Artigo de Revisão

REVISTA BRASILEIRA DE CIÊNCIA & MOVIMENTO
ISSN 0103-1716

A review of the statistical methods used in sports and exercise sciences PhD theses: a case study in a single post-graduate program

Uma revisão dos métodos estatísticos usados em teses de doutorado nas ciências do esporte e do exercício: um estudo de caso em um único programa de pós-graduação

FIGUEIREDO P, MARQUES EA, BACKES OP, GAYA AR, CASTRO FS. **A** review of the statistical methods used in sports and exercise sciences PhD theses: a case study in a single post-graduate program. **R. bras. Ci. e Mov** 2018;26(4):191-200.

ABSTRACT: This study aimed to describe the statistical methods used in PhD thesis in sports and exercise sciences, and to examine the time trends of statistical methods prevalence. We analyzed 50 PhD theses supported by a post-graduate program recognized by CAPES and defended between January 2003 and December 2013. Most theses used more than two statistical methods, and T-test, ANOVA, multiple comparisons, correlation and reliability analysis were the most used approaches. Less than 50% of both experimental and observational theses reported the effect size (ES), confidence interval (CI), sample size calculation and observed power analysis. The use of multiple comparisons in experimental studies significantly increased in the last 5-years (63.2%) compared to 2004-2008 period. No other significant changes were observed. Results suggest that several statistical methods are used in quantitative design studies. Due to the observed large prevalence of statistical deficiencies (lack of reporting ES, CI, power analysis), postgraduate programs should have qualified statisticians with a major say in checking the statistical quality of PhD theses and subsequent articles, besides supervisors more concerned to the statistical methods.

Key Words: Physical education; Statistics; Research design.

***ABSTRACT*:** Este estudo teve como objetivo descrever os métodos estatísticos utilizados em teses de doutorado em ciências do esporte e do exercício e examinar as tendências temporais da prevalência desses métodos. Foram analisadas 50 teses de doutorado desenvolvidas em programa de pós-graduação reconhecido pela CAPES e defendidas entre janeiro de 2003 e dezembro de 2013. A maioria das teses utilizou mais de dois métodos estatísticos. Teste t, ANOVA, comparações múltiplas, correlação e análise de confiabilidade foram as abordagens mais utilizadas. Menos de 50% das teses experimentais e observacionais relataram o tamanho do efeito (ES), intervalo de confiança (CI), cálculo do tamanho da amostra e análise de potência observada. O uso de comparações múltiplas em estudos experimentais aumentou significativamente (63,2%) nos últimos 5 anos em relação ao período de 2004-2008. Não foram observadas outras alterações significativas. Os resultados sugerem que vários métodos estatísticos são usados em estudos de design quantitativo. Devido à grande prevalência observada de deficiências estatísticas (falta de notificação de ES, IC, análise de potência), os programas de pós-graduação devem ter estatísticos qualificados com grande influência no controle da qualidade estatística das teses de doutorado e artigos subsequentes, além disso, orientadores deveriam ter maior preocupação em relação aos métodos.

Palavras-chave: Educação física; Estatística; Desenho de pesquisa.

Pedro Figueiredo[1,2]
Elisa A. Marques[1]
Othavio Porto Backes[3]
Anelise Reis Gaya[3]
Flávio de Souza Castro[3]

[1]Research Center in Sports Sciences, Health Sciences and Human Development, CIDESD, University Institute of Maia
[2]Portugal Football School, Portuguese Football Federation
[3]Universidade Federal do Rio Grande do Sul

Contato: Flávio de Souza Castro - souza.castro@ufrgs.br

Recebido: 24/12/2017
Aceito: 22/05/2018

Introdução

Appropriate use of statistical methods is becoming increasingly important in sports and exercise sciences research[1]. There is consensus that inappropriate study designs and statistical methodology lead to incorrect results, poor interpretation of study findings and wrong conclusions. Many journals provide authors with statistical guidelines and have a dedicated statistical committee, which analyses the methods used. In addition, few review articles[1] provide a view of best statistical practices specifically applicable in sports and exercise field.

