European Journal of Orthodontics, 2016, 32–38
doi:10.1093/ejo/cjv004
Advance Access publication 4 March 2015

Original article

Statistical analysis in orthodontic journals: are we ignoring confounding?

Anastasia Spanou*, Despina Koletsi, Padhraig S. Fleming***, Argy Polychronopoulou****and Nikolaos Pandis*****

*Laboratory of Research of the Musculoskeletal System, School of Medicine, University of Athens, Athens, Greece, **Department of Orthodontics, School of Dentistry, University of Athens and Private Practice in Athens, Greece, ***Oral Growth and Development, Barts and The London School of Medicine and Dentistry, Institute of Dentistry, Queen Mary University of London, London, UK ****Department of Preventive and Community Dentistry, School of Dentistry, University of Athens, Athens, Greece, *****Department of Orthodontics and Dentofacial Orthopedics, Dental School/Medical Faculty, University of Bern, Switzerland and Private Practice in Corfu, Greece

Correspondence to: Nikolaos Pandis, Private Practice, 29 p. Zafiropoulou Street, Corfu 49100, Greece. E-mail: npandis@yahoo.com

Summary

Aim: To assess the prevalence of adjustment for confounding within statistical analysis and matching at the design stage in leading orthodontic journals and to explore potential associations between accounting for confounding and publication characteristics.

Materials and methods: Twenty-four issues of four leading orthodontic journals with the highest impact factor were searched from July 2014 backwards. Confounding adjustment through statistical analysis and study characteristics including journal, study design, region of origin, number of authors, number of centres, involvement of a statistician, significance of results, and type of analysis were recorded. Reporting of matching at the design stage was also recorded.

Results: Of 426 studies identified, only 71 (17 per cent) accounted for confounding in the statistical analysis. There was evidence that journal, country of authorship, and involvement of a statistician (odds ratio = 3.91, 95 per cent confidence interval: 2.16–7.10; $P < 0.001$) were significant predictors of accounting for confounding at the analysis level. Reporting of matching at the design stage was identified in 111 of 426 (26 per cent) studies in which 9 studies adjusted for confounding at the analysis level.

Conclusions: Appropriate adjustment for confounding in orthodontic literature either at the design or at the analysis stage was identified in less than half of studies overall (41 per cent), suggesting lack of expertise and awareness in design, conduct, analysis, and reporting of non-randomized studies in this field. This is a critical limitation that can potentially result in biased estimates and associations between examined exposures and outcomes.

Introduction

Orthodontic research is composed of both non-randomized and randomized studies (1) with non-randomized studies predominating (2, 3). Non-randomized studies are also commonly termed observational studies and include studies such as cohort, case–control studies, cross-sectional studies, and ecological studies.

Non-randomized studies are placed at a lower level on the evidence pyramid due to limitations in their design. Common issues hampering observational studies include selection bias, information bias, and confounding (1). In particular, confounding can overestimate or underestimate the association between an exposure and an outcome. A factor can be considered a confounder if it is associated with the exposure of interest, is an independent outcome predictor, and is not in the causal pathway between the exposure and the outcome (Figure 1). For example, patient age can be a confounder in a study aiming to investigate the

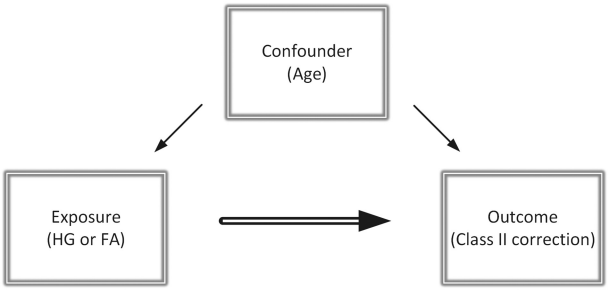

Figure 1. Age is a potential confounder of the association between the exposure (type of Class II corrector) and outcome (Class II correction).

association between skeletal Class II correction (outcome) and the type of appliance used [either functional appliance (FA) or headgear (HG)] (exposure). Let us assume that age differs among the exposed groups, with FAs provided to younger patients. Age is also a risk factor associated with the outcome and not in the causal pathway between exposure and outcome. If we accept that younger patients are more responsive to Class II mechanics and that younger participants predominate in the FA group compared to the HG group, then it is possible that the effect of the FA is overestimated and that of the HG underestimated.

