Reporting of Statistical Results in Prosthodontic and Implantology Journals: *P* Values or Confidence Intervals?

Dimitrios Kloukos, DDS, Dr Med Dent[a]/Spyridon N. Papageorgiou, DDS[b]/
Padhraig S. Fleming, BDent Sc, MSc, PhD, FDS (Orth)[c]/Haralampos Petridis, DDS, MS, Cert Prosthod, PhD[d]/
Nikolaos Pandis, DDS, Dr Med Dent, MSc, DLSHTM[e]

Purpose: Confidence intervals (CIs) are integral to the interpretation of the precision and clinical relevance of research findings. The aim of this study was to ascertain the frequency of reporting of CIs in leading prosthodontic and dental implantology journals and to explore possible factors associated with improved reporting. ***Materials and Methods:*** Thirty issues of nine journals in prosthodontics and implant dentistry were accessed, covering the years 2005 to 2012: *The Journal of Prosthetic Dentistry*, *Journal of Oral Rehabilitation*, *The International Journal of Prosthodontics*, *The International Journal of Periodontics & Restorative Dentistry*, *Clinical Oral Implants Research*, *Clinical Implant Dentistry and Related Research*, *The International Journal of Oral & Maxillofacial Implants*, *Implant Dentistry*, and *Journal of Dentistry*. Articles were screened and the reporting of CIs and *P* values recorded. Other information including study design, region of authorship, involvement of methodologists, and ethical approval was also obtained. Univariable and multivariable logistic regression was used to identify characteristics associated with reporting of CIs. ***Results:*** Interrater agreement for the data extraction performed was excellent (kappa = 0.88; 95% CI: 0.87 to 0.89). CI reporting was limited, with mean reporting across journals of 14%. CI reporting was associated with journal type, study design, and involvement of a methodologist or statistician. ***Conclusions:*** Reporting of CI in implant dentistry and prosthodontic journals requires improvement. Improved reporting will aid appraisal of the clinical relevance of research findings by providing a range of values within which the effect size lies, thus giving the end user the opportunity to interpret the results in relation to clinical practice. *Int J Prosthodont 2014;27:427–432. doi: 10.11607/ijp.4011*

Accurate and transparent research conduct and reporting are a foundation of health care decision making and delivery. Numerous guidelines have been developed in an effort to improve reporting, with a plethora of guidelines specific to individual study designs. For example, the Consolidated Standards of Reporting Trials (CONSORT) statement has been produced in an attempt to standardize reporting of randomized controlled trials (RCTs).[1] Among the 25 items in the checklist, it is suggested that for each primary and secondary outcome, results are reported as a summary of the outcome in each group (eg, mean or proportion) together with the effect size (risk ratio, relative risk, odds ratio, risk difference, hazard ratio, difference in median survival time, or difference in means) and its precision, such as a 95% confidence interval (CI). Unfortunately, however, compliance with reporting guidelines has been exposed as suboptimal both in medical and dental research.[2–5]

Inferential statistical tests comparing outcomes among two or more groups are an integral part of most research studies, with results typically reported with corresponding two-sided *P* values. The *P* value represents the probability of finding the observed treatment effect/difference or a more extreme one among the

[a]*Research Associate, Department of Orthodontics and Dentofacial Orthopedics, Faculty of Medicine, University of Bern, Bern, Switzerland.*
[b]*Postgraduate Student, Department of Orthodontics, School of Dentistry, University of Bonn, Bonn, Germany; Doctoral Fellow, Department of Oral Technology, School of Dentistry, University of Bonn, Bonn, Germany; Research Assistant, Clinical Research Unit 208, University of Bonn, Bonn, Germany.*
[c]*Senior Clinical Lecturer/Consultant, Barts and the London School of Medicine and Dentistry, Institute of Dentistry, Queen Mary University of London, United Kingdom.*
[d]*Senior Clinical Lecturer, Prosthodontic Unit, Department of Restorative Dentistry, University College London Eastman Dental Institute, London, United Kingdom.*
[e]*Private Practice, Corfu, Greece; Assistant Professor, Department of Orthodontics and Dentofacial Orthopedics, Faculty of Medicine, University of Bern, Bern, Switzerland.*

