

# Analysis of the use of sample size calculation and error of method in researches published in Brazilian and international orthodontic journals

David Normando\*, Marco Antonio de Oliveira Almeida\*\*, Cátia Cardoso Abdo Quintão\*\*\*

## Abstract

**Introduction:** Reliable sample size and an appropriate analysis of error are important steps to validate the data obtained in a scientific study, in addition to the ethical and economic issues. **Objective:** To evaluate, quantitatively, how often the researchers of orthodontic science have used the calculation of sample size and evaluated the method error in studies published in Brazil and in the United States of America. **Methods:** Two major journals, according to CAPES (Brazilian Federal Agency for Support and Evaluation of Graduate Education), were analyzed through a hand search: *Revista Dental Press de Ortodontia e Ortopedia Facial* and the *American Journal of Orthodontics and Dentofacial Orthopedics* (AJO-DO). Only papers published between 2005 and 2008 were examined. **Results:** Most of surveys published in both journals employed some method of error analysis, when this methodology can be applied. On the other hand, only a very small number of articles published in these journals have any description of how sample size was calculated. This proportion was 21.1% for the journal published in the United States (AJO-DO), and was significantly lower ( $p= 0.008$ ) for the journal of orthodontics published in Brazil (3.9%). **Conclusion:** Researchers and the editorial board of both journals should drive greater concern for the examination of errors inherent in the absence of such analyses in scientific research, particularly the errors related to the use of an inadequate sample size.

**Keywords:** Biostatistics. Sample size. Error of methods.

**How to cite this article:** Normando D, Almeida MAO, Quintão CCA. Analysis of the use of sample size calculation and error of method in researches published in Brazilian and international orthodontic journals. *Dental Press J Orthod.* 2011 Nov-Dec;16(6):33.e1-9.

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

\* Professor of Orthodontics, Federal University of Pará (UFPA). PhD in Orthodontics, State University of Rio de Janeiro (UERJ). MSc in Dentistry, University of São Paulo (FOUSP). Specialist in Orthodontics, PROFIS-USP.

\*\* MSc and PhD in Orthodontics, Federal University of Rio de Janeiro (UFRJ). Head Professor of Orthodontics, UERJ.

\*\*\* MSc and PhD in Orthodontics, UFRJ. Associate Professor of Orthodontics, UERJ.

## INTRODUCTION

Scientific studies are more reliable when they are carefully planned. The problem to be investigated must be well defined and operationalized from samples randomly obtained from appropriate populations. The methods should be followed carefully using measurements obtained with a reliable and previously calibrated instrument.<sup>6</sup> Finally, the study should have a sample size appropriate to its goals and large enough so that a clinically important effect would also be statistically significant.

The sample sizing is a mathematical process of decision of how many individuals or specimens should be included in the investigation and it must be carried out before collecting the data.<sup>5</sup> The sample sizing is important for ethical and economic purposes.<sup>7</sup> A study using a small sample cannot produce useful results, exposing the participants (sample) to unnecessary risks, while a sample with an excessive size uses more resources than necessary, besides putting in risk an unnecessary number of individuals.

Unfortunately, the natural distancing of health researchers and clinicians from the understanding of mathematics and, consequently, from the interest about statistical methods, coupled with the inaccessible language used by statisticians when communicating with health professionals, make this subject uninteresting to most researchers and clinicians<sup>12</sup>. The ignorance and disinterest increase the likelihood of errors in the design and analysis of the results of a scientific study, reducing the reliability of the data. Understanding these errors would help researchers to observe the reasons that led to their emergence and the necessary steps needed to minimize them.

## BASIC CONCEPTS

Assuming a natural distance of a majority of researchers and clinicians from statistical methods, we will outline in a brief review the concepts necessary for understanding the

calculations required to obtain an adequate sample size, and the errors inherent in the method used to measure the data.<sup>4,10</sup>

### Sample size calculation

When conducting the statistical analysis of results, a researcher, after choosing a statistical test, is subjected to two types of errors inherent to this phase: The type I error (alpha) and type II error (beta).<sup>7</sup> To facilitate the understanding of both types of errors we should consider, as an example, a statistical test used to determine whether two samples have significant differences, and compare it to a trial that determines whether a defendant is guilty or innocent.

### Type I error ( $\alpha$ ) or false positive

At the end of all judicial proceedings the jury should give its sentence, finding the defendant innocent or guilty, considering that, in principle, everyone is innocent until proven otherwise. Statistically, this principle is defined as the “null hypothesis” ( $H_0$ ), i.e., the data obtained in the study did not produce sufficient evidence to consider the samples as significantly different.