Studies focusing on analyzing the Physical Education Brazilian journals have reported that a high amount of research does not follow the basic assumptions for statistical use[2,3]. Considering that one of the key indicators of scientific production is the number of papers published in international journals, better designs and statistical procedures should be developed.

In Brazil, the number of postgraduate students and scientific publication has grown 22% in 2010-2012 when compared with the previous triennium. Also the quality of the programs improved, as researchers are publishing in journals with higher impact factors. In fact, the proportion of articles published in higher quality journals accounted for 1/3 of the total publications[4]. Nevertheless, data from three Physical Education postgraduate programs in 1998 and 2002 showed that 72% of the theses and dissertations did not result in article publication and only about 15% of the articles were published in indexed journals[5].

Considering the science progress, it is necessity to publish the results from PhD theses, as publishing findings, hypotheses, and theories. In addition, the postgraduate programs where these PhD theses are developed are evaluated regularly by national entities stimulating the need to increase the number of publications with high quality (including appropriate reporting and interpretation). For example, each researcher, in regularly evaluated programs, should publish a minimum of articles in journals considered of superior quality.

To accomplishing these goals, sound training at the postgraduate level in statistical concepts and methods of analysis are crucial to overcome the documented inadequacies in reporting important aspects of research design and statistical analyses[2,3]. A specialized supervisor inclusion, regarding statistics, could be either an option.
Although it is imperative to understand which study designs and statistical techniques are used in sports sciences, limited information is available. Thus, this study reviews PhD theses in sports and exercise sciences field with the aim to describe the statistical methods used and if they report commonly recommended statistical techniques. As a secondary analysis, we examined the time trends of statistical methods prevalence.

Methods

Criteria for selection of PhD Theses

The inclusion criteria for this study was PhD theses solely based in quantitative research, defined as a formal, objective, systematic process which generates numerical data and usually seeks to establish causal relationships between two or more variables, using statistical methods to test the strength and significance of the relationships. Mixed methods thesis combine qualitative and quantitative approaches to research were included. We excluded qualitative research-based dissertations, and *in vitro* and animal studies.

Data collection and analysis

In this analysis, we searched all PhD theses, in the University digital repository, supported by a post-graduate program recognized by CAPES, defended between January 2003 and December 2013. This period was further split into three periods (2003-2006; 2007-2010; 2011-2013) to summarize the data. Then, to assess changes over time, further split into two periods (January 2004-December 2008 and January 2009-December 2013). Thus, the year of 2003 was removed for the time trend analysis in order to form two comparable periods (5-year length).

Data extraction and management

Data extraction was completed using a pilot-tested and revised coding frame to record information on several details. The major categories of variables coded included publication year, study design, subgroups of study designs, and sample characteristics.

All data were coded and reviewed for accuracy and consistency by two researchers in an independent way. If they disagree, a third researcher was consulted. There was no judgment on the level of appropriateness of the statistical methods used. Statistical data were recorded based on the classification describe in Table 1 modified from previously proposed classification[6,7]. When the same statistical technique was used more than once in a same thesis, we recorded it as one single event.

Table 1. Categories of statistical procedures used to assess the statistical content and statistical practice classification of theses.