In randomized controlled trials (RCTs), the problem of confounding is handled via the process of randomization, which aims to create treatment groups that differ only with regard to the allocated intervention. Randomization, therefore, limits confounding by equalizing known and unknown confounders among the comparison groups, notwithstanding chance happenings. RCTs are considered the gold standard for assessing the efficacy and safety of the intervention of interest; however, they are not always possible or ethical. In the latter instances, non-randomized studies may be appropriate. Consequently, observational studies are used extensively to describe the distribution of disease and exposure in populations and can also be useful for hypotheses generation and testing.

Non-randomized studies are more prone to systematic biases than RCTs, while it is also more difficult to make causal inference concerning the effect of an intervention (4, 5). Consequently, in non-randomized studies, it is important that the role and potential sources of bias and confounding are appraised before commencing the study. Failure to do this risks an inability to discount alternative explanations for any association (or otherwise) observed. Potential confounders may be identified by carrying out a detailed literature review. For example, in a cohort study investigating the skeletal response to FA therapy, a literature search would reveal a negative association related to condylion–gonion–menton (Co–Go–Me) angulation (6); therefore, either Co–Go–Me or a proxy, such as maxillo-mandibular planes angle (MMPA), may be used in an adjusted analysis.

Once confounding variables have been identified, their effects can be mitigated by carrying out matching at the design stage. For example, in a comparative design, participants within either group may be matched in respect of MMPA. Matching may, however, introduce further issues such as difficulty in identifying suitable controls. Consequently, statistical techniques to address the effects of confounding are recommended in non-randomized studies. Approaches include assessment of the association between the exposure and outcome within different levels (or strata) of a confounding factor, known as stratification. The basis for stratification is that individuals within each stratum are similar with respect to the confounding factor; therefore, the measure of effect obtained for each stratum would not be affected by the confounder. However, this approach may result in residual confounding as differences in respect of confounders are inevitable.

Where the observed measures of effect do not differ markedly across the strata of the potential confounder, it is preferable to combine these into a single summary estimate, rather than present a separate value for each stratum as the data within each stratum may be thin, making the individual estimates imprecise. Moreover, it is better to summarize data as succinctly as possible to facilitate interpretation. This adjusted or pooled measure of effect represents the best estimate of the association between the exposure and outcome, controlling for the confounding factor. There is no statistical test to pinpoint whether a factor is necessarily a confounder. Instead, judgement is required to compare the adjusted measure of effect against the crude estimate. If the difference between these is considered important, confounding is likely to be present. Conversely, if the estimates appear similar, this would imply that confounding is not present and an adjusted analysis is unnecessary. There are no universally agreed rules for making this decision. Overlap of the confidence intervals for the adjusted and unadjusted estimates may be assessed; it is also suggested that the estimates should differ by at least 10 per cent for confounding to be suspected (7).

Clearly, accounting for potential confounders at the design and/or analysis stage is an important step when analysing and interpreting data from non-randomized studies. There is no previous report on the handling of confounding within statistical analyses in dental journals. The objectives of this study were therefore to assess the prevalence of adjustment for confounding within statistical analysis within a subset of leading orthodontic journals and to explore the relationship between statistical testing and variables including the journal and region of publication and study design. Further, reporting on matching between study groups at the design stage was also recorded.

Materials and methods

The most recent 24 issues of the 4 orthodontic journals with the highest impact factor, namely *American Journal of Orthodontics and Dentofacial Orthopedics* (*AJODO*), *Angle Orthodontist* (*AO*), *European Journal of Orthodontics* (*EJO*), and *Orthodontics and Craniofacial Research* (*OCR*), were accessed by searching their respective electronic archives from July 2014 backwards.

