of studies with two-phase/unclear treatment. Therefore, factors from comparative two-group cohort studies were assessed only for those studies that strictly reported one-phase fixed appliance treatment, which were free from bias (Table 3). Orthodontic treatment with extraction of four premolars was associated with a slight improvement in occlusal outcomes, as indicated by the OGS score (MD, −4.9 points; 95% CI, −11.8 to 1.9 points; $p = 0.159$), and a moderate increase in the treatment duration (MD, 6.4 months; 95% CI, 1.4 to 11.5 months; $p = 0.013$). However, only the increase in treatment duration was statistically significant at the 5% level. Finally, no considerable differences in occlusal outcomes could be found between patients treated in the orthodontic department at a university and those treated in a private orthodontic clinic.

Methodological overview

Additionally, the methodological status of all available comparisons included in the studies identified from this systematic review was assessed, regardless of whether they were eligible for the clinical part of the systematic review (Table 4). From the 26 comparisons regarding various treatment factors reported in the included studies, 10 (38.5%) used matching to form patient groups that were comparable in terms of the severity of the baseline malocclusion. However, in one case, the pre-treatment ABO OGS score was used to match the severity of the baseline pre-treatment malocclusion, and this was identified as problematic. In four (15.4%) of the 26 identified comparisons, the severity of the baseline malocclusion in the compared groups was considered by using it as a covariate in the statistical analyses. Overall, baseline confounding was adequately assessed, in one way or the other, in only 10 (38.5%) of the included comparisons.

Among the available comparisons, two included both matched and nonmatched studies and enabled an assessment of the influence of matching on the