By declaring a defendant guilty, the jury may be incurring an error if the defendant is innocent indeed. This jury’s mistake illustrates, in the statistical analysis, the type I error or false positive. In the case exemplified, the samples are considered to be statistically different, but in fact are similar, or were obtained from the same population.

In statistical analysis the probability of type I error (alpha error, or false-positive) is defined by the p-value, obtained through any statistical test. The lower the p-value, the lower the possibility (probability) of error by stating, in the study, that the difference obtained when comparing the two samples is statistically significant or has occurred by random.<sup>10</sup> Great is the difficulty of beginner researchers to understand p-value, since they do not understand the value obtained as a

probability of error. Thus, the lower its value, the greater is the statistical significance.

The level of type I error ( $\alpha$ ) must be defined before the beginning of the study, and in dentistry it is usually defined as less than or equal to 5%. Therefore, the maximum probability of error accepted to reject the null hypothesis is 5% ( $p < 0.05$ ).

### Type II error ( $\beta$ ) or false negative

Back to the judicial proceeding...

Another possibility of error would occur if the jury found the defendant not guilty (accepting the null hypothesis), and the defendant was truly guilty. Statistically the null hypothesis (equality) would be accepted when, in fact, the samples were different, or were drawn from different populations. This type of error, known as beta ( $\beta$ ) or false negative, may occur by chance, but is more common when the sample size is very small to achieve a statistically significant difference.<sup>10,13</sup> Statistical analysis did not have enough power to read a real difference.<sup>6</sup>

Commonly, we find scientific articles in orthodontic literature in which researchers observe a clinically significant difference between two groups. However, when evaluating the data statistically, a significant difference ( $p > 0.05$ ) is not achieved. Also, a very common explanation is that “probably with a larger sample size we could have obtained a statistically significant difference between the groups.” When researchers observe a clinically significant difference between groups, but without statistical significance, the study is defined as “underpower.”

Power in statistics is the probability of a study to avoid a false negative result,<sup>10</sup> i.e., in the example cited before, to define the samples as similar when they really are similar. Thus, statistical power expresses the probability of detecting a true effect. Conventionally, power has been established as 80% or 90% (0.8 or 0.9) and is equal to  $1 - \beta$ . So for the power to be at least equal to 0.8,

the probability of the type II error ( $\beta$ ) must not be greater than 20% (0.2). Power =  $1 - 0.2 = 0.8$ . Small sample size reduces the power of a study, however if you provide excessively large samples a statistically significant studied is obtained, even when the difference is too small to be considered clinically important.<sup>7</sup>

This review about common errors in statistics reinforces the need to design studies with adequate sample size in order to avoid misinterpretations of the results and, therefore, inappropriate clinical treatment. However, beyond the level of  $\alpha$  and  $\beta$ , other factors are part of the recipe that defines the sample size:<sup>1,6,10</sup>

a) The minimum clinical effect that the researcher expects on the primary outcome. The lower the effect to be examined, a larger sample size (n) will be required.

b) How the data will be measured (continuous scale, ordinal, or nominal). Nominal (e.g., Presence x Absence; Class I x Class II x Class III) or ordinal data (e.g., severe pain, moderate, mild) require larger samples when compared to continuous numeric data (e.g., cephalometric measurements).

c) The type of statistical test being used, and whether it is uni- or bilateral. The type of test is chosen from the type of variable measured and the type of data distribution, normal or free. Parametric tests, which use continuous data with normal distribution (e.g., t test / ANOVA / Pearson correlation) require samples with a smaller number of participants, compared with non-parametric tests (e.g., Chi-square / Fisher's exact test / Mann-Whitney / Wilcoxon / Spearman correlation).<sup>8</sup> A tutorial on selecting the appropriate statistical test<sup>8</sup> for a given study can be obtained in [www.dentalpress.com.br/bioestatistica](http://www.dentalpress.com.br/bioestatistica).

d) The variability of the data. The higher the data variability (standard deviation, variance, inter-quartile deviation), the greater must be the size of the sample. In general, paired outcomes assume better control of data variability.