Original studies other than editorial, commentaries, and special type articles (i.e. book reviews and resident's journal reviews) were screened for eligibility. Systematic reviews or other types of reviews, case reports, and one-group (non-comparative) studies were further excluded. RCTs were also excluded as this design is assumed largely to account for potential confounders. One of the authors (AS) screened all titles, abstracts, and, if necessary, full texts to identify eligible studies. Initial piloting of data collection was done for 25 per cent of eligible studies. A second author (DK) was consulted in cases of uncertainty with disagreements resolved through discussion.

Data were extracted for the following variables:

1. Journal type.
2. Type of study design (prospective clinical trial, retrospective cohort, cross-sectional, retrospective case–control).
3. Orthodontic-related topic.
4. Continent of authorship and number of centres (based on authors' affiliations).
5. Number of authors involved.
6. Involvement of a methodologist/statistician (also based on information on affiliations).
7. Significance/non-significance of the results (with respect to primary outcomes).
8. Type of analysis conducted (the most complex analysis addressing the primary outcome was considered).

9. Accounting for confounding factors through statistical analysis.
10. Reporting of matching between study groups.

Descriptive statistics on the characteristics of the included studies and cross-tabulations were conducted to investigate associations between confounding adjustment through statistical analysis and study characteristics. Chi-square and Fisher's exact tests were applied where appropriate. Univariable and multivariable logistic regression was undertaken to identify potential predictor variables of adjusting for confounding factors or otherwise in the statistical analysis, including journal, study design, continent, number of authors, number of centres, statistician involvement, and statistical significance of primary outcome. The model fit was assessed using the Hosmer–Lemeshow test. The level of significance was pre-specified at $P < 0.05$. All analyses were performed with Stata version 12.0 software (Stata Corporation, College Station, Texas, USA).

Results

Four hundred and twenty-six studies were included, with only 71 (17 per cent) of those having considered confounding in their statistical analysis (Figure 2). The *OCR* (7/24, 29 per cent), followed by the *AJODO* (20/88, 23 per cent) had the highest proportion of articles using adjusted analysis to account for confounders in the statistical analysis. Retrospective case–control studies (3/8, 38 per cent), although poorly represented in the subset of orthodontic articles, and cross-sectional studies (44/219, 20 per cent) were more likely to involve adjusted analyses compared to prospective clinical trials (11/105, 10 per cent) or retrospective cohort studies (13/94, 14 per cent; $P = 0.04$). In addition, articles with involvement of a methodologist/statistician (32/86, 37 per cent) were more likely to use adjusted analyses ($P < 0.001$), while studies of Asian authorship were less likely to do so (15/175, 9 per cent). The most prevalent type of analysis performed in studies accounting for confounding factors was some type of multivariable regression ($n = 60/71$, 85 per cent; Table 1).

Of the most prevalent orthodontic-related topics, such as growth ($n = 88$), aesthetics/behavioural effects ($n = 66$), and Class II treatment ($n = 41$), a relatively small proportion (15–20 per cent) accounted for confounders in the statistical analysis (Table 2). In the univariable analysis, journal, study design, continent of authorship, and statistician involvement were identified as significant predictors for handling confounding in the statistical analysis. In the multivariable model, journal [*AJODO* versus *AO*—odds ratio (OR) = 2.36, 95 per cent confidence interval (CI): 1.16–4.83, $P = 0.02$; *EJO* versus *AO*—OR = 0.98, 95 per cent CI: 0.47–2.04, $P = 0.96$; *OCR* versus *AO*—OR = 2.10, 95 per cent CI: 0.70–6.30, $P = 0.18$], continent (Europe versus Asia—OR = 2.51, 95 per cent CI: 1.21–5.21, $P = 0.01$; America versus Asia: OR = 2.06, 95 per cent CI: 0.97–4.40, $P = 0.06$; other versus Asia—OR = 2.61, 95 per cent CI: 0.65–10.42, $P = 0.17$), and involvement of a statistician in authorship (OR = 3.91, 95 per cent CI: 2.16–7.10, $P < 0.001$) remained significant predictors (Table 3). Reporting of matching between study groups and distribution of study characteristics is shown in Table 4, irrespective of whether they had accounted for confounders. One hundred and eleven studies reported matching overall (26 per cent). Only 9 of 71 studies (13 per cent) that accounted for confounding through statistical analysis reported on matching of the groups under comparison, while 102 of 355 (29 per cent) that did not use adjusted analysis performed matching ($P = 0.01$). Therefore, the number of studies that have accounted to some degree for confounding either at the design or analysis stage was 173 (41 per cent).

