TURKISH JOURNAL of ORTHODONTICS

DOI: 10.5152/TurkJOrthod.2021.21030

Original Article

Reporting Sample Size Calculation in Randomized Clinical Trials Published in 4 Orthodontic Journals

Marialicia Calderon-Augusto[1], Carlos Flores-Mir[2], Luis Ernesto Arriola-Guillén[3]

[1]Division of Orthodontics, School of Dentistry, Universidad Científica del Sur, Lima, Perú
[2]Department of Orthodontics, University of Alberta, Edmonton, Alberta, Canada
[3]Division of Orthodontics, School of Dentistry, Universidad Científica del Sur, Lima, Perú

Cite this article as: Calderon-Augusto M, Flores-Mir C, Arriola-Guillén LE. Reporting sample size calculation in randomized clinical trials published in 4 orthodontic journals. *Turk J Orthod.* 2021; 34(4): 214-219.

Main Points

- Inadequate sample size calculation includes the failure to report the confidence level, test power, effect size, and expected variability.
- Some parameters of sample calculations were more often reported in specific journals.
- RCTs published in orthodontic journals frequently do not adequately report the parameters used for sample calculations.

ABSTRACT

Objective: The purpose of this study was to describe sample size calculations in randomized clinical trials (RCTs) published in 4 orthodontic journals.

Methods: This cross-sectional study evaluated 142 RCTs published between 2015 and 2019 in the 4 journals with the highest impact factor in orthodontics according to the SCIMAGO 2018 ranking. In the study, 2 trained and experienced orthodontists assessed whether the RCTs evaluated reported their sample size calculations, and whether they adequately described the criteria for the calculations, including the level of significance, test power, precision or effect size (clinically relevant difference), and expected variability. The sample size calculation was considered adequately reported when the above 4 criteria were described.

Results: We identified 120 publications (84.5%) reporting the sample size calculation, but only 70 (58.3%) fully described the above parameters. Inadequate calculation included failure to report the confidence level (ranging from 0% to 12.9%), test power (ranging from 0% to 20%), effect size (ranging from 0% to 22.5%), and expected variability (ranging from 22.6% to 80%). According to the journal, some parameters of sample size calculation were more frequently reported.

Conclusions: RCTs published in the 4 leading orthodontic journals frequently do not report the parameters used for sample size calculations.

Keywords: Sample size calculations, randomized clinical trials, orthodontic journals

INTRODUCTION

Randomized clinical trials (RCTs) in orthodontic scientific literature are useful for answering clinical research questions through quasi-scientific experimentation and facilitating therapeutic decision making.[1-5] Correctly executed RCTs likely provide the best evidence on the results of health interventions.[1,6] Likewise, the ability to extrapolate the results of an RCT into different populations depends on the control of biases that may be present during trial.[7,8] Furthermore, every orthodontic researcher who designs an RCT expects to be able to extrapolate their results to clinical practice.[3,4] Therefore, in addition to randomization, one of the conditions that a study requires to be extrapolated is the use of adequate sample size to provide an adequately powered study.[1,7,9-11]

Corresponding author: Luis Ernesto Arriola-Guillén, e-mail: luchoarriola@gmail.com

Received: July 31, 2019
Accepted: June 13, 2020

The samples of RCTs should be representative of the study populations in which a clinically relevant effect is to be tested.[6,10,12] The representativeness of a sample is achieved by an adequate sample size determination and by the type of sampling used, which ideally, should be probabilistic.[13,14] The use of an appropriate sample size calculation is only the starting point for controlling the external validity of a study,[15-18] since while the use of a specific sample might be reported, the wrong assumptions regarding the parameters used during the sample size calculation may have been made.[12] Hence, it is important for the criteria used for the sample calculation to be clearly described in the materials and methods section of the articles to assess whether the calculations were properly estimated.[9,13,15,16] Moreover, sample size calculation in RCTs should have enough power to detect a clinically important difference, if present, or to confirm the lack of a difference between treatment groups. The investigators should conduct appropriate sample size calculations based on clinical importance and reasonable assumptions.[19,20]