Correspondence to: *Dr Haralampos Petridis, Prosthodontic Unit, Department of Restorative Dentistry, UCL Eastman Dental Institute, 256 Gray's Inn Road, London WC1X 8LD, UK. Email: c.petridis@ucl.ac.uk*

study groups, when no difference exists among these groups, ie, when the null hypothesis is true. If the observed *P* value is small, typically below .05, the null hypothesis is rejected, and a conclusion is drawn that there is evidence of a true difference beyond chance, ie, a statistically significant effect. To quantify definitively the difference among groups or the size of effect, the entire target population must be considered. However, as this is unfeasible, a representative sample is used to make inferences about the entire population. While *P* values are indicative of a statistically significant difference or outcome, their value does not extend beyond the dichotomy of the presence or absence of significance. Of greater clinical relevance is the between-groups difference or effect size and its range[6–8]; CIs are required to provide this added information.

The size of the observed *P* value is influenced by sample size and variance of the sample, while *P* values, an arbitrary cut-off, may have limited relationship with the observed effect size and clinical importance. For example, a parallel-group clinical trial involving 1,000 patients in each arm is conducted to evaluate the time to failure of restorative materials A and B. The mean survival period is 60 and 61 months, respectively, with an SD of 5 months for both groups. A *t* test between the two groups would produce a highly significant result ($P = .001$). However, by analyzing the difference in survival between the two groups in conjunction with the CI (mean difference: 1 month; 95% CI: 0.32 to 1.62 months), the interpretation of the outcome is altered, as the observed difference lacks clinical significance. CIs display the range of the plausible difference of effect or association between or among groups, providing insight into whether the observed differences are important and likely to reflect the population in general.

CIs always contain the effect estimate and allow assurance, at a predefined level (usually 95% or 99%), that the intervals contain the true population value. The 95% CIs imply that if 100 samples were drawn from the target population, 95% of them would contain the true population value. Increasing the sample size leads to narrowing of the width of the CIs around the same size of effect, thus increasing precision; this is not the case with *P* values, where increasing the sample size simply lowers the *P* value.

It has previously been shown both in dental and medical research that *P* values are commonly reported in isolation, leading to conclusions concerning treatment effectiveness while disregarding the direction of the effect, its size, its range, and the clinical importance of the observed results.[9] There is, therefore, a tendency for the result to be interpreted in terms of significance or nonsignificance based solely on *P* values. Consequently, any significant result may be regarded as important, irrespective of its clinical importance or plausibility, while nonsignificant results, regardless of clinical importance, are interpreted as indicating no difference of effect.[10–12] There is a lack of studies evaluating the quality of statistical reporting in the prosthodontic and implant dentistry literature, with no previous assessment of whether only *P* values and/or CIs are included in the results. The objective of this study was, therefore, to assess recent prosthodontic and implant dentistry research evaluating the frequency of CI reporting and exploring possible associations between CI reporting and publication characteristics, such as journal of publication, study type, region of publication, ethical approval, and involvement of a statistician or methodologist in the research.

Materials and Methods

Four dental journals with an emphasis on prosthodontics (*The International Journal of Prosthodontics, Journal of Oral Rehabilitation*, *The Journal of Prosthetic Dentistry*, *The International Journal of Periodontics & Restorative Dentistry*), four dental implantology journals (*Clinical Implant Dentistry and Related Research*, *Clinical Oral Implants Research*, *Implant Dentistry*, *The International Journal of Oral & Maxillofacial Implants*) and one general dental journal with a predilection for prosthodontics (*Journal of Dentistry*) were included in the study. The selected journals had the highest impact factors of prosthodontic and implant dentistry journals based on 2009 data.

The contents of 30 issues of each journal from March 2012 backwards were searched. Supplemental issues included in this time frame were included in the analysis but were not counted as an issue. Online-only articles, editorials and letters, case reports, reviews, and descriptive articles with no statistical comparisons for the main research question, were excluded from the analysis. Both clinical (animal or human) and laboratory studies were included. Screening and selection of studies were conducted independently by two authors (DK, SNP). After the studies were selected, a calibration exercise with 80 randomly selected studies between the authors conducting the data extraction (DK, SNP) was performed. Interrater agreement was evaluated for all extracted data, and any disagreements were resolved with discussion.