Discussion

Waste within biomedical research is increasingly topical with both randomized and non-randomized studies affected both by conduct and reporting issues (8). A host of reporting and methodological shortcomings have been exposed particularly in relation to RCTs and systematic reviews within dentistry and orthodontics (8–10). The conduct and reporting of statistical analyses has been exposed as deficient in respect of over-reliance on *P* values and failure to treat clustered data appropriately (11–14). The present study, however, is the first to

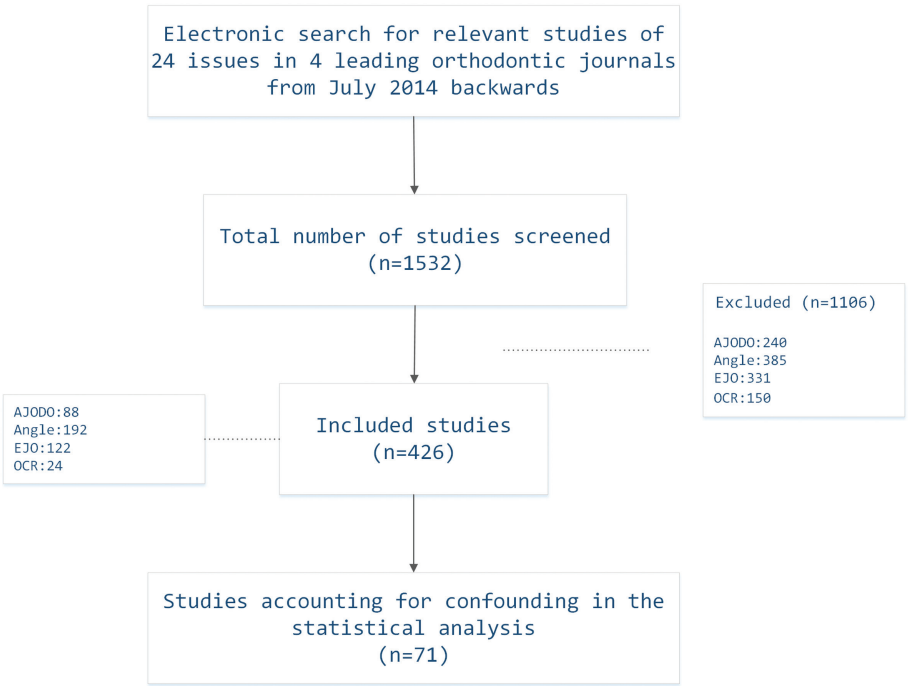

Figure 2. Flowchart of study selection.

Table 1. Distribution of included studies based on consideration of confounding in the statistical analysis or otherwise (n = 426). *AJODO, American Journal of Orthodontics and Dentofacial Orthopedics*; *AO, Angle Orthodontist*; *EJO, European Journal of Orthodontics*; *OCR, Orthodontics and Craniofacial Research*.

	Confounding considered						
	No		Yes		Total		
	n	%*	*n*	%*	*n*	%	*P* value
Journal							0.08**
AJODO	68	77	20	23	88	100	
AO	166	86	26	14	192	100	
EJO	104	85	18	15	122	100	
OCR	17	71	7	29	24	100	
Study design							0.04***
Prospective clinical trial	94	90	11	10	105	100	
Retrospective cohort	81	86	13	14	94	100	
Cross-sectional	175	80	44	20	219	100	
Retrospective case–control	5	63	3	38	8	100	
Continent							0.001***
America	90	80	23	20	113	100	
Europe	95	77	29	23	124	100	
Asia	160	91	15	9	175	100	
Other	10	71	4	29	14	100	
No. authors							0.28**
1–3	111	86	18	14	129	100	
4	84	79	23	21	107	100	
≥5	160	84	30	16	190	100	
No. centres							0.05**
Single centre	127	88	17	12	144	100	
Multi-centre	228	81	54	19	282	100	
Statistician involvement							<0.001**
No	301	89	39	11	340	100	
Yes	54	63	32	37	86	100	
Significance							0.21**
No	41	77	12	23	53	100	
Yes	314	84	59	16	373	100	
Type of analysis****							<0.001***
t-Test	287	96	11	4	298	100	
Chi-square test	33	100	0	0	33	100	
Regression	35	37	60	63	95	100	
Total	355	83	71	17	426	100	

*Row percentage.