When wrong parameters are introduced in the sample size calculation, they are frequently not related to the confidence level or test power considerations.[21,22] The problems in the case of a quantitative outcome variable begin with the decision of the expected effect size (precision), which is the minimum difference that is desired to be detected between the groups to be compared, or from a clinical point of view, the minimally significant difference to decide that one treatment is better than another. The second source of problems could be the estimated variability (standard deviation or variance).[11] Concerning a qualitative outcome variable, possible errors may be more related to poor decisions of the estimated differences in the proportions of the groups compared.[8,23]

Evidence-based dentistry describes RCTs as study designs that are near the top of the evidence pyramid, and these trials are usually the primary source for supporting the conclusions of systematic reviews answering therapeutic questions. Therefore, a good description of their sample calculations should be made to allow for adequate external validity.[5,6,15,16,18,24] Several studies have already evaluated sample size reporting in orthodontics and have highlighted that despite improvements, the quality of reporting sample size parameters remains suboptimal, and further studies are needed, especially in relation to RCTs.[25-28] Therefore, the purpose of this descriptive study was to determine the frequency of application and pertinence of sample size calculations of RCTs published from 2015 to 2019 in 4 leading orthodontic journals according to the 2018 SCIMAGO ranking.

METHODS

This cross-sectional study evaluated 142 RCTs published in 4 orthodontic specialty journals with a high-impact factor in orthodontics. The study was approved by the Ethics Committee of the Cientifica del Sur University (Lima-Perú) with protocol number 669-2019-POS8.

All the articles included were RCTs published from January 2015 to December 2019, in the 4 journals with the highest impact factor in the field of orthodontics according to the SCIMAGO 2018 journal ranking, https://www.scimagojr.com/journalrank.php?area=3500&category=3505 (accessed on April 25th, 2020). "Angle Orthodontist" (AO), "American Journal of Orthodontics and Dentofacial Orthopedics" (AJODO), "European Journal of Orthodontics" (EJO), and "Korean Journal of Orthodontics" (KJO).

The inclusion criteria were RCTs as defined by their authors in the title of their publications or in the methods section, and the inclusion of human samples. Animal trials, quasi-experiments, and observational and pilot studies (defined by the authors in the title) were excluded.

For the study, 2 orthodontists (MCA and LEAG) were trained in sample size calculation parameters and were calibrated during a pilot test including 30 RCT evaluations, obtaining Kappa intraobserver and interobserver agreements of 0.93 to 1 in both measurements for all the considered variables.

The evaluators then searched for articles that met the inclusion criteria in the selected journals. Once identified, we evaluated whether the authors performed a sample size calculation (to compare 2 means or to compare 2 proportions). This allowed verification of the articles that described sample calculations, and whether these defined the respective parameters including the confidence level (the measure of certainty regarding how exactly a sample reflects the population studied within a chosen confidence interval), test power (1-ß (type II errors)), effect size or precision (the difference desired to detect, referencing where these data were obtained) and the standard deviations or variances of the control groups (variability). Only the information reported in the publication was used.

A study was considered adequate when all of the 4 above parameters were reported; otherwise, the description was deemed inadequate. All the articles were evaluated twice by the 2 evaluators with an interval of 1 month between evaluations. In cases of discrepancies in the definition of any criterion, the consensus of the 2 evaluators defined whether or not the criterion was met.

Statistical Analyses

All the analyses were performed with the Statistical Package for Social Sciences (SPSS) Windows software, Version 24.0 (IBM SPSS Corp., Armonk, NY, USA). We described the reporting of sample size in all the journals, and then in each of the 4 journals evaluated. Mainly we evaluated whether the description of the sample calculations met the parameters required for sample size calculations. We assessed the criteria fulfilled, the number of criteria fulfilled, the orthodontic journal, and the year of publication.

RESULTS

One hundred forty-two RCTs were identified in the 4 journals evaluated, 120 of which (84.5%) reported the sample calculation. Sample calculation was described in > 80% of the articles in