Category	Brief description
Statistical procedures	
Descriptive statistics only	No statistical content, or descriptive statistics only (e.g., percentages, means, standard deviations, standard errors, histograms)
t-tests	Procedures for comparing two means (one-sample, matched-pair, and two-sample t-tests)
ANOVA	Comparing several means (e.g., Simple and repeated one-way Analysis of variance, analysis of covariance, factorial ANOVA, mixed ANOVA, F-tests)
Multiple comparisons	Procedures for handling multiple inferences (e.g., Newman-Keuls, Bonferroni, Tukey, Scheffé's contrasts, Duncan tests)
Non-parametric tests	Non-parametric statistics/distribution-free tests (e.g., Sign test, Wilcoxon signed-rank test, Mann-Whitney test, Kruskal-Wallis, Friedman tests)
Correlation	Pearson correlation coefficient and testing
Non-parametric correlation	Spearman's rho, Kendall's tau test, test for trend
Simple Linear regression	Least-squares regression with one predictor and one outcome variable
Multiple regression	Complex model with several predictors (Hierarchical regression, Forced entry or Enter regression, and Stepwise methods)
Contingency/Multiway tables	Chi-square tests, Fisher's exact test, McNemar's test, Mantel-Haenzel procedure, log-linear methods
Reliability	Intra-class correlations coefficient, Cronbach's alpha, Cohen's kappa
Epidemiology statistics	Measures of association (e.g., relative risk, odds ratio, rate ratio)
Other	Anything not fitting above headings, includes Factorial analysis, ROC curves, cluster analysis, structural equation modeling
Statistical practice classification	
NHST	When p values, or mention of null hypothesis or statistical significance was included
Effect size	At least one measure of effect size was reported and recognized as such by authors.
Confidence interval	At least one confidence interval was reported (tables, figures, or text)
Sample size calculation	Sample size calculation was described throughout the text
Observed power analysis	Observed power (or post-hoc power) was reported

ANOVA - Analysis of variance; ROC - receiver-operating characteristics; NHST- null hypothesis significance testing.

Statistical analysis

Descriptive statistics were used to describe the characteristics of the included studies. Frequency analysis was used to determine the number of studies reporting each statistical category. For each study design, methods reported in the two time periods were compared using the Chi-square statistics or Fisher's exact test, and Mann-Whitney test when appropriate. Data were analyzed using SPSS 23.0 software and p-values <0.05 were considered significant.

Results

From the 74 PhD theses screened, 21 were excluded because they used only a qualitative research methods approach. Fifty-three quantitative research theses were identified for potential inclusion and full-text reports were analyzed. Two animal-based and one in vitro-based theses were then excluded, and 50 theses entered in the analysis. One thesis was classified simultaneously as observational and experimental, because it included 2 distinct studies, each with a different study approach. All other theses were classified in one type of study design only. Therefore, the final number of analyzed studies was 51, where 26 and 25 were classified as observational and experimental studies, respectively.

Of the 26 observational studies, 22 studies (84.6%) were analytical studies while a minority were descriptive studies (n=4). In addition, most of the analytical studies were cross-sectional (77.3%), while the other main types of observational studies were marginally represented (two cohort studies, two case report studies and one case-control study). Regarding the experimental studies, 48% included a control group and 13 (52%) studies compared two groups receiving different interventions.

The main characteristics of the 51 included studies are presented in Table 2. The sample sizes of the individual studies varied. As expected, the experimental studies had smaller sample sizes compared with the observational studies. In both types of studies participants were mostly children and/or adolescents and adults, while older adults were only included in the experimental studies. All together, 32 studies included men and women (62.7%) and 29.4% focused exclusively on men.

Table 2. Characteristics of included theses.

	Study Design	
Characteristic	**Observational (n=26)**	**Experimental (n=25)**
Study participants, n (%)		
Female	0 (0)	4 (16.0)
Male	7 (26.9)	8 (32.0)
Both	19 (73.1)	13 (52.0)
Age groups, n (%)		
Children and/or adolescents	12 (46.2)	8 (32.0)
Adults	7 (26.9)	6 (24.0)
Older adults	0 (0)	5 (20.0)
≥2 age groups	7 (26.9)	6 (24.0)
Sample size, n (%)		
Up to 25	7 (26.9)	11 (44.0)
26 to 50	3 (11.5)	5 (20.0)
51 to 150	3 (11.5)	5 (20.0)
151 to 250	1 (3.8)	1 (4.0)