**Pearson's chi-square test.

***Fisher's exact test.

****t-Test category includes t-test, k-way analysis of variance (ANOVA), multivariate analysis of variance (MANOVA), and non-parametric equivalents; *regression* category includes mixed models and Friedman/repeated measures ANOVA, analysis of covariance (ANCOVA), multivariate analysis of covariance (MANCOVA), linear regression, logistic regression, Cox regression, and survival analysis.

consider the specifics of handling of confounding within non-randomized studies in orthodontics, exposing rather disappointing results.

The results indicated that confounding is considered at the analysis stage less frequently than should be the case, being undertaken in just 17 per cent of this cross-section with a high number of studies resorting in 'naive' statistical comparisons, while handling of confounding at the design stage was undertaken in 26 per cent of the studies. This is an important finding as the influence of confounders can be substantial in the resultant estimates. Confounding can be negative or positive, which implies that the observed associations between exposures and outcomes can be either overestimated or underestimated or can result in non-existent associations. This blurring of true associations may influence the interpretation of findings and can be misleading or resulting in inappropriate recommendations. We could not identify a directly comparable study within the dental literature previously; however, in an analysis of a random subset of dental articles published within 10 leading dental journals from 1995 and 2009, at least one methodological error was identified in 51.5 per cent of studies (15). These figures were also consistent with error rates within even high-impact medical journals, where discrepancy rates of up to 60 per cent were exposed (16). It is, therefore, unsurprising that disappointing results were found in the present analysis. Moreover, each of the identified journals has similar peer review processes with dedicated statistical review being the exception rather than the rule. Authors are, however, encouraged to adhere to the more established reporting guidelines in submissions to each of the four orthodontic journals considered. Kim and Hong (15) attributed inappropriate use of statistics to lack of expertise both among dental researchers and reviewers. In the present analysis, the involvement of a statistician or methodologist was associated with more frequent adjustment for confounding suggesting that more detailed analyses are more likely with increasing knowledge and expertise.

Table 2. Distribution of articles using adjusted analyses to account for confounding or otherwise per orthodontic subject (n = 426).

	Confounding considered					
	No		Yes		Total	
Subject	*n*	%*	*n*	%*	*n*	%
Growth	73	83	15	17	88	100
Aesthetics/behavioural	53	80	13	20	66	100
Class II treatment	35	85	6	15	41	100
Class III treatment	34	89	4	11	38	100
Expansion/crossbite	29	85	5	15	34	100
Cleft	20	83	4	17	24	100
Extractions	19	86	3	14	22	100
Bonding/materials	18	86	3	14	21	100
Impacted canines	12	71	5	29	17	100
Oral hygiene	12	71	5	29	17	100
Implants	11	85	2	15	13	100
Retention	8	89	1	11	9	100
Open bite	7	100	0	0	7	100
Agenesis/autotransplantation	5	83	1	17	6	100
Alignment	6	100	0	0	6	100
Tooth movement/resorption	4	67	2	33	6	100
Other (temporomandibular joint/apnoea)	4	67	2	33	6	100
Anchorage	5	100	0	0	5	100
Total	355	83	71	17	426	100

*Row percentage.

Table 3. Univariable and multivariable logistic regression derived odds ratios (ORs) and 95% confidence intervals (CIs) for consideration of confounding in the statistical analysis in the identified studies (n = 426). *AJODO, American Journal of Orthodontics and Dentofacial Orthopedics*; *AO, Angle Orthodontist*; *EJO, European Journal of Orthodontics*; *OCR, Orthodontics and Craniofacial Research*.