An important question then arises: Will the orthodontic science researchers have actually bothered to properly size the samples used in their studies? Recently the Institute of Dental and Craniofacial Research in the United States required the inclusion of all this information in clinical research submissions to obtain funding from this institution.<sup>9</sup>

Although absent in the dental literature, few studies have been published in medical literature about the adequacy of sample size or power of statistical tests used in clinical research. In general, the analysis of studies that fail to observe statistically significant differences between a treated group and control reveals the use of underpowered sample size to read differences clinically important, indicating the usual presence of the type II error or false negative.<sup>2,11</sup>

A study analyzing 2000 clinical trials on the treatment of schizophrenia<sup>10</sup> found that only 1% of a published research (n=20) presented an evaluation of the statistical power or sample size calculation. The authors noted that only 3% of the studies analyzed (n=60) had a sample size powered enough to read 20% of difference between the results of the treatments employed.

### **Error of the method**

Another type of error that can occur in scientific studies is related to the reliability of the data obtained, i.e., the ability of these data to represent the truth about the phenomenon that is being examined, and to be reproducible at a later date.<sup>4,13</sup>

Although common among studies with continuous variables (e.g. cephalometric measurements),<sup>13</sup> the calculation of error of measurements should also be examined for non-parametric data (e.g. epidemiological studies on the prevalence of malocclusion). However, as it is not always possible to collect the appropriate sample size for a given study, for ethical or economic issues, in some other situations it is impossible to

make a proper analysis of the error of the method. For example, when the loss of the specimen is inevitable during data collecting. This loss is often common during the shear bond tests of orthodontic brackets, since it is impossible to repeat the exact procedures after bracket debonding.

The error of a particular method is achieved using doubled measurements made at a given time interval. Two types of errors are typically examined: The random error, which measures the degree of precision for a specific variable and the systematic error, which evaluates the reproducibility of a given measurement.<sup>4</sup> A large part of the variance of an outcome is related to the inaccuracy of measurements, which also has a huge implication for the sample size.<sup>3</sup> The smaller the errors of method, the greater the validity and reliability of data.<sup>3</sup> Measures should therefore have “precision and be reproducible.” Precision is defined as the ability of the method to make an accurate reading of the measurement examined, “reproducibility” is the possibility of this measure to be repeated at different times under the same conditions.

Although widely required in the most renowned and highest impact journals, we can still observe numerous scientific papers without proper examination of sample sizes, and the errors of methods used to obtain the data.

### **PROPOSITION**

Taking into account the importance of an adequate sample size in scientific research, as well as the reliability and reproducibility of the data obtained from these samples, this study aims to analyze the frequency of reported use of the sample size calculation and error of the method in Brazilian and international orthodontic literature.

### **MATERIAL AND METHODS**

Two orthodontic journals were selected and evaluated in this study. One edited in Portuguese and published in Brazil, the *Revista*

*Dental Press de Ortodontia e Ortopedia Facial* (Dental Press, Maringá, Brazil) and the other edited in English and published in the United States, the *American Journal of Orthodontics and Dentofacial Orthopedics*, AJO-DO (Elsevier, Saint Louis, USA). These two journals were selected because they represent the orthodontic journals of the highest scientific impact in Brazil and worldwide respectively, according to CAPES (Brazilian Federal Agency for Support and Evaluation of Graduate Education).

The analysis of published articles was initially made by their abstract, where it was observed whether the paper described any statistical results. If so, a single examiner began reading the Materials and Methods and Results sections. A close reading of these sections was performed to assess whether or not the article presented an analysis of the sample size calculation, or “power” of the statistical test used, and the description of the error of method, where possible to be applied. There was no judgment on the appropriateness of the statistical methods used and/or reliability of the results.

Both in the Dental Press journal and in the AJO-DO only articles published in the “original articles” section were evaluated. In these sections, we have examined only papers with some descriptive and/or analytical statistics. Case reports and literature reviews were not examined, for obvious reasons.

### Sample size calculation

Initially a pilot study was done from the reading of the 20 original articles published in the Brazilian journal (Dental Press) in 2007 (v. 12, n. 4, 5 and 6) in order to obtain the proportion of published articles describing the sample size calculation in this journal. In this initial search only one article, among the 20 examined, described the sample size calculation, which corresponds to a ratio of 5%. Reproducibility was possible in 16 (80%) of the

20 studies analyzed. The analysis of these 16 articles revealed the assessment of the method error in 8 papers, i.e. a ratio of 50%.

The sample size calculation was obtained by BioEstat 5.0 software (a freeware available at [www.mamiraua.org.br](http://www.mamiraua.org.br)), considering the use of a binomial two-tailed statistics with 80% power (0.8) and alpha level of 5% able to read a 20% difference between the articles published between the two journals. Therefore, the sample size was defined as 49 items for each group for sample size calculation and an analysis of 93 articles per group for the reading of the use of method error.