>250	12 (46.2)	3 (12.0)
Mean sample size (min-max)	628.9 (1-5083)	114.7 (7-988)
Year, n (%)		
2003-2006	3 (11.5)	2 (8.0)
2007-2010	11 (42.3)	13 (52.0)
2011-2013	12 (46.2)	10 (40.0)

Table 3 presents the distribution of the statistical methods and practice used in the analyzed theses. The number of theses reporting use of ANOVA and multiple comparisons were higher among those using an experimental design, while simple linear regression and multiple regression were more often used in observational theses. The use of t-tests was 22.2% more frequent in the experimental studies. In addition, simple linear regression and epidemiological statistics where never used, while the ANOVA was the most common procedure. Although with lower frequency compared with the experimental studies (50% vs. 80%), the ANOVA was also the statistical category most applied in the observational studies. Most PhD theses (~85%) used between 2 and 5 different statistical tests. Less than 50% of both experimental and observational theses reported the effect size (ES), confidence interval (CI), sample size calculation and observed power analysis.

Table 3. Statistical content and practice classification by study design.

	Study Design			
	Observational (n=26)		**Experimental (n=25)**	
	No.	**%**	**No.**	**%**
Statistical methods				
Descriptive statistics only	1	4.0	1	4.0
T-tests	11	45.8	17	68.0
Analysis of variance	12	50.0	20	80.0
Multiple comparisons	7	29.2	17	68.0
Non-parametric tests	5	20.8	9	36.0
Correlation	8	33.3	6	24.0
Non-parametric correlation	5	20.8	2	8.0
Simple Linear regression	4	16.7	0	0
Multiple regression	9	37.5	1	4.0
Contingency/Multiway tables	8	33.3	4	16.0
Reliability	8	33.3	9	36.0
Epidemiology statistics	2	8.3	0	0
Other	6	25.0	2	8.0
Complexity of statistical analyses				
No. of different inferential methods				
Only 1 method	1	4.0	1	4.2
2 or 3 methods	11	44.0	11	45.8
4 or 5 methods	10	40.0	10	41.7
More than 5 methods	3	12.0	2	8.3

Statistical practice				
NHST	23	95.8	24	96.0
Effect size	5	20.8	3	12.0
Confidence interval	8	33.3	4	16.0
Sample size calculation	10	41.7	12	48.0
Observed power analysis	2	8.3	3	12.0

In the time-trend analysis (Table 4) only the number of experimental theses reporting multiple comparisons was significantly higher in the most recent years compared with the earlier 5-year period. Although between 2009 and 2013 the use of multiple comparisons and ANOVA was equally high (88.2%) in the experimental studies, a significant increase (63.2%) in the latest method was observed in the last period. Despite no other statistical significant differences were found, in the observational theses the use of ANOVA, contingency/multiway tables, and reporting the effect size decreased 22% in the last 5-year period. In the experimental theses the higher decrease (20.6%) was observed for the non-parametric tests, while the use of ANOVA methods and sample size calculation increased more than 25%. Moreover, only 20% of the theses using an experimental approach were in accordance with the general principles of randomized controlled trial (RCT).

Table 4. Statistical content and practice classification by study design and year of publication.

	Observational					**Experimental**				
	2004-2008 (n=6)		**2009-2013 (n=18)**			**2004-2008 (n=8)**		**2009-2013 (n=17)**		
Statistical methods	**No.**	**%**	**No.**	**%**	**Δ**	**No.**	**%**	**No.**	**%**	**Δ**
Descriptive statistics only	0	0	1	5.6	+5.6	0	0	1	5.9	+5.9
t-tests	3	50.0	8	44.4	-5.6	5	62.5	12	70.6	+8.1
Analysis of variance	4	66.7	8	44.4	-22.3	5	62.5	15	88.2	+25.7
Multiple comparisons	2	33.3	5	27.8	-5.5	2	25.0	15	88.2	**+63.2***
Non-parametric tests	1	16.7	4	22.2	+5.5	4	50.0	5	29.4	-20.6
Correlation	2	33.3	6	33.3	0	2	25.0	4	23.5	-1.5
Non-parametric correlation	1	16.7	4	22.2	+5.5	1	12.5	1	5.9	-6.6
Simple Linear regression	1	16.7	3	16.7	0	0	0	0	0	–
Multiple regression	3	50.0	6	33.3	-16.7	1	12.5	0	0	-12.5
Contingency/Multiway tables	3	50.0	5	27.8	-22.2	1	12.5	3	17.6	+5.1
Reliability	2	33.3	6	33.3	0	3	37.5	6	35.3	-2.2
Epidemiology statistics	1	16.7	1	5.6	-11.1	0	0	0	0	–
Other	1	16.7	5	27.8	+11.1	1	12.5	1	5.9	-6.6
Complexity of statistical analyses										
No. of different inferential methods										