Data extracted included reporting of *P* values and/or notation of significance or nonsignificance or reporting of CIs for between-group differences. Other data obtained included journal and year of publication; region of publication (Europe, Americas, or Other region, based on the first author); ethical approval; involvement of a statistician or methodologist; single or multicenter study; research design (in vitro, interventional, or observational design); and whether,

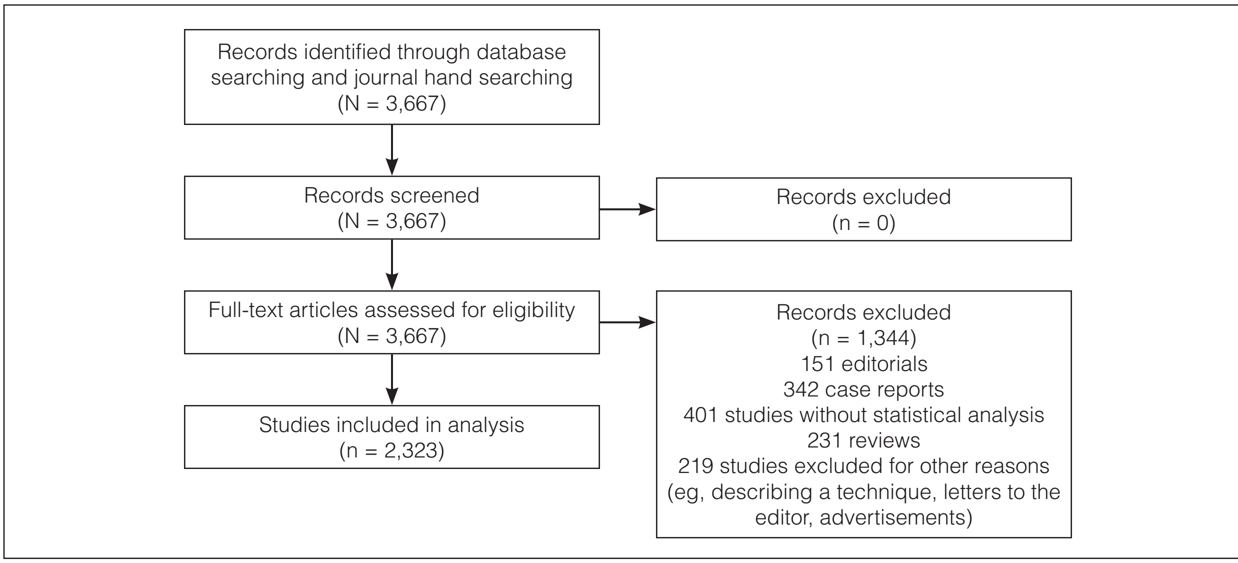

Fig 1 Study selection flow diagram.

if interventional, the study was an RCT. Involvement of a statistician or methodologist was ascertained by checking author affiliations (public health or epidemiology departments were considered as providing statistical assistance), author degrees (where provided), and information in the methods or acknowledgment section of each paper.

The frequency of CI reporting was the dependent variable in the analyses. Univariable logistic regression was used to identify characteristics associated with CI reporting, followed by multivariable logistic regression for simultaneous investigation of a number of predictors. Model fit was assessed using the Hosmer-Lemeshow test. All analyses were performed with Stata software (version 13.1, StataCorp). A two-sided *P* value of .05 was considered statistically significant with a 95% CI.

Results

Overall, 3,667 articles were screened, and 2,323 were included in this study (Fig 1). Only 314 (14%) of them presented CIs. The interrater agreement was excellent (Cohen's kappa: 0.88; 95% CI: 0.87 to 0.89). Included papers were published between 2005 and 2012. The journals contributing the highest percentage of papers toward the final included studies were *The International Journal of Oral & Maxillofacial Implants* (19%), *Clinical Oral Implants Research* (16%), and *The International Journal of Prosthodontics* (15%). Among these journals, CI reporting was most prevalent in *The International Journal of Prosthodontics* (22%), *Clinical Oral Implants Research* (20%), *Journal of Dentistry*, and *Journal of Oral Rehabilitation* (13% each). Reporting of CIs was uniformly relatively poor, as shown in Table 1.

The univariable analysis indicated that there was evidence of association between CI reporting and journal type, region of publication, involvement of a statistician/methodologist, number of research centers, and study design (Tables 1 and 2). Similar findings were identified in the multivariable analysis, although number of research centers was no longer statistically significant (Table 2). In the multivariable analysis, *The International Journal of Periodontics & Restorative Dentistry* was the least likely journal to report CIs. The highest odds for CI reporting were identified for the *Journal of Dentistry* (odds ratio [OR]: 3.33; 95% CI: 1.60 to 6.93) followed by the *The International Journal of Prosthodontics* (OR: 3.31; 95% CI: 1.66 to 6.60) compared to the baseline group. No difference in the odds of CI reporting was observed with respect to ethical approval (OR: 1.06; 95% CI: 0.86 to 1.39) or publication year (OR: 1.02; 95% CI: 0.93 to 1.13).