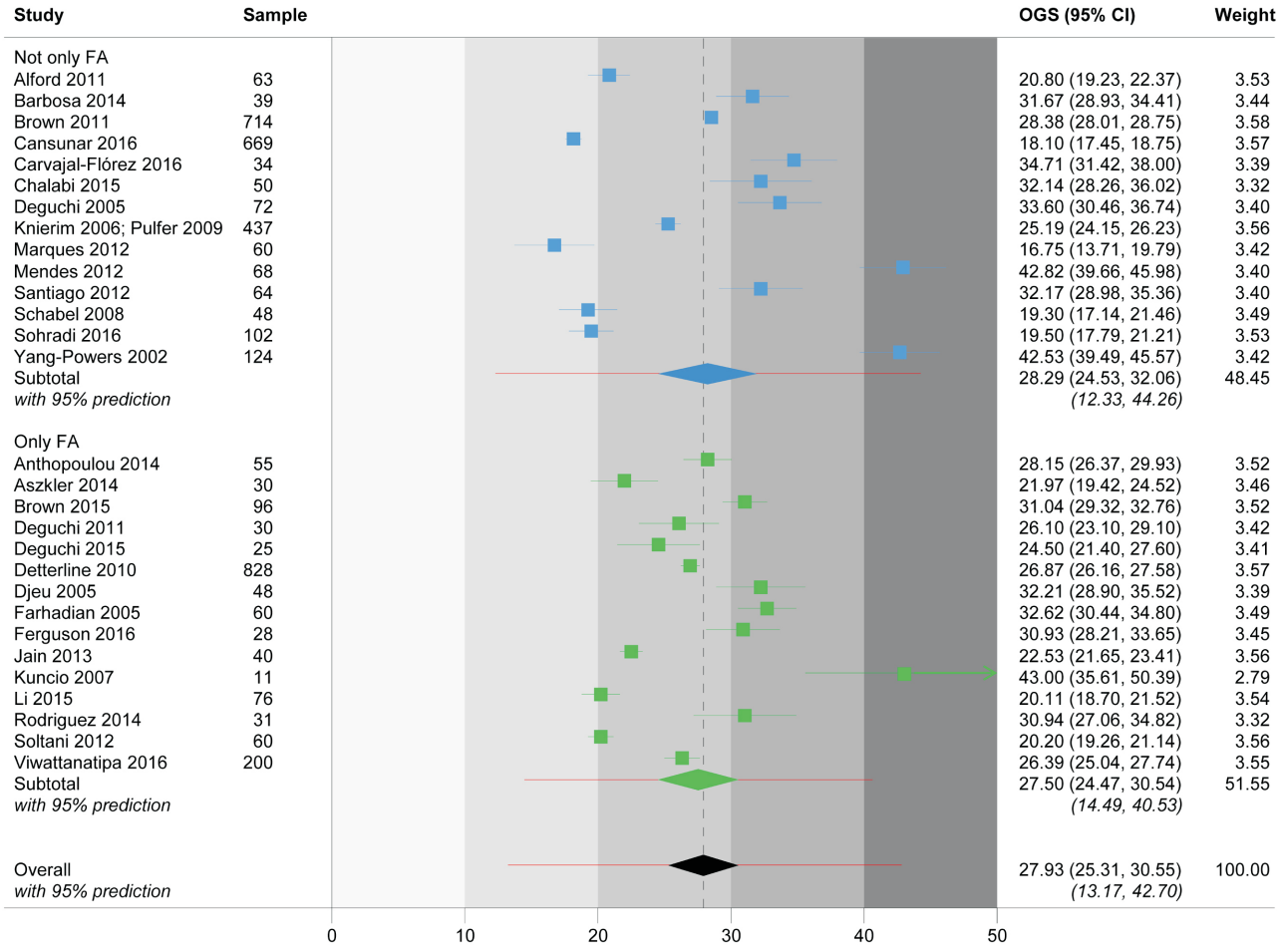

Figure 3. Overall pooling for occlusal outcomes of fixed appliance (FA) treatment assessed using the Orthodontic Grading System proposed by the American Board of Orthodontics Mean Orthodontic Grading System scores and their corresponding 95% confidence intervals (CIs) for each included study are given as boxes with horizontal lines, respectively. The weighted pooled summary estimates with and their corresponding 95% CIs for the two subgroups or overall are given as diamonds. Horizontal lines at the diamonds represent the 95% prediction that gives a range of possible values to be clinically seen, while incorporating existing heterogeneity.

results (Appendix H). In the comparison of aligner versus fixed appliance treatment, studies with matched patient samples tended to find considerably greater differences in occlusal outcomes. Moreover, studies with baseline matching tended to find considerably smaller differences in occlusal outcomes between extraction and nonextraction treatment groups compared with studies without matching. Finally, the absolute pooled difference in the OGS score between matched and nonmatched patient samples across studies was calculated as ΔMD = 7.20 OGS points (95% CI, −2.16 to 16.57 points; p = 0.132; Appendix I). This could possibly have clinical implications, although evidence was very limited.

DISCUSSION

Summary of evidence

This systematic review summarizes evidence from 34 clinical cohort studies including a total of 6,207 patients who received comprehensive orthodontic fixed appliance treatment. The pooled analysis for the primary outcome, which was occlusal outcomes as measured using the OGS score, indicated an average OGS score of 27.9 OGS points (95% CI, 25.3–30.6 points), which was relatively consistent regardless of one-phase or two-phase treatment (p = 0.800; Table 2).

Analysis of the secondary outcome, which was the treatment duration, revealed an average treatment duration of 24.9 months (95% CI, 24.6–25.1 months). However, a considerable difference of 13.2 months