The more recent volumes of the two journals were selected for evaluation because they represent the most current trend of the articles published in both journals. In order to obtain the sample size to evaluate the use of sample size calculation or statistical power, it was necessary to read all articles published in the Dental Press journal in 2007 (v. 12, n. 1, 2, 3, 4, 5 and 6), and one number published in 2006 (v. 11, n. 6), totaling 51 articles published in seven issues of the journal. For the AJO-DO journal, it was necessary to read issues of the journal published in five months (v. 132, n. 3, 4, 5, 6, year 2007) and one issue published in January 2008 (v. 133, n. 1) with a total of 57 articles.

Among the 51 articles examined in Dental Press journal and 57 articles in the AJO-DO, it was considered possible to evaluate the error of the method in 43 studies published in the Dental Press and 41 of the AJO-DO. Thus to reach the number determined on the sample size calculation for reading a difference of 20% ( $\alpha=5\%$ ,  $\beta=20\%$ ), it was required to read six more issues of the Dental Press journal (v. 11, n. 1, 2, 3, 4 and 5; v. 10, n. 6) and 7 issues of the AJO-DO (v. 132, n. 1 and 2; and v. 131, n. 2, 3, 4, 5 and 6), totaling 49 more articles published in the Dental Press journal (total  $n=92$ ) and 53 more in the AJO-DO (total  $n=94$ ).

### **Error of method**

In order to assess the reproducibility of the method, 15 articles published in each journal (n=30) were randomly reevaluated one week later by the same examiner. The Kappa test with a confidence level of 95% ( $\alpha=0.05$ ) was used to analyze the replicability of the study. The level of agreement was set for the presence of sample size calculation as to the error of method analysis.

### **Statistical analysis**

Frequencies observed for the use of sample size calculation and error of method in both journals were compared using the Binomial test for two proportions. The confidence level was set at 95% ( $\alpha=0.05$ ).

## **RESULTS AND DISCUSSION**

### **Reproducibility for the evaluation of the articles**

The error analysis revealed an excellent reproducibility in re-reading the articles for the use of sample size calculation, with 100% agreement (Kappa = 0.9,  $p < 0.001$ , 95CI = 1.0–0.53). Regarding the analysis of the reproducibility to describe the use of the method error, a concordance level of 97% (Kappa= 0.79,  $p < 0.001$ , 95CI = 1.0–0.44) was observed.

These preliminary results show an excellent reproducibility of the method used to evaluate the papers.

### **The use of a sample size calculation**

Planning the sample size is often important and difficult to run, requiring careful consideration in choosing the scientific objectives and appropriate information even before beginning the study.

The results of the present study revealed that only 3.9% of articles published in the Brazilian orthodontic journal, in the years 2005-2007, presented some information regarding

sample size calculation or power. This percentage was significantly lower ( $p=0.008$ ) when compared to articles published in AJO-DO which had just over 21% of the articles describing the methods used for sample size calculation (Table 1). These data reflect a worrying lack of attention of researchers and reviewers to this important factor, leading to a strong potential of introducing errors in the statistical evaluation of data published in these two important journals.

### **Study of method error**

Regarding the concern about the assessment of method error, our findings revealed a different picture from that obtained for the use of sample size calculation. For the Dental Press journal, in just over 15% of published articles it was not possible for authors to reproduce the measurements made in the study and thus to carry out error analysis. For the AJO-DO, the study of error was considered impossible to run on 28.1% of published articles (Table 2).

Among the papers where it was possible to run the error analysis, most reported to use some type of analysis (Table 2). For the journal published in Brazil, 60.9% of the articles (n=56) in which it was possible to study the error of the method, had performed such analysis, while for the AJO-DO, 76.6% of the articles (n=72) presented the study of error of method (Table 2).

Comparatively, for papers published in the AJO-DO, the authors (76.6%) seem to be a little more concerned with the study of the error of the method than in those articles published in Dental Press (60.9%,  $p=0.02$ ). However, we emphasize the large number of scientific papers published in the Dental Press journal that have used a tool to evaluate the error of method. We must also consider the fact that the articles analyzed in the AJO-DO have been published in the last 12 months (Feb. 2007 to Jan. 2008), while in the Dental Press journal it

was necessary to include articles published in 2005 and 2006 in order to compose a sample with an appropriate size.