Only 1 method	0	0	1	5.9	+5.9	1	12.5	0	0	-12.5
2 or 3 methods	2	33.3	8	47.1	+13.8	4	50.0	7	43.8	-6.3
4 or 5 methods	3	50.0	6	35.3	-14.7	3	37.5	7	43.8	+6.3
More than 5 methods	1	16.7	2	11.8	-4.9	0	0	2	12.4	+12.4
Statistical practice										
NHST	6	100	17	94.4	-5.6	8	100	16	94.1	-5.9
Effect size	1	16.7	4	22.2	+5.5	1	12.5	2	11.8	-0.7
Confidence interval	3	50.0	5	27.8	-22.2	1	25.0	1	11.8	-13.2
Sample size calculation	3	50.0	7	38.9	-11.1	2	25.0	10	58.8	+33.8
Observed power analysis	0	0	2	11.1	-11.1	1	12.5	2	11.8	-0.7

Δ = Difference between %; * significantly different from 2004-2008 (p=0.004).

Discussion

Our findings show that cross-sectional study designs were the most common type of observational design studies, and experimental studies were similarly represented by studies with a control group and studies comparing two different intervention groups. Both study designs (observational and experimental) were similarly used in the last decade, and between the two time periods. Most theses used at least two statistical tests. In the experimental studies, ANOVA and multiple comparisons techniques account for most of the statistical tests reported, and t-test and ANOVA were commonly used in the observational studies. The use of regression techniques was higher in observational studies. In both study designs the frequency of use of NHST without CI, ES, and power estimation was high (~75%). Only the frequency of use of multiple comparisons significantly increased between the two time periods.

Experimental design studies were mostly based in small samples (n <25). Consistent with our result, Zlowodzky and Bhandari[8] observed a tendency to use small samples in most studies. This sample size restriction has the effect of lowering the statistical power, decreasing the capacity of proving the hypotheses being tested. The observed difference in sample size between theses based on different study designs was expected, due to several common constraints associated with experimental studies such as cost, selection of participants, space and equipment available. Although the most powerful type of experimental study is RCT, only 20% of the theses using an experimental approach were in accordance with the general principles of RCT. Efforts have been made to improve the quality of the reports of trial findings. For example, the Consolidated Standards of Reporting Trials group developed a Statement, which is an evidence-based, minimum set of recommendations for a complete and transparent report of randomized trials[9]. Similar findings were also reported in a recent retrospective survey examining the research design and statistical methods in Indian medical journals[10], showing a even lower rate of RCT published in 2013 (7.3%).

We confirmed a significant use of inferential statistics, which was expected due to the quantitative nature of all selected theses. Therefore, discordant results have been found when considering all types of study designs[3]. Approximately 46% of all articles published in Physical Education Brazilian Journals between 2009 and 2011 did not used any type of statistical analysis. Consistent with our finding, in medical research the majority of the articles published in top journal contained inferential statistics[11].