https://doi.org/10.4041/kjod.2017.47.6.401

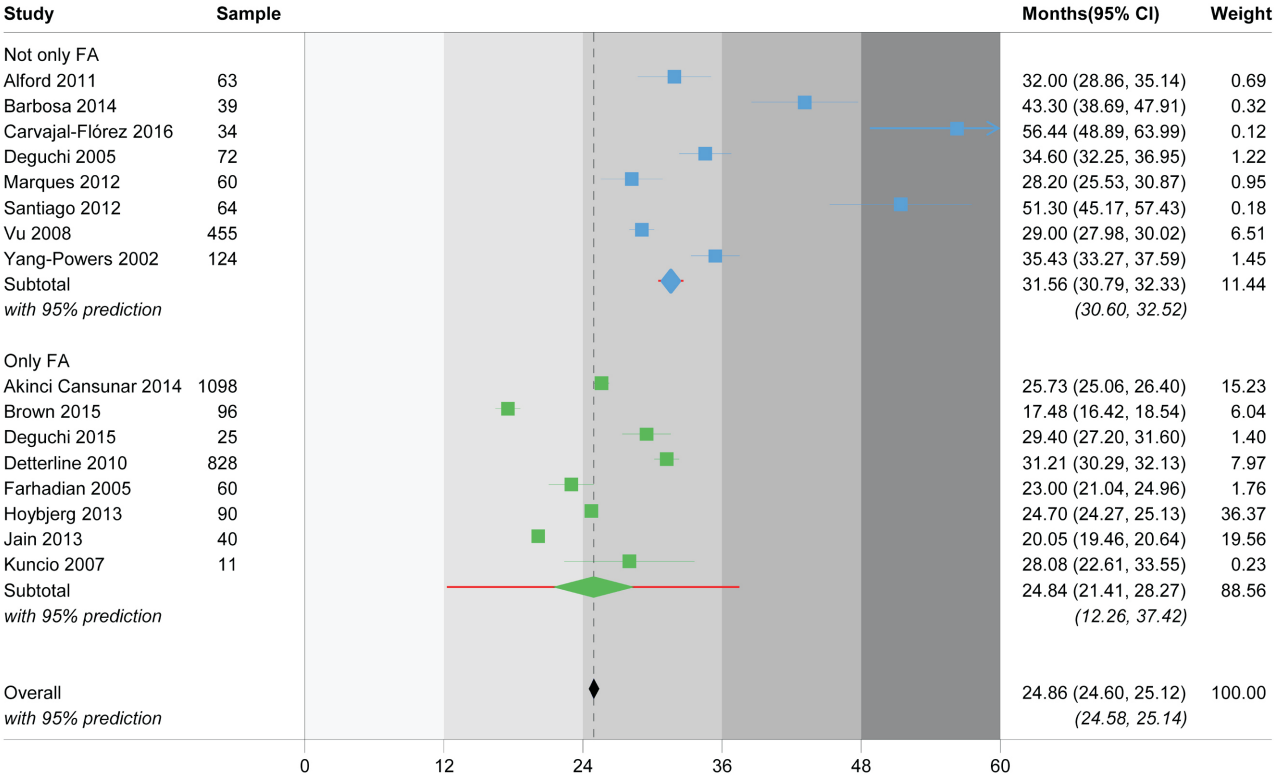

Figure 4. Overall pooling for the fixed appliance (FA) treatment duration in months. Mean treatment durations and their corresponding 95% confidence intervals (CIs) for each included study are given as boxes with horizontal lines, respectively. The weighted pooled summary estimates with and their corresponding 95% CIs for the two subgroups or overall are given as diamonds. Horizontal lines at the diamonds represent the 95% prediction that gives a range of possible values to be clinically seen, while incorporating existing heterogeneity.

(4.8–21.6 months; p = 0.004; Table 2) in treatment duration was found between studies that strictly reported one-phase fixed appliance treatment and those that reported two-phase or unclear treatment. Therefore, this systematic review focuses on the clearly defined subsample of studies on one-phase fixed appliance treatment with an average treatment duration of 24.8 months (95% CI, 21.4–28.3 months). This is slightly higher than the average treatment duration of 19.9 months reported by Tsichlaki et al.[22] However, a fixed-effect model was used by the authors of that study, which cannot be easily justified in such a broad clinical scenario[18] and could, in combination with the reported statistical inconsistency, have had a profound impact on the meta-analysis results.[19] In addition, contrary to the studies included in the previous systematic review of Tsichlaki et al.,[22] the studies assessed in the present review included the assessment of final occlusal outcomes using the ABO OGS in their scope and could possibly have paid extra attention to the finishing stage of orthodontic treatment. Finally, the number of studies included was considerably higher in the previous systematic review than in the present review (22 and 8 studies, respectively), because the use of the ABO OGS was not an inclusion criterion in the former.