Table 1. Evaluation of the parameters reported for sample size calculation

	AO = 49		AJODO = 35		EJO = 52		KJO = 6		Total = 142	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Parameter	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)
Description of a sample calculation	40 (81.6)	9 (18.4)	31 (88.6)	4 (11.4)	44 (84.6)	8 (15.4)	5 (83.3)	1 (16.7)	120 (84.5)	22 (15.5)
Description of specific reports	AO = 40		AJODO = 31		EJO = 44		KJO = 5		Total = 120	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
	n (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)
Confidence level (%)	37 (92.5)	3 (7.5)	27(87.1)	4 (12.9)	41 (93.2)	3 (6.8)	5 (100.0)	0 (0.0)	110 (91.7)	10 (8.3)
Test Power	38 (95.0)	2 (5.0)	31 (100.0)	0 (0.0)	44 (100.0)	0 (0.0)	4 (80.0)	1 (20.0)	117 (97.5)	3 (2.5)
Effect size	31 (77.5)	9 (22.5)	31 (100.0)	0 (0.0)	35 (79.5)	9 (20.5)	4 (80.0)	1 (20.0)	101 (84.2)	19 (15.8)
Variability	18 (45.0)	22 (55.0)	24 (77.4)	7 (22.6)	28 (63.6)	16 (36.4)	1 (20.0)	4 (80.0)	71 (59.2)	49 (40.8)

AO, Angle Orthodontist; AJODO, American Journal of Orthodontics and Dentofacial Orthopedics; EJO, European Journal of Orthodontics; KJO, Korean Journal of Orthodontics.

the 4 journals. The criteria for sample calculation most frequently described in the 4 journals (> 90%) were the power of the test (97.5%) and the confidence level (91.7%) (Table 1).

When we evaluated the use of the 4 parameters for describing sample calculations in the 4 journals, only 70 (58.3%) studies described all the parameters, and 30 (25%) reported using at least 3 parameters. The 4 sample calculation parameters were most frequently reported in the journals AJODO and EJO (Table 2).

When evaluating adequate description of the sample calculation, it was observed that 10 (8.3%) studies did not describe the confidence level, the test power was not specified in 3 (2.5%), the effect size was not specified in 19 (15.8%), and the expected variability was not described in 49 (40.8%) studies.

The test power was described in 100% of the articles in AJODO and EJO, and in 95% of the articles in AO. Furthermore, a description of effect size was more frequent in AJODO (100%), EJO (77.5%), and AO (79.5%).

Table 2. Number of parameters reported during sample size calculation in the journals evaluated

Number of Parameters Reported	AO	AJODO	EJO	KJO	Total = 122 (100%)
	n (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)
4	17 (42.5)	23 (74.2)	29 (65.9)	1 (20.0)	70 (58.3)
3	13 (32.5)	8 (25.8)	6 (13.6)	3 (60.0)	30 (25.0)
2	9 (22.5)	0 (0.0)	9 (20.5)	1 (20.0)	19 (15.8)
1	1 (2.5)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.8)

AO, Angle Orthodontist; AJODO, American Journal of Orthodontics and Dentofacial Orthopedics; EJO, European Journal of Orthodontics; KJO, Korean Journal of Orthodontics.

The expected variability was more frequently reported in the AJODO (77.4%) and in EJO articles (63.6%) (Table 1).

Full reporting of sample size calculations was observed with 2 of the 4 journals, AJODO (74.2%) and EJO (63.6%), and was similar regarding the year of publication (Table 3).

Additionally, only 9 (7.5%) studies used a previous pilot test to perform the sample calculation, while 38 (31.7%) did not describe how the precision/effect size of their calculation was determined. Finally, among the articles performing sample size calculation, 96 (80%) exceeded the minimum sample size required, mainly in 3 of the journals evaluated (AJODO, EJO, KJO) (Table 4).

Table 3. Efficiency of sample calculation reporting (reported 4 parameters) according to the type of journal and the year of publication

Journal	Efficient Reporting, *n* (%)	Without Efficient Reporting, *n* (%)	Total, *n* (%)
AO	17 (42.5)	23 (57.5)	40 (100)
AJODO	23 (74.2)	8 (25.8)	31 (100)
EJO	29 (65.9)	15 (34.1)	44 (100)
KJO	1 (20.0)	4 (80.0)	5 (100)
All Journals	70 (58.3)	50 (41.7)	120 (100)
Year of publication			
2015	14 (77.8)	4 (22.2)	18 (100)
2016	14 (58.3)	10 (41.7)	24 (100)
2017	10 (62.5)	6 (37.5)	16 (100)
2018	14 (50.0)	14 (50.0)	28 (100)
2019	17 (50.0)	17 (50.0)	34 (100)

AO, Angle Orthodontist; AJODO, American Journal of Orthodontics and Dentofacial Orthopedics; EJO, European Journal of Orthodontics; KJO, Korean Journal of Orthodontics.