Discussion

The interpretation of *P* values derived from statistical testing is typically qualitative, with study results presented as either significant or not significant, whereas CIs provide a range of values, within which the true difference among the study groups is believed to exist, thus giving the reader the opportunity to interpret the clinical relevance of the results. Furthermore, *P* values have no units, whereas CIs assume the units of the outcome variable, making interpretation of the results easier. Presentation of CIs is particularly informative when nonsignificant results are obtained, permitting judgment of the importance of nonsignificant differences based on clinical relevance. The importance of the latter has been confirmed in research by Freiman

Table 1 Reporting of CIs for the Assessed Studies (n = 2,323) per Journal of Publication, Year of Publication, Authorship Region, Ethical Approval, Statistician Involvement, Number of Research Centers, Study Design, and Overall Number of RCTs

	Confidence interval reporting			
	No (%)	Yes (%)	Overall (%)	*P* value
Journal				
Clin Implant Dent Relat Res	167 (85)	29 (15)	196 (100)	< .001
Clin Oral Implants Res	303 (83)	63 (17)	366 (100)	
Implant Dent	125 (93)	9 (7)	134 (100)	
Int J Oral Maxillofac Implants	398 (92)	33 (8)	431 (100)	
Int J Periodontics Restorative Dent	174 (94)	11 (6)	185 (100)	
Int J Prosthodont	280 (80)	68 (20)	348 (100)	
J Dent	246 (85)	42 (15)	288 (100)	
J Oral Rehabil	177 (81)	42 (19)	219 (100)	
J Prosthet Dent	139 (89)	17 (11)	156 (100)	
Year of publication				
2005	10 (100)	0 (0)	10 (100)	.56
2006	12 (100)	0 (0)	12 (100)	
2007	92 (90)	10 (10)	102 (100)	
2008	182 (87)	27 (13)	209 (100)	
2009	247 (88)	35 (12)	282 (100)	
2010	524 (86)	88 (14)	612 (100)	
2011	590 (86)	99 (14)	689 (100)	
2012	352 (86)	55 (14)	407 (100)	
Region of authorship				
Europe	910 (85)	160 (15)	1,070 (100)	.05
Americas	481 (86)	78 (14)	559 (100)	
Other	618 (89)	76 (11)	694 (100)	
Ethical approval				
No	1,030 (51)	154 (49)	1,184 (51)	.46
Yes	979 (49)	160 (51)	1,139 (49)	
Statistician/methodologist involvement				
No	1,692 (89)	218 (11)	1,910 (100)	< .001
Yes	317 (77)	96 (23)	413 (100)	
Single- or multicenter				
Single-center	613 (89)	79 (11)	692 (100)	.05
Multicenter	1,396 (86)	235 (14)	1,631 (100)	
Design				
In vitro	762 (92)	70 (8)	832 (100)	< .001
Interventional	648 (89)	80 (11)	728 (100)	
Observational	599 (79)	164 (21)	763 (100)	
RCT				
No	1,896 (87)	279 (13)	2,175 (100)	< .001
Yes	113 (76)	35 (24)	148 (100)	
Total	2,009 (86)	314 (14)	2,323 (100)	

RCT = randomized controlled trial.

et al,[13] who reanalyzed the results of 71 negative studies based on significance testing, using CIs to reinterpret the study results. The reanalysis indicated that many of the treatments might have erroneously been labeled as ineffective initially due to the undue isolated emphasis on *P* values.

The mean overall reporting frequency of CIs (14%) is in keeping with findings both from research within medical and dental journals. Vavken et al,[14] in a recent publication, found the CI reporting in orthopedic research to be around 20%. Moreover, the authors identified that the probability of statistically significant results representing a between-group difference of more than 10% was only 69%; a high proportion of statistically significant results may therefore not reflect large treatment effects. Concurrent reporting of CIs

Table 2 Results of Univariable and Multivariable Logistic Regression of CI Reporting Odds Ratios Relating to CI Reporting for Journal, Year of Publication, Authorship Region, Ethical Approval, Statistician Involvement, Number of Research Centers, and Study Design