Interestingly, we found no association of the average outcome of orthodontic treatment with the mean treatment duration, mean severity of the initial malocclusion as assessed using the DI, and various patient- or treatment-related characteristics (Appendix F). Although this is in agreement with the findings of two included studies[23,24] that found nonsignificant correlation coefficients of −0.18 to −0.30 for the association between the ABO OGS and DI, this does not mean that the DI is not a crucial component of the ABO OGS framework in clinical investigations of treatment effects.

The only factor that appeared to considerably influence the outcomes of orthodontic treatment was the inclusion of tooth extractions. First, on a study level, the mean OGS score was significantly associated with the extraction rate in each study (Appendix F). On an average, every 10% increase in the extraction rate was significantly associated with a decrease in the OGS

Table 3. Results of the meta-analyses regarding the effect of characteristics from included comparative case–control studies reporting one-phase fixed appliance treatment on the primary (OGS score) and secondary outcome (treatment duration)

Factor	ABO OGS score							Treatment duration						
	n	MD	95% CI	Tau^2 (95% CI)	I^2 (95% CI), %	95% PrI	*p*	n	MD	95% CI	Tau^2 (95% CI)	I^2 (95% CI), %	95% PrI	*p*
Extraction of four premolars	2	−4.94	−11.82 to 1.94	21.02 (NA)	85.3 (NA)	NA	0.159	2	6.41	1.38–11.45	11.21 (NA)	83.6 (NA)	NA	0.013
Private practice vs. university clinic	1	0.5	−3.77 to 4.77	-	NA	NA	0.818	-	-	-	-	-	-	-

ABO, American Board of Orthodontics; OGS, Objective Grading System; n, number of studies; MD, mean difference; CI, confidence interval; PrI, predictive interval; NA, not applicable.

score by 0.7 point, which indicated better occlusal outcomes. In addition, analysis of within-study data from case-control studies indicated that comprehensive treatment involving extraction of the four premolars was associated with improved treatment outcomes, as indicated by a decrease in the OGS score (MD, −4.9 OGS points; 95% CI, −11.8 to 1.9 OGS points; p = 0.159), and a prolonged treatment duration (MD, 6.4 months; 95% CI, 1.4 to 11.5 months; p = 0.013). Although only the increase in the treatment duration was statistically significant at the 5% level, this was most probably due to imprecision caused by a small sample size; the addition of future studies may rectify this.

Finally, a methodological overview was conducted of all identified clinical case-control studies that assessed occlusal outcomes according to various treatment-related factors (Table 4). This also included study arms that assessed novel interventions (aligners and individualized or lingual appliances) that were excluded from the clinical part of the systematic review because of their basic design.[25,26] The results indicated that the majority of studies neither matched the compared patient groups according to their baseline malocclusion severity nor used the baseline malocclusion severity as a covariate in the statistical analyses (Table 4). As a result, only 10 (38.5%) of the available comparisons were free from baseline confounding. This might be important, as meta-epidemiological analysis indicated that matching of experimental groups according to the baseline malocclusion severity may considerably influence the observed results (Appendixes H and I).

Some additional methodological flaws were found among the included studies. First, a large number (n = 29) of possibly relevant clinical studies identified from the literature search did not assess all eight components of the ABO OGS and were consequently excluded from the present review because of pooling incompatibility. Second, an included study used the ABO OGS to measure the baseline malocclusion severity and match the compared groups,[27] and this contradicts the rationale behind this index which might be problematic[28,29] and does not justify substitution of the DI.[9] Finally, some included studies measured the baseline severity with the DI and performed statistical tests to determine baseline differences in DI among the compared groups. This practice is inherently wrong[30] because the results can be easily distorted by increasing the sample size; furthermore, it cannot substitute proper matching or covariate adjustment.