Table 4. Description of other specific parameters reported during sample size calculation

	AO	AJODO	EJO	KJO	Total = 122 (100%)
Other Specific Parameters Evaluated	***n*** **(%)**	***n*** **(%)**	***n*** **(%)**	***n*** **(%)**	***n*** **(%)**
Details of obtaining the effect size					
Not precise	19 (47.5)	4 (12.9)	14 (31.8)	1(20.0)	38 (31.7)
Previous pilot test	2 (5.0)	4 (12.9)	3 (6.8)	0 (0.0)	9 (7.5)
Obtained from literature	19 (47.5)	23 (74.2)	27 (61.4)	4 (80.0)	73 (60.8)
Exceeded the minimum sample size calculated					
Yes	26 (65.0)	26 (83.9)	39 (86.9)	5 (100.0)	96 (80.0)
No	14 (35.0)	5 (16.1)	5 (11.4)	0 (0.0)	24 (20.0)

AO, Angle Orthodontist; AJODO, American Journal of Orthodontics and Dentofacial Orthopedics; EJO, European Journal of Orthodontics; KJO, Korean Journal of Orthodontics.

DISCUSSION

A description of the calculation of sample size is necessary in scientific studies in order to determine that a representative number of individuals from a related potential population is included in the study.[6,10,12] Ideally, studies should evaluate the entire study population, but due to financial reasons and time limits, this rarely happens. A large number of investigations use unrepresentative samples and mistakenly seek to extrapolate their results to their study population.[25] To achieve external validity of the results, the study must have sample representativeness that is first related to the calculation of the sample size; and, second, to the type of sampling that should be a probabilistic, ensuring that all individuals in a study population have the same probability of being chosen for the study.[6,10-14] Thus, errors in the calculation of sample size may be widespread due to a lack of knowledge of the methodological importance of this calculation, which consequently affects the representativeness of the samples and thereby decreases the real value of the study.[12,17] In the studies in which a sample size calculation was provided, it was noted that the determination of effect size (the clinically relevant difference between groups) or variability (the amount of data dispersion in the groups) raised the greatest difficulties. These 2 parameters are the main criteria for calculating sample size, since the selection of inaccurate precision or variability in sample calculation will likely increase or decrease the *P* value, precluding the identification of significant differences between the groups compared or even oversizing the study groups.

Currently, RCTs are considered the best study design to address therapeutic clinical questions. It is important to note that for the present study, in addition to RCT being included in the title or the methods section of the studies evaluated, we confirmed compliance to this condition in the papers analyzed to avoid any type of bias. It is important for RCTs to comply with reporting representative samples.[29,30] We therefore determined the frequency of application and the pertinence of sample size calculation in RCTs published from 2015 to 2019 in the 4 journals with the highest orthodontic impact factor according to the SCIMAGO 2018 ranking.

Few studies in other areas of dentistry have evaluated whether sample size calculations of RCTs were adequately developed.[29-32] A review of 42 RCTs in prosthodontics published between 2008 and 2017 in the leading prosthodontic journals concluded that 50% of the publications did not adequately describe the calculation of their sample sizes.[31] Likewise, in endodontics, a review of 50 RCTs published in the 2 high-impact journals in this field from 2000 to 2001 and 2009 to 2010 concluded that 60% of the studies did not indicate how the sample sizes were determined. Although detailed reporting significantly improved between 2000 and 2010,[29] these results were still not encouraging. To the best of our knowledge, only a few related studies were conducted in orthodontics in 2011,[26] 2014,[27] and more recently in 2019,[28] but did not consider only RCTs. This first study[26] evaluated the frequency of reporting sample size calculation in studies published in journals in Brazil and the United States. The authors suggested that the researchers and the editorial committee of these journals should be more concerned about errors related to the use of an inappropriate sample size. The second study[27] concluded that although sample size calculations are often reported in trials published as RCTs in orthodontic specialty journals, reporting is usually suboptimal and in need of significant improvement. Nonetheless, this trend has yet to be addressed, with more than 40% of RCTs presenting deficiencies in the reporting of sample size calculation.