	Univariable			Multivariable		
Predictor	OR	95% CI	*P* value	OR	95% CI	*P* value
Journal						
Clin Implant Dent Relat Res	2.75	1.33–5.68	.01	2.48	1.17–5.24	.02
Clin Oral Implants Res	3.29	1.69–6.41	< .001	2.75	1.38–5.49	.001
Implant Dent	1.14	0.46–2.83	.78	1.19	0.47–3.01	.72
Int J Oral Maxillofac Implants	1.31	0.65–2.66	.45	1.26	0.61–2.60	.54
Int J Periodontics Restorative Dent	Reference					
Int J Prosthodont	3.84	1.98–7.47	< .001	3.31	1.66–6.60	.001
J Dent	2.70	1.35–5.39	.01	3.33	1.60–6.93	.001
J Oral Rehabil	3.75	1.87–7.53	< .001	2.56	1.23–5.31	.01
J Prosthet Dent	1.93	0.88–4.26	.10	2.79	1.20–6.48	.02
Year						
(per unit)	1.07	0.98–1.17	.13	1.02	0.93–1.13	.62
Country						
Europe	1.43	1.07–1.91	.02	1.23	0.90–1.67	.20
Americas	1.32	0.94–1.85	.11	1.48	1.04–2.12	.03
Asia/other	Reference					
Ethics						
No	Reference					
Yes	1.06	0.86–1.39	.46			
Statistician						
No	Reference					
Yes	2.35	1.80–3.07	< .001	1.99	1.50–2.65	< .001
Single- or multicenter						
Single-center	Reference					
Multicenter	1.31	1.00–1.71	.05	1.15	0.86–1.53	.36
Design						
In vitro	Reference					
Interventional	1.34	0.96–1.88	.09	1.13	0.74–1.75	.57
Observational	2.98	2.21–4.02	< .001	3.14	2.23–4.41	< .001
RCT						
No	Reference					
Yes	2.12	1.42–3.17	< .001	2.68	1.62–4.42	< .001

OR = odds ratio; RCT = randomized controlled trial.

would help to qualify these results accordingly. An assessment of statistical methodology in dental research indicated that only 20 out of 307 (6.5%) identified papers reported on CIs.[15] Reporting of CIs in endodontic leakage studies was found to be extremely limited, with 1.1% of the papers reporting them.[16] Results from a previous study relating to orthodontic journals highlighted less satisfactory reporting, with CIs displayed in just 6% of assessed studies. In that study, papers involving multivariable statistical analysis were associated with significantly greater likelihood of CI reporting.[9]

In the multivariable analysis, an array of factors were found to be of significance in relation to the likelihood of CI reporting. In particular, CI reporting was found to differ significantly among the assessed journals; this may reflect editorial policy and the degree of active editorial and reviewer intervention. Similarly, where statisticians or methodologists were involved, CI reporting was more prevalent; this improvement may reflect greater awareness of the value of CIs among this group. Although the input of statistical expertise goes often unrecognized by either authorship or acknowledgment, it has been noted that research without methodological assistance is more likely to be rejected without review and less likely to be accepted for publication.[17] CI reporting was also found to be more prevalent in RCTs than other designs. The latter may reflect the influence of the espousal and implementation of CONSORT guidelines in the constituent journals, with all of the assessed journals being CONSORT endorsers. Nevertheless, the prevalence of CI reporting even among RCTs was low, reflecting limited influence of guideline endorsement. This is a feature common to many reporting guidelines within branches of both medicine and dentistry.[18] While previous research has linked improved reporting of CIs to use of advanced statistical analyses including multivariable analyses,[9] this association was not assessed in the present study.

Among the different study designs, observational was the most likely to report CIs and the same was evident in RCTs vs nonRCTs. The latter was in agreement with previous studies.[19]

The limitations of the present study may be related to possible misclassification of recorded variables, given that double data extraction was not implemented. However, given the large differences in reporting vs nonreporting of CIs, possible errors are unlikely to alter the direction of the effects and conclusions of this study.