The strengths of this systematic review include the *a priori* registration in PROSPERO, the extensive unrestricted literature search, which included studies in languages other than English, the use of robust methodology pertaining to the qualitative and

https://doi.org/10.4041/kjod.2017.47.6.401

Table 4. Methodological overview of comparison-specific characteristics obtained from of identified case-control studies

No	Study*	Intervention	Comparison	Initial malocclusion severity			Selection bias addressed?
				Matching	How	Used as covariate	
1	Detterline, 2010	18”-slot CB	22”-slot CB	Yes	DI	Yes	Yes
2	Ferguson, 2016	Corticotomy & CB	CB	No	-	No	
3	Anthopoulou, 2014	Ex	Non-Ex	Yes	Discriminant analysis	No	Yes
4	Farhadian, 2005	Ex	Non-Ex	No	-	No	No
5	Hoybjerg, 2013	Ex	Non-Ex	No	-	No	No
6	Knierim, 2006	Ex	Non-Ex	No	-	No	No
7	Pinskaya, 2004	Ex	Non-Ex	No	-	No	No
8	Sohradi, 2016	Ex	Non-Ex	No	-	No	No
9	Vu, 2008	Ex	Non-Ex	No	-	No	No
10	Akinci Cansunar, 2014	Ex	Non-Ex	No	-	No	No
11	Carvajal-Flórez, 2016	Finishing protocol	Conventional finishing	No	-	No	No
12	Park, 2008	Finishing protocol	Conventional finishing	Yes	DI	Yes	Yes
13	Deguchi, 2011	Skeletal anchorage	No skeletal anchorage	No	–(ΔDI = 14.5)	No	No
14	Brown, 2015	Indirect bracket placement	Direct bracket placement	Yes	DI	No	Yes
15	Brown, 2015	Insignia appliance	CB	Yes	DI	No	Yes
16	Djeu, 2005	Invisalign aligners	CB	Yes	DI	No	Yes
17	Kuncio, 2007	Invisalign aligners	CB	No	-	No	No
18	Li, 2015	Invisalign aligners	CB	Yes	DI + RCT	No	Yes
19	Deguchi, 2015	Lingual appliance	CB	Yes	DI	No	Yes
20	Jain, 2013	MBT prescription CB	Roth prescription CB	No	–(ΔDI = 3.8)	Yes	Yes
21	Soltani, 2012	MBT prescription CB	SE prescription CB	No	-	No	No
22	Marques, 2012	Orthodontist	General dentist	Yes	ABO OGS	No	No
23	Hoybjerg, 2013	3 Retention protocols	-	No	-	No	No
24	Alford, 2011	SureSmile appliance	CB	No	ΔDI = 2.6	Yes	Yes
25	Mislik, 2016	University clinic	Private practice	Yes	Discriminant analysis	No	Yes
26	Yang-Powers, 2002	University clinic	ABO-submitted cases	No	-	No	No

*Mendes et al. (2012) was excluded because patients were matched in terms of the final ABO OGS score.
CB, Conventional brackets; DI, discrepancy index; Ex, extraction treatment; Non-Ex, nonextraction treatment; RCT, randomized clinical trial; MBT, McLaughlin–Bennett–Trevisi; SE, standard edgewise; ABO, American Board of Orthodontics; OGS, Objective Grading System.

quantitative synthesis of data,[31] transparent reporting of quantitative data for all outcomes from the included studies, the use of the new robust Paule–Mandel random-effects estimator,[17] and the use of subgroup, meta-regression, and sensitivity analyses to check the robustness of the results. However, some limitations cannot be overlooked. First and foremost, this systematic review included mostly observational, nonrandomized, retrospective clinical studies, and this is bound to have influenced the results of the meta-analyses.[25] Therefore, we planned *a priori* not to focus on the comparative effectiveness of various interventions, considering it would require experimental prospective controlled studies. Instead, we provided an overview of expected treatment outcomes and possible influencing factors and assessed methodological issues in existing studies.