Our study evaluated 142 RCTs published from 2015 to 2019 in the 4 major orthodontic journals according to the SCIMAGO 2018 ranking, which is highly recognized in the academic field. We found that although 84.5% of these publications reported the use of sample size calculations, only 58.3% of the publications complied with the description of the 4 criteria for sample size calculation: confidence level, test power, effect size or precision, and variability of the results. It is important to note that the confidence level used in all the RCTs evaluated was 95%, with most reporting a test power of 80% or 90%. It is clear that a variation in these values may increase or decrease the sample size, and consequently, the possibility of obtaining a representative sample. Moreover, it is necessary for this information to be adequately described for good understanding of sample size calculation, analysis of external validity, and finally to guide new RCTs.

In a previous study comparing the percentages of reporting sample calculations (2005 and 2008), only 3% of the studies in Brazilian journals and 21% in American journals described this calculation.[26] According to our findings, these results have considerably increased (84.5%). This information indicates that despite the increase in the description of sample size calculations between 2008 and 2019 in the orthodontic journals evaluated, sample size calculations are still under-reported. It should also be noted that the descriptions assessed were not the same in all the journals.

As mentioned above, the actual external validity was questionable in almost all of the published RCTs. Even in the case of split-mouth designs, accurate description of the parameters used for sample size calculation should be included and described in the method section of scientific articles, but this does not always occur. Furthermore, it is important to specify that the sample calculation achieves a minimum required sample size, considering that a larger number should be included in order to obtain an adequate final power for the study due to possible sample loss to follow-up. All of these aspects should be taken into account by all orthodontic journal editors in order to ensure and increase the external validity of the RCTs published.[29,32,33,34]

The main purpose of this study was not to compare the criteria of different orthodontic journals for calculating the sample size of RCTs but rather to describe how many clinical trials meet the requirements of good sample calculation according to the most representative parameters reported in the scientific literature. This is important in order to promote the practice of better sample calculations in RCTs, regardless of the type of journal.

We found that only a little more than half (58.3%) of the studies in the 4 journals evaluated complied with the 4 parameters for sample size calculation, with 25% fulfilling 3 parameters. Full reporting of sample size calculations was performed in AJODO (74.2%) and EJO (63.6%), but it was not related to the year of publication.

It should be noted that of the 142 RCTs selected, all included a quantitative outcome variable, and therefore, only specific sample size characteristics to compare 2 means were evaluated. It was observed that a low percentage of publications did not report the confidence level, although it is the most straightforward criterion to report. On the other hand, while not reported in all the publications, the power of the sample size power was described in 100% of the articles in AJODO and EJO, and in 95% of the articles in AO. Likewise, the confidence level was the second most frequently described criterion (> 91%) in all orthodontic journals.

Moreover, the effect size (clinically relevant difference) was not reported in 15.8% of the included RCTs. It is essential to specify and quantify the effect size for reviewers and readers to know whether the reported difference is clinically relevant and how the effect size was calculated.[25] In the RCTs published in AJODO, EJO, and AO, the effect sizes were 100%, 77.5% and 79.5%, respectively. While many studies reported effect size values that were thereafter not considered when discussing their study results, these values should have been used as the clinical relevance threshold. In addition, these results are sometimes not related to an actual clinically relevant difference but are presented in the article as if they really were. Future studies should take into account the effect size in terms of small, medium, or large clinical impact. Likewise, almost half of the publications did not describe the expected variability for their sample calculations (40.8%), and therefore, the amount of data dispersion these publications expected remained unknown. In this regard, this criterion was best described in RCTs published in AJODO and EJO (77.4% and 63.6%, respectively) compared to the studies in the other 2 journals.

In the evaluation of precision in sample size calculation, 60.8% of the studies obtained this parameter from the scientific literature, while only 7.5% obtained the study precision from a previous pilot test, and 31.7% did not report sample precision, despite it being an important parameter to consider in sample calculations.

On the other hand, it is important to note that the majority of the publications evaluated (80%) exceeded the minimum sample size required by their sample calculations, mainly in 3 of the journals assessed (AJODO, EJO, KJO), being an excellent methodological finding that should be noted.

In summary, despite the large number of RCTs published in the field of orthodontics reporting the use of sample size calculation, there are still deficiencies in describing the method of sample size determination, leading to uncertainty regarding the external validity of the results. Effective, in-depth descriptions of sample size calculations should be provided in future publications to improve the extrapolation of the results reported to clinical scenarios.