Regarding the most commonly used statistical tests, our findings are consistent with previous reports showing that simple methods such as t-tests, ANOVA, multiple comparisons and chi-square tests remain in common usage[3,11,12]. However, the use of some inferential statistics such as methods of survival analysis were absent from our reports, although this methods are frequently reported in top medical journals[11]. The reasons for these differences may include differences in research topics and study purposes. Survival analysis is very common in medical, clinical and epidemiology research, where the major events that the trial subjects suffer are death, relapse, adverse drug reaction or development of a new disease.

Our findings showed that non-parametric tests are used in more than 20% of the observational studies and in 36% of the experimental studies. Non-parametric tests are usually less powerful than corresponding parametric tests, thus the probability to reject the null hypothesis when it is false is higher[13]. However, selecting rank based non-parametric test may indicate that more attention is being given to the assumptions of parametric test. Of note, in the time-trend analysis we noticed a relevant decrease in the use of these statistics in the experimental studies. In addition, the trend analysis also showed a significantly higher use of multiple comparisons in the experimental studies, which may suggest that the interpretation of multiple P-values is being correctly addressed. When a large number of statistical tests are performed, some will have *P*-values less than 0.05 purely by chance, even if all null hypotheses are really true. Methods such as Bonferroni, Tukey, Dunnett (and more) are usual approaches available to correct for multiple comparisons[14].

Our results concur with those reported earlier showing that most articles apply more than one statistical test6,11. Maybe the "Scandinavian" format in theses is an explanation for this case.

Regarding statistical practice outcomes, sample size calculation was reported in less than 50%, both for observational and experimental theses. In studies that do not use sample calculation, samples are often based on inaccurate and arbitrary assumptions, mainly driven by the study feasibility (convenience)[15]. Sample size calculation provides a suitable number of observations for certain variable, minimizing the probability of error in the interpretation, statistical power and validity of the results[2,16]. Despite the low values observed in the PhD theses, the frequency was higher than that reported in scientific studies published in national journals in the Physical Education area in the period 2010-2012[2]. Also in the latter no significant difference between publication years in the proportions of use of the sample size calculation was reported, as in the present study. Nevertheless, an increase of about 34% was observed with time in the experimental studies, suggesting a trend of improvement in the requirements for quality results and interpretation.

In Sport Sciences theses, NHST was very often used without CI, ES or Power estimation, as reported for high impact journals[7,11], despite the limitations of NHST. Thus, it would be beneficial to use ES and CI to support rejection or acceptance of the null hypothesis or to support results from descriptive statistics. The proportion of theses reporting ES, CI or Power estimation was low for both study designs. Values for CI and Power estimation were higher and ES lower than those reported by Fritz, Sherndl and Kühberger[17] in psychological journals in the period 1990-2010. Overall they found that approximately 10% used CI, 38% ES and only 3% power analysis.

Contrarily to the trends of some journals[7,11,12] CI, ES and Power analysis did not increment with time in our study, showing that this approaches are not being used as the basis of discussion and interpretation. However, reporting CI and ES does not guarantee that researchers use them in their interpretation of results, which means that the observed percentages reported may overestimate its effective use.

The primary strength of this study is the inclusion of all PhD theses based in quantitative approach from a postgraduate program in Sports and Exercise Sciences, which enabled us for an up-to-date in-depth description of the statistical methods and practice used. This analysis also includes a time trend analysis for monitoring possible changes in statistical usage.

There are, however, a number of limitations to be considered when interpreting the results. A potential weakness concerns the absence of a more detailed analysis of statistical methods used, i.e., it was not possible to determine whether they were appropriately used. It is important to realize that the quality of a study cannot be judged by the frequent use of statistical tests or their degree of complexity. The use of the appropriate study design and the use of careful methods to avoid the many sources of bias are similarly important as the choice of a particular statistical test. We emphasize the importance of caution in generalizing our evidence to other disciplines, postgraduate programs or countries.