Finally, as considerable heterogeneity was found in both the primary and secondary outcomes, which remained unexplained even after the investigation of possible sources through subgroup, meta-regression, and sensitivity analyses, readers should be cautioned that the pooled estimates of the meta-analyses may be imprecise. Clinicians are instead advised to base their conclusions on the range of possible values indicated by the 95% CIs and 95% PrIs to identify cases that deviate from these ranges.

CONCLUSION

Recommendations for clinical practice

With regard to clinical relevance, this systematic review cannot provide robust evidence on the comparative effectiveness of various interventions. The range of expected occlusal outcomes and treatment duration are provided on the basis of the identified studies, and clinicians are advised to consider these two in conjunction and take care to identify cases with extreme deviations from this range. Comprehensive treatment with extraction of the four premolars may be associated with possibly improved occlusal outcomes and a longer treatment duration than non-extraction treatment. However, the available evidence is limited and not free from bias.

Recommendations for future research

The use of the ABO OGS can be very helpful for objective evaluation and comparison of the occlusal outcomes of orthodontic treatment with different fixed appliances, as well as several surgical and nonsurgical treatment outcomes through randomized controlled trials. Furthermore, researchers should consider both occlusal outcomes and the treatment duration in their trials to draw robust conclusions regarding the treatment efficiency. Researchers comparing various interventions should match compared patients according to the severity of the baseline malocclusion using the DI or any other robust method. Finally, covariate adjustment according to the severity of the baseline malocclusion can aid in achieving the most reliable statistical estimates[30] and improving their statistical power.[32] However, it must be stressed that *post hoc* matching of compared patients does not substitute proper prospective trial planning and cannot alleviate the inherent biases that can be found in nonrandomized and, particularly, retrospective study designs.[25,26,33]

SUPPLEMENTARY MATERIALS

Appendices A–I is available at https://doi.org/10.4041/kjod.2017.47.6.401.

REFERENCES

1. Pandis N, Polychronopoulou A, Eliades T. Active or passive self-ligating brackets? A randomized controlled trial of comparative efficiency in resolving maxillary anterior crowding in adolescents. Am J Orthod Dentofacial Orthop 2010;137:12.e1-6; discussion 12-3.
2. Pandis N, Polychronopoulou A, Katsaros C, Eliades T. Comparative assessment of conventional and self-ligating appliances on the effect of mandibular intermolar distance in adolescent nonextraction patients: a single-center randomized controlled trial. Am J Orthod Dentofacial Orthop 2011;140:e99-105.
3. Papageorgiou SN, Konstantinidis I, Papadopoulou K, Jäger A, Bourauel C. A systematic review and meta-analysis of experimental clinical evidence on initial aligning archwires and archwire sequences. Orthod Craniofac Res 2014;17:197-215.
4. Papageorgiou SN, Konstantinidis I, Papadopoulou K, Jäger A, Bourauel C. Clinical effects of pre-adjusted edgewise orthodontic brackets: a systematic review and meta-analysis. Eur J Orthod 2014;36:350-63.
5. Papageorgiou SN, Gölz L, Jäger A, Eliades T, Bourauel C. Lingual vs. labial fixed orthodontic appliances: systematic review and meta-analysis of treatment effects. Eur J Oral Sci 2016;124:105-18.
6. Casko JS, Vaden JL, Kokich VG, Damone J, James RD, Cangialosi TJ, et al. Objective grading system for dental casts and panoramic radiographs. American Board of Orthodontics. Am J Orthod Dentofacial Orthop 1998;114:589-99.
7. Mislik B, Konstantonis D, Katsadouris A, Eliades T. University clinic and private practice treatment outcomes in Class I extraction and nonextraction patients: a comparative study with the American Board of Orthodontics Objective Grading System. Am J Orthod Dentofacial Orthop 2016;149:253-8.
8. Pinskaya YB, Hsieh TJ, Roberts WE, Hartsfield JK. Comprehensive clinical evaluation as an outcome assessment for a graduate orthodontics program. Am J Orthod Dentofacial Orthop 2004;126:533-43.
9. Cangialosi TJ, Riolo ML, Owens SE Jr, Dykhouse VJ, Moffitt AH, Grubb JE, et al. The ABO discrepancy index: a measure of case complexity. Am J Orthod Dentofacial Orthop 2004;125:270-8.
10. Dougherty HL. The orthodontic standard of care. Am J Orthod Dentofacial Orthop 1991;99:482-5.
11. Vig KWL, Firestone A, Wood W, Lenk M. Quality of orthodontic treatment. Semin Orthod 2007;13:81-7.
12. Mavreas D, Athanasiou AE. Factors affecting the duration of orthodontic treatment: a systematic review. Eur J Orthod 2008;30:386-95.
13. Tsichlaki A, Chin SY, Pandis N, Fleming PS. How