One limitation of this study was that it only analyzed a specific aspect of the development of an RCT, and it is clear that the methodological quality of RCTs also involves many other aspects. We focused on sample calculation since it is associated with the representativeness of a sample and the possibility of achieving greater external validity. Additionally, other important aspects can influence the calculation of sample size. These include, among others, the necessary information to fulfill the sample size formulas, whether derived from a pilot study or not, the statistical analysis to obtain the data to be introduced in the formulas, and the number of study groups, especially when there are more than 2 groups and a specific group is required. However, we only analyzed the main factors involved in sample calculation based on the main reports of the scientific literature. Nevertheless, more studies considering these different points of analysis are needed.

CONCLUSION

Adequate description of sample size calculation in RCTs published in 4 leading orthodontic journals is largely deficient, and that should be taken into account in future publications to improve the quality of orthodontic RCTs.

Ethics Committee Approval: Ethics committee approval was received for this study from the Ethics Committee of the Cientifica del Sur University (Lima-Perú) with protocol number 669-2019-POS8.

Informed Consent: N/A.

Peer Review: Externally peer-reviewed.

Author Contributions: Concept - M.C.A., C.F.M., L.E.A.G.; Design - M.C.A., C.F.M., L.E.A.G.; Supervision - M.C.A., C.F.M., L.E.A.G.; Fundings - M.C.A., C.F.M., L.E.A.G.; Materials - M.C.A., C.F.M., L.E.A.G.; Data Collection and/or Processing - M.C.A., C.F.M., L.E.A.G.; Analysis and/or Interpretation - M.C.A., C.F.M., L.E.A.G.; Literature Review - M.C.A., C.F.M., L.E.A.G.; Writing - M.C.A., C.F.M., L.E.A.G.; Critical Review - M.C.A., C.F.M., L.E.A.G.

Conflict of Interest: The authors have no conflicts of interest to declare.

Financial Disclosure: The authors declared that this study has received no financial support.