	Univariable			Multivariable		
Category/unit	OR	95% CI	*P* value	OR	95% CI	*P* value
Journal						
AO	Reference			Reference		
AJODO	1.88	0.98–3.59	0.06	2.36	1.16–4.83	0.02
EJO	1.11	0.58–2.11	0.76	0.98	0.47–2.04	0.96
OCR	2.63	0.99–6.95	0.05	2.10	0.70–6.30	0.18
Study design						
Prospective clinical trial	Reference			Reference		
Retrospective cohort	1.37	0.58–3.23	0.47	1.41	0.56–3.60	0.47
Cross-sectional	2.14	1.06–4.36	0.03	1.93	0.89–4.19	0.10
Retrospective case–control	5.13	1.08–24.44	0.04	2.80	0.48–16.30	0.25
Continent						
Asia	Reference			Reference		
America	2.73	1.35–5.49	0.005	2.06	0.97–4.40	0.06
Europe	3.26	1.66–6.38	0.001	2.51	1.21–5.21	0.01
Other	4.27	1.19–15.26	0.03	2.61	0.65–10.42	0.17
Number of authors						
1–3	Reference			Reference		
4	1.69	0.86–3.33	0.13	1.16	0.54–2.46	0.71
≥5	1.16	0.61–2.18	0.65	0.73	0.36–1.50	0.40
Number of centres						
Single centre	Reference			Reference		
Multi-centre	1.77	0.98–3.18	0.06	1.48	0.76–2.88	0.25
Statistician involvement						
No	Reference			Reference		
Yes	4.57	2.64–7.93	<0.001	3.91	2.16–7.10	<0.001
Significance						
Yes	Reference			Reference		
No	1.56	0.77–3.14	0.22	1.60	0.71–3.58	0.26

Confounding can be handled at the design stage, at the analysis stage, or at both time points. The present findings suggest that a greater number of studies attempted to account for confounders at the design stage through matching between study groups (26 per cent), while the corresponding figure for studies handling confounding at the analysis stage was lower (17 per cent). Only a very limited

Table 4. Reporting of matching between groups across study characteristics (n = 426). *AJODO, American Journal of Orthodontics and Dentofacial Orthopedics; AO, Angle Orthodontist; EJO, European Journal of Orthodontics; OCR, Orthodontics and Craniofacial Research.*

	Matching						
	No		Yes		Total		
	n	%*	*n*	%*	*n*	%	*P* value
Journal							0.6**
AJODO	70	80	18	20	88	100	
AO	139	72	53	28	192	100	
EJO	89	73	33	27	122	100	
OCR	17	71	7	29	24	100	
Study design							0.01***
Prospective clinical trial	73	70	32	30	105	100	
Retrospective cohort	64	68	30	32	94	100	
Cross-sectional	175	80	44	20	219	100	
Retrospective case–control	3	38	5	63	8	100	
Continent							0.88***
America	81	72	32	28	113	100	
Europe	94	76	30	24	124	100	
Asia	130	74	45	26	175	100	
Other	10	71	4	29	14	100	
No. authors							0.02**
1–3	102	79	27	21	129	100	
4	68	64	39	36	107	100	
≥5	145	76	45	24	190	100	
No. centres							0.2**
Single centre	112	78	32	22	144	100	
Multi-centre	203	72	79	28	282	100	
Statistician involvement							0.35**
No	248	73	92	27	340	100	
Yes	67	78	19	22	86	100	
Significance							0.55**
No	41	77	12	23	53	100	
Yes	274	73	99	27	373	100	
Confounding considered in the analysis							0.01**
No	253	71	102	29	355	100	
Yes	62	87	9	13	71	100	
Type of analysis****							0.01**
t-Test	208	70	90	30	298	100	
Chi-square test	25	76	8	24	33	100	
Regression	82	86	13	14	95	100	
Total	315	74	111	26	426	100	

*Row percentage.