However the data show that the articles published in the two journals have a relatively small level of concern for the application of methods to calculate the sample size. The result raises concerns for both journals, but is most critical for that published in Brazil where,

among 51 articles that were analyzed and published between late 2005 and late 2007, only 2 (3.9%) examined the sample size.

On the other hand, the data are better regarding the use of the error of the method. Most studies published in both journals, in which such analysis was possible, reported its use although that number is slightly higher for the AJO-DO.

TABLE 1 - Absolute and relative frequency (%), relative difference, p-value and power for papers with statistical analysis (A) and the description of sample size calculation (B), published in the *Revista Dental Press de Ortodontia e Ortopedia Facial* and in the *American Journal of Orthodontics and Dentofacial Orthopedics* (AJO-DO).

Description	Dental Press 2005-2007	AJO-DO 2007-2008	Difference (%)	p-value	Power
Original articles with statistical analysis (A)	51 (100%)	57 (100%)	----		
Articles with sample size calculation (B)	2 (3.9%)	12 (21.1%)	-17.2	0.008**	0.78

\*\* p<0.01.

TABLE 2 - Absolute and relative frequency (%), relative difference, p-value and power for papers with possible analysis of error of the method (C) and the description of this analysis (D), published in the *Revista Dental Press de Ortodontia e Ortopedia Facial* and in the *American Journal of Orthodontics and Dentofacial Orthopedics* (AJO-DO).

Description	Dental Press 2005-2007	AJO-DO 2007-2008	Difference (%)	p-value	Power
Original articles with statistical analysis (C)	92 (84.3%)	94 (71.9%)	12.4		
Articles with sample size calculation (D)	56 (60.9%)	72 (76.6%)	-14.6	0.02 *	0.64

\* p<0.05.

## CONCLUSION

Most researches published in the *Revista Dental Press de Ortodontia e Ortopedia Facial* and in the *American Journal of Orthodontics and Dentofacial Orthopedics* adopt some method to evaluate method error, however only a very small number of these articles showed a sample size calculation. This ratio, already small (21.1%)

for the journal published in the United States (AJO-DO) is significantly lower ( $p=0.008$ ) in the orthodontics journal edited in Brazil (3.9%). Researchers, reviewers and the editorial board of both journals should address this major concern regarding the evaluation of the limitations inherent in analysis of scientific research, especially the errors when an improper sample size is used.



## REFERENCES

1. Browner WS, Newman TB. Sample size and power based on the population attributable fraction. *Am J Public Health*. 1998;79(9):1289-94.
2. 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(13):690-4.
3. Houston WJ. The analysis of errors in orthodontic measurements. *Am J Orthod*. 1983;83(5):382-90.
4. Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. *Designing clinical research — an epidemiologic approach*. 2<sup>nd</sup> ed. Baltimore: Williams and Wilkins; 2001.
5. Last JM. Making the dictionary of epidemiology. *Int J Epidemiol*. 1996;25(5):1098-101.
6. Lenth RV. Some practical guidelines for effective sample size determination. *Am Statistician*. 2001;55(3):187-93.
7. MacFarlane TV. Sample size determination for research projects. *J Orthod*. 2003;30(2):99-100.
8. Normando D, Tjäderhane L, Quintão CCA. A PowerPoint-based guide to assist in choosing the suitable statistical test. *Dental Press J Orthod*. 2010;15(1):101-6.
9. National Institute of Dental and Craniofacial Research. Policies and Procedures for Investigator Initiated Clinical Trials. [Cited 2008 Feb 10]. Available from: [www.nidcr.nih.gov/ClinicalTrials/ClinicalTrialsProgram/DataCoordinator.htm](http://www.nidcr.nih.gov/ClinicalTrials/ClinicalTrialsProgram/DataCoordinator.htm).
10. Phillips C. Sample size and power. What is enough? *Semin Orthod*. 2002;8:67-76.
11. Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ*. 1998;31(317):1181-4.
12. Torgerson DJ, Miles JN. Simple sample size calculation. *J Eval Clin Pract*. 2007;13(6):952-3.
13. Valladares JVN, Domingues MHMS, Capelloza Filho L. Pesquisa em Ortodontia: bases para a produção e a análise crítica. *Rev Dental Press Ortod Ortop Facial*. 2000;5(4):89-105.

Submitted: August 7, 2008

Revised and accepted: August 31, 2008

---

### Contact address

David Normando

Rua Boaventura da Silva, 567-1201

Zip code: 66.055-090 – Belém / PA, Brazil

E-mail: [davidnor@amazon.com.br](mailto:davidnor@amazon.com.br)