Conclusion

In summary, our results suggest that several statistical methods are used in quantitative design studies in sports and exercise sciences, and the type of study design is associated with the use of some particular statistical approaches. Thus, the statistical education among exercise sciences postgraduates in Brazil should address a variety of statistical concepts that play an essential role in sports research and it cannot treat all broad groupings of statistical methods shown in Table 1. Rather, it may focus more appropriately on those statistical techniques most frequently used. Moreover, this approach should provide the student with sufficient skills to interpret a large proportion of journal articles using more advanced statistics. The use of several statistical tests in the same study is extremely high, which is favored by the availability of multifaceted statistical software packages that must be encourage during statistic courses (including both theoretical and practical training). In addition, these statistical software help toward the usage of more sophisticated techniques.

The large prevalence of statistical deficiencies (lack of reporting ES, CI, power analysis) were also observed, which in turn suggests that postgraduate programs should have qualified statisticians who should have a major say in checking the statistical quality of PhD theses and subsequent articles. In addition, supervisors most concerned with statistical methods might be needed, along with specific statistical supervisors. Finally, we recommend that observational studies should focus on studying the female population and older populations, as both groups were absent.

Funding: Programa Nacional de Pós-doutorado – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)

References

1. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc. 2009; 41(1): 3-13.

2. Crozara LF, Sponton ACdS, Jarrete AP, Araujo HN, Huamaní OG, Fernandes RA. Use of sample size calculation in scientific research in Brazilian Physical Education journals. Rev Bras Cineantropom Desempenho Hum. 2014; 16(5): 514-21.

3. Teixeira IP, Smirmaul BPC, Luchini PEH, Gobbi RB, Moura LP, Fernandes RA. Uso da estatística na Educação Física: análise das publicações nacionais entre os anos de 2009 e 2011. Rev Bras Educ Fís Esporte. 2015; 29(1): 139-47.

4. CAPES. Relatório de Avaliação 2010-2012. Available from: http://capes.gov.br/component/content/article?id=4666:educacao-fisica [2016 jun 15].

5. Ramos PS, Furtado EC, Carvalho ERF, Campos MO, Souza DVBC, Almeida LDP, *et al*. Does thesis and dissertations generate scientific papers? Analysis based in 3 physical education graduate programs. Brazilian Journal of Biomotricity. 2009; 3(4): 315-24.

6. Otwombe KN, Petzold M, Martinson N, Chirwa T. A review of the study designs and statistical methods used in the determination of predictors of all-cause mortality in HIV-infected cohorts: 2002-2011. PLoS One. 2014; 9(2): e87356.

7. Tressoldi PE, Giofre D, Sella F, Cumming G. High impact = high statistical standards? Not necessarily so. PLoS One. 2013; 8(2): e56180.

8. Zlowodzki M, Bhandari M. Outcome measures and implications for sample-size calculations. J Bone Joint Surg Am. 2009; 91(Suppl 3): 35-40.

9. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010; 340: c332.

10. Hassan S, Yellur R, Subramani P, Adiga P, Gokhale M, Iyer MS, *et al*. Research design and statistical methods in Indian medical journals: a retrospective survey. PLoS One. 2015; 10(4): e0121268.

11. Strasaka AM, Zamana Q, Marinella G, Pfeiffera KP, Ulmera H. The Use of Statistics in Medical Research: A Comparison of The New England Journal of Medicine and Nature Medicine. The American Statistician. 2007; 61(1): 47-55.

12. 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(2): 119-26.

13. Whitley E, Ball J. Statistics review 6: Nonparametric methods. Crit Care. 2002; 6(6): 509-13.

14. Ludbrook J. Multiple comparison procedures updated. Clin Exp Pharmacol Physiol. 1998; 25(12): 1032-7.

15. Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. BMJ. 2009; 338: b1732.

16. Torgerson DJ, Miles JN. Simple sample size calculation. J Eval Clin Pract. 2007; 13(6): 952-3.

17. Fritz A, Sherndl T, Kühberger A. A comprehensive review of reporting practices in psychological journals: Are effect sizes really enough? Theory & Psychology. 2013; 23(1): 98-122.