**Pearson's chi-square test.

***Fisher's exact test.

****t-Test category includes t-test, k-way analysis of variance (ANOVA), multivariate analysis of variance (MANOVA), and non-parametric equivalents; *regression* category includes mixed models and Friedman/repeated measures ANOVA, analysis of covariance (ANCOVA), multivariate analysis of covariance (MANCOVA), linear regression, logistic regression, Cox regression, and survival analysis.

number of studies, constituting less than 1 per cent, reported on handling at both stages. However, both approaches to controlling for confounding factors have utility and may be complementary. Matching at the design stage is performed when identifying potential known risk factors that are likely to have a bearing on the outcome. In addition, handling of confounding within the statistical analysis deals with residual confounding potentially arising from risk factors that may not be controlled at the design stage. For example, it may be difficult to obtain similar groups at the onset of a study, with regard to multiple risk factors, as it may be necessary to increase the required sample size markedly. Various types of adjusted analyses or more sophisticated methods, such as propensity score or inverse probability of treatment weighting, have been proposed for handling of confounding (17, 18). The vast majority of studies identified in the present research that did control for confounding used either mixed model analysis or adjusted linear/logistic regression, while none reported on more advanced analysis methods.

Limitations of this study include its restriction to four leading orthodontic journals. These journals were included as they are prominent, well regarded, and have the highest impact factors. It was therefore hoped that these might represent best practice with more significant discrepancies arising in less prominent journals. However, while journals with higher impact factor tend to be more methodologically sound with respect to RCTs and systematic reviews, it is unclear whether this applies to adjustment for confounding (19, 20). Furthermore, in view of the lack of similar research either in dentistry or medicine, it is difficult to place the present findings in context. Notwithstanding this, it is reasonable

to conclude that further knowledge and understanding of the importance of adjustment for confounding within non-randomized studies within orthodontics is important. A further shortcoming is the difficulty in assessing whether statisticians or methodologists were involved in the research based purely on author qualifications and affiliations. Direct liaison with trial authors to establish this definitively was not, however, considered worthwhile due to the likelihood of a limited response rate.

Guidelines to facilitate the reporting and indirectly conduct of observational studies and *in vitro* experiments have been developed (21, 22); however, their adoption has been less concerted than is the case with guidelines pertaining to either clinical trials or SRs. In view of the preponderance of non-randomized studies within orthodontic journals (2, 3), there is an onus on increasing awareness both of conduct and reporting of these studies, in conjunction with an appreciation of the associated guidelines among researchers, peer reviewers, and editorial staff (23). Moreover, deeper appreciation of the importance of accounting for confounding variables within these analyses would appear to be important. Dedicated statistical review of manuscripts with potential statistical issues is already ingrained within one orthodontic journal (*Journal of Orthodontics*). It may be sensible to considering extending this to other leading orthodontic journals; such initiatives have proven instructive and moderately successful within medicine (24, 25). Comparing groups without accounting for potential confounders is a common and important methodological limitation that should be addressed when planning and interpreting non-randomized studies in orthodontics. Failure to do so is likely to result in erroneous presentation and interpretation of important clinical findings.

Conclusions

Based on this cross-sectional survey of four leading orthodontic journals with high impact factor, appropriate adjustment for confounding was found to be present in less than half of the published non-randomized studies and therefore effects of exposures/interventions can potentially be over- or underestimated. Enhanced conduct and reporting of statistical tests in orthodontics is needed to ensure that inferences from research studies are appropriate and not blurred by unidentified factors.