REFERENCES

1. Berger VW, Bears JD. When can a clinical trial be called 'randomized'? *Vaccine*. 2003;21(5-6):468-472. [CrossRef]
2. Dechartres A, Charles P, Hopewell S, Ravaud P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. *J Clin Epidemiol*. 2011;64(2):136-144. [CrossRef]
3. Lempesi E, Koletsi D, Fleming PS, Pandis N. The reporting quality of randomized controlled trials in orthodontics. *J Evid Based Dent Pract*. 2014;14(2):46-52. [CrossRef]
4. Benson P. *Systematic Reviews–Professional Masochism or Important Process?* London, England: SAGE Publications; 2008.
5. Zuccati G, Clauser C, Giorgetti R. Randomized clinical trials in orthodontics: reality, dream, or nightmare? *Am J Orthod Dentofacial Orthop*. 2009;136(5):634-637. [CrossRef]
6. Hill CL, LaValley MP, Felson DT. Discrepancy between published report and actual conduct of randomized clinical trials. *J Clin Epidemiol*. 2002;55(8):783-786. [CrossRef]
7. Manríquez M J, Valdivia C G, Rada G G, Letelier S LM. Critical assessment of randomized controlled trials published in biomedical Chilean journals. *Rev Med Chil*. 2005;133(4):439-446. [CrossRef]
8. Schulz KF, Altman DG, Moher D, CONSORT Group, CONSORT. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*. 2010;152(11):726-732. [CrossRef]
9. Lazcano-Ponce E, Salazar-Martínez E, Gutiérrez-Castrellón P *et al.* Randomized clinical trials: variants, randomization methods, analysis, ethical considerations and regulation. *Salud Publica Mex*. 2004;46(6):559-584. [CrossRef]
10. Peduzzi P, Henderson W, Hartigan P, Lavori P. Analysis of randomized controlled trials. *Epidemiol Rev*. 2002;24(1):26-38. [CrossRef]
11. Shea BJ, Reeves BC, Wells G *et al.* AMSTAR 2: a critical appraisal tool for systematic reviews that include randomized or non-randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:j4008. [CrossRef]
12. Brookes ST, Whitely E, Egger M *et al.* Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol*. 2004;57(3):229-236. [CrossRef]
13. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA*. 1994;272(2):122-124.
14. Moher D, Cook DJ, Jadad AR et al. Assessing the quality of reports of randomized trials: implications for the conduct of meta-analyses. *Health Technol Assess*. 1999;3(12):i-iv, 1, 1-98. [CrossRef]
15. Higgins JP, Altman DG, Gøtzsche PC et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928. [CrossRef]
16. Higgins JP, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons; Chichester; 2011.
17. Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, Flores-Mir C. Randomized clinical trials in dentistry: risks of bias, risks of random errors, reporting quality, and methodologic quality over the years 1955-2013. *PLOS ONE*. 2017;12(12):e0190089. [CrossRef]
18. Bigby M, Williams H. Appraising systematic reviews and meta-analyses. *Arch Dermatol*. 2003;139(6):795-798. [CrossRef]
19. Pandis N, Fleming PS, Hopewell S, Altman DG. The CONSORT Statement: application within and adaptations for orthodontic trials. *Am J Orthod Dentofacial Orthop*. 2015;147(6):663-679. [CrossRef]
20. Pandis N, Polychronopoulou A, Eliades T. Sample size estimation: an overview with applications to orthodontic clinical trial designs. *Am J Orthod Dentofacial Orthop*. 2011;140(4):e141-e146. [CrossRef])
21. Deeks JJ, Dinnes J, D'Amico R et al. Evaluating non-randomised intervention studies. *Health Technol Assess*. 2003;7(27):iii-x, 1-173. [CrossRef]
22. Freiman JA, Chalmers TC, Smith H Jr., Kuebler RR. The importance of beta, the type II error, and sample size in the design and interpretation of the randomized controlled trial. Medical uses of statistics. *N Engl J Med.* 1978;299(13):690-694. [CrossRef]
23. Begg C, Cho M, Eastwood S et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996;276(8):637-639. [CrossRef]
24. Harrison JE. Clinical trials in orthodontics II: assessment of the quality of reporting of clinical trials published in three orthodontic journals between 1989 and 1998. *J Orthod*. 2003;30(4):309-15; discussion 297. [CrossRef]
25. Sandhu SS, Sandhu J, Kaur H. Reporting quality of randomized controlled trials in orthodontics--what affects it and did it improve over the last 10 years? *Eur J Orthod*. 2015;37(4):356-366. [CrossRef]
26. Normando D, Almeida MAdO, Quintão CCA. Análise do emprego do cálculo amostral e do erro do método em pesquisas científicas publicadas na literatura ortodôntica nacional e internacional. *Dent Press J Orthod*. 2011;16(6):33-35. [CrossRef]
27. Koletsi D, Pandis N, Fleming PS. Sample size in orthodontic randomized controlled trials: are numbers justified? *Eur J Orthod*. 2014;36(1):67-73. [CrossRef]
28. Gratsia S, Koletsi D, Fleming PS, Pandis N. A priori power considerations in orthodontic research: a 3-year meta-epidemiologic study. *Eur J Orthod*. 2020;42(4):454-459. [CrossRef]
29. Bondemark L, Ruf S. Randomized controlled trial: the gold standard or an unobtainable fallacy? *Eur J Orthod*. 2015;37(5):457-461. [CrossRef]
30. Shahravan A, Haghdoost AA, Rad M, Hashemipoor M, Sharifi M. Sample size calculation of clinical trials published in two leading endodontic journals. *Iran Endod J*. 2014;9(1):56-60.
31. Dumbrigue HB, Dumbrigue EC, Dumbrigue DC, Chingbingyong MI. Reporting of sample size parameters in randomized controlled trials published in prosthodontic journals. *J Prosthodont*. 2019;28(2):159-162. [CrossRef]
32. Koletsi D, Fleming PS, Seehra J, Bagos PG, Pandis N. Are sample sizes clear and justified in RCTs published in dental journals? *PLOS ONE*. 2014;9(3);9(1):e85949. [CrossRef]
33. Dechartres A, Trinquart L, Atal I et al. Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: research on research study. *BMJ*. 2017;357:j2490. [CrossRef]
34. Koletsi D, Pandis N, Polychronopoulou A, Eliades T. What's in a title? An assessment of whether randomized controlled trial in a title means that it is one. *Am J Orthod Dentofacial Orthop*. 2012;141(6):679-685. [CrossRef]