Limited reporting of CIs in prosthodontic and implant dentistry journals may reflect a lack of awareness or understanding of their value in statistical testing. Similar deficiencies in dental journals have been reported in respect to other statistical issues[4,5,20] and in reporting both of clinical trials and systematic reviews.[2,3] The implication of failure to report CIs may be misinterpretation of dental research findings and subsequent application of the results in clinical practice. It is, therefore, important that greater editorial emphasis is placed on the reporting of estimation and CIs in prosthodontic and implant research, rather than relying on *P* values in isolation to make inferences from study results. However, while editorial intervention has proven successful in improving reporting of CIs in medicine,[21] intervention has been less successful in inducing associated interpretation and discussion of results, including precision and size of effect based on CIs.[22] It is, therefore, important that researchers are adequately versed both in the need to present CIs and in the appreciation of the rationale for so doing; this knowledge and understanding will permit a necessary shift in emphasis from hypothesis testing to estimation of effect size and direction.

Conclusions

CI reporting in prosthodontic and implant dentistry journals is low and broadly similar to that encountered in medical research. Increased awareness of and emphasis on the reporting of CIs is needed to ensure that findings from dental research can be interpreted correctly by educators, clinicians, and patients.

Acknowledgments

The authors reported no conflicts of interest related to this study.

References

1. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. BMC Med 2010;24:8–18.
2. Gagnier JJ, Kellam PJ. Reporting and methodological quality of systematic reviews in the orthopaedic literature. J Bone Joint Surg Am 2013;95:e771–e777.
3. Adie S, Harris IA, Naylor JM, Mittal R. CONSORT compliance in surgical randomized trials: Are we there yet? A systematic review. Ann Surg 2013;258:872–878.
4. Fleming PS, Buckley N, Seehra J, Polychronopoulou A, Pandis N. Reporting quality of abstracts of randomized controlled trials published in leading orthodontic journals from 2006 to 2011. Am J Orthod Dentofacial Orthop 2012;142:451–458.
5. Fleming PS, Seehra J, Polychronopoulou A, Fedorowicz Z, Pandis N. A PRISMA assessment of the reporting quality of systematic reviews in orthodontics. Angle Orthod 2013;83:158–163.
6. Altman DG. Confidence intervals in practice. In: Altman DG, Machin D, Bryant TN, Gardner MJ, eds. Statistics with Confidence: Confidence Intervals and Statistical Guidelines. 2nd ed. London: British Medical Journal Books, 2000:6–14.
7. Gardner MJ, Altman DG. Confidence intervals rather than p values: Estimation rather than hypothesis testing. Br Med J (Clin Res Ed) 1986;292:746–750.
8. Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med 1999;130:995–1004.
9. Polychronopoulou A, Pandis N, Eliades T. Appropriateness of reporting statistical results in orthodontics: The dominance of *P* values over confidence intervals. Eur J Orthod 2011;33:22–25.
10. Savitz DA. Is statistical significance testing useful in interpreting data? Reprod Toxicol 1993;7:95–100.
11. Barnett ML, Mathisen A. Tyranny of the p-value: The conflict between statistical significance and common sense. J Dent Res 1997;76:534–536.
12. Chia KS. “Significant-itis”—An obsession with the *P*-value. Scand J Work Environ Health 1997;23:152–154.
13. 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 control trial. Survey of 71 “negative” trials. N Engl J Med 1978;299:690–694.
14. Vavken P, Heinrich KM, Koppelhuber C, Rois S, Dorotka R. The use of confidence intervals in reporting orthopaedic research findings. Clin Orthop Relat Res 2009;467:3334–3339.
15. Kim JS, Kim DK, Hong SJ. Assessment of errors and misused statistics in dental research. Int Dent J 2011;61:163–167.
16. Lucena C, Lopez JM, Pulgar R, Abalos C, Valderrama MJ. Potential errors and misuse of statistics in studies on leakage in endodontics. Int Endod J 2013;46:323–331.
17. Altman DG, Goodman SN, Schroter S. How statistical expertise is used in medical research. J Am Med Assoc 2002;287:2817–2820.
18. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Syst Rev 2012;1:1–60.
19. Pandis N, Polychronopoulou A. Eliades T. An assessment of quality characteristics of randomized control trials published in dental journals. J Dent 2010;38:713–721.
20. Polychronopoulou A, Eliades T, Taoufik K, Papadopoulos MA, Athanasiou AE. Knowledge of European orthodontic postgraduate students on biostatistics. Eur J Orthod 2011;33:434–440.
21. Fidler F, Thomason N, Cumming G, Finch S, Leeman J. Editors can lead researchers to confidence intervals, but can't make them think: Statistical reform lessons from medicine. Psychol Sci 2004;15:119–126.
22. Rothman, K.J. A show of confidence. N Engl J Med 1978(b); 299:1362–1363.