References

1. Pandis, N., Tu, Y.K., Fleming, P.S. and Polychronopoulou, A. (2014) Randomized and nonrandomized studies: complementary or competing? *American Journal of Orthodontics and Dentofacial Orthopedics*, 146, 633–640.
2. Harrison, J.E., Ashby, D. and Lennon, M.A. (1996) An analysis of papers published in the British and European Journals of Orthodontics. *British Journal of Orthodontics*, 23, 203–209.
3. Gibson, R. and Harrison, J.E. (2011) What are we reading now? An update on the papers published in the orthodontic literature (1999–2008). *Journal of Orthodontics*, 38, 196–207.
4. Vandenbroucke, J.P. (2011) Why do the results of randomised and observational studies differ? *British Medical Journal*, 343, d7020.
5. Vandenbroucke, J.P. (2009) The HRT controversy: observational studies and RCTs fall in line. *Lancet*, 373, 1233–1235.
6. Franchi, L. and Baccetti, T. (2006) Prediction of individual mandibular changes induced by functional jaw orthopedics followed by fixed appliances in Class II patients. *Angle Orthodontist*, 76, 950–954.
7. Kirkwood, B.R. and Sterne, J.A.C. (2003) *Essential Medical Statistics. Chapter 18: Controlling for Confounding: Stratification*. Blackwell Science Ltd, Malden, MA.
8. Glasziou, P., Meats, E., Heneghan, C. and Shepperd, S. (2008) What is missing from descriptions of treatment in trials and reviews? *British Medical Journal*, 336, 1472–1474.
9. Fleming, P.S., Seehra, J., Polychronopoulou, A., Fedorowicz, Z. and Pandis N. (2013) A PRISMA assessment of the reporting quality of systematic reviews in orthodontics. *Angle Orthodontist*, 83, 158–163.
10. Fleming, P.S., Buckley, N., Seehra, J., Polychronopoulou, A. and Pandis, N. (2012) Reporting quality of abstracts of randomized controlled trials published in leading orthodontic journals from 2006 to 2011. *American Journal of Orthodontics and Dentofacial Orthopedics*, 142, 451–458.
11. Papageorgiou, S.N., Papadopoulos, M.A. and Athanasiou, A.E. (2012) Evaluation of methodology and quality characteristics of systematic reviews in orthodontics. *Orthodontics and Craniofacial Research*, 14, 116–137.
12. Fleming, P.S., Koletsi, D., Polychronopoulou, A., Eliades, T. and Pandis, N. (2013) Are clustering effects accounted for in statistical analysis in leading dental specialty journals? *Journal of Dentistry*, 41, 265–270.
13. Polychronopoulou, A., Pandis, N. and Eliades, T. (2011) Appropriateness of reporting statistical results in orthodontics: the dominance of P values over confidence intervals. *European Journal of Orthodontics*, 33, 22–25.
14. Kloukos, D., Papageorgiou, S.N., Fleming, P.S., Petridis, H. and Pandis, N. (2014) Reporting of statistical results in prosthodontic and implantology journals: p values or confidence intervals? *International Journal of Prosthodontics*, 27, 427–432.
15. Kim, D.K. and Hong, S.J. (2011) Assessment of errors and misused statistics in dental research. *International Dental Journal*, 61, 163–167.
16. Kuo, Y. (2002) Extrapolation of correlation between 2 variables in 4 general medical journals. *Journal of the American Medical Association*, 287, 2815–2817.
17. Castillo, R.C., Scharfstein, D.O. and MacKenzie, E.J. (2012) Observational studies in the era of randomized trials: finding the balance. *The Journal of Bone and Joint Surgery American*, 94, 112–117.
18. Patorno, E., Grotta, A., Bellocco, R. and Schneeweiss, S. (2013) Propensity score methodology for confounding control in health care utilization databases *Epidemiology Biostatistics and Public Health*, 10, e8940-1–e8940-15.
19. Fleming, P.S., Koletsi, D., Seehra, J. and Pandis, N. (2014) Systematic reviews published in higher impact clinical journals were of higher quality. *Journal of Clinical Epidemiology*, 67, 754–759.
20. Bala, M.M., *et al.et al.* (2013) Randomized trials published in higher vs. lower impact journals differ in design, conduct, and analysis. *Journal of Clinical Epidemiology*, 66, 286–295.
21. von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C. and Vandenbroucke, J.P. (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Medicine*, 4, e296.
22. http://www.equator-network.org/ (11 December 2014, date last accessed).
23. Hannigan, A. and Lynch, C.D. (2013) Statistical methodology in oral and dental research: pitfalls and recommendations. *Journal of Dentistry*, 41, 385–392.
24. Gore, S.M., Jones, G. and Thompson, M.A. (1992) The Lancet's statistical review process; areas for improvement by authors. *The Lancet*, 340, 100–102.
25. Gardner, M.J., Altman, D.G., Jones, D.R. and Machin, D. (1983) Is the statistical assessment of papers submitted to the "British Medical Journal" effective? *British Medical Journal*, 286, 1485–1488.