long does treatment with fixed orthodontic appliances last? A systematic review. Am J Orthod Dentofacial Orthop 2016;149:308-18.
14. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [Internet]. The Cochrane Collaboration; 2011 [updated 2011 Mar; cited 2016 Dec 3]. Available from: http://handbook.cochrane.org/.
15. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009;62:e1-34.
16. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998;52:377-84.
17. Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res Synth Methods 2016;7:55-79.
18. Papageorgiou SN. Meta-analysis for orthodontists: Part I--How to choose effect measure and statistical model. J Orthod 2014;41:317-26.
19. Papageorgiou SN. Meta-analysis for orthodontists: Part II--Is all that glitters gold? J Orthod 2014; 41:327-36.
20. IntHout J, Ioannidis JP, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open 2016;6:e010247.
21. Ioannidis JP. Interpretation of tests of heterogeneity and bias in meta-analysis. J Eval Clin Pract 2008; 14:951-7.
22. Tsichlaki A, Chin SY, Pandis N, Fleming PS. How long does treatment with fixed orthodontic appliances last? A systematic review. Am J Orthod Dentofacial Orthop 2016;149:308-18.
23. Djeu G, Shelton C, Maganzini A. Outcome assessment of Invisalign and traditional orthodontic treatment compared with the American Board of Orthodontics objective grading system. Am J Orthod Dentofacial Orthop 2005;128:292-8; discussion 298.
24. Viwattanatipa N, Buapuean W, Komoltri C. Relationship between discrepancy index and the objective grading system in thai board of orthodontics patients. Orthod Waves 2016;75:54-63.
25. Papageorgiou SN, Xavier GM, Cobourne MT. Basic study design influences the results of orthodontic clinical investigations. J Clin Epidemiol 2015; 68:1512-22.
26. Papageorgiou SN, Koretsi V, Jäger A. Bias from historical control groups used in orthodontic research: a meta-epidemiological study. Eur J Orthod 2017;39:98-105.
27. Marques LS, Freitas Junior Nd, Pereira LJ, Ramos-Jorge ML. Quality of orthodontic treatment performed by orthodontists and general dentists. Angle Orthod 2012;82:102-6.
28. CoHong M, Kook YA, Baek SH, Kim MK. Comparison of treatment outcome assessment for class I malocclusion patients: peer assessment rating versus American board of orthodontics-objective grading system. J Korean Dent Sci 2014;7:6-15.
29. Hong M, Kook YA, Kim MK, Lee JI, Kim HG, Baek SH. The Improvement and Completion of Outcome index: A new assessment system for quality of orthodontic treatment. Korean J Orthod 2016;46:199-211.
30. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med 2002;21:2917-30.
31. Papageorgiou SN, Papadopoulos MA, Athanasiou AE. Reporting characteristics of meta-analyses in orthodontics: methodological assessment and statistical recommendations. Eur J Orthod 2014;36:74-85.
32. Hernández AV, Steyerberg EW, Habbema JD. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. J Clin Epidemiol 2004;57:454-60.
33. Papageorgiou SN, Kloukos D, Petridis H, Pandis N. Publication of statistically significant research findings in prosthodontics & implant dentistry in the context of other dental specialties. J Dent 2015;43:1195-202.