

# Biostatistical Analysis on Recent Issues : A Sampling study

Satyanarayan Choudhary

Saraswati Multiple Campus, Tribhuvan University

---

## Abstract

This paper aims at examining the statistical methodology used in the research articles published in the journals of high repute. Conducting the correct statistical analysis is of paramount importance in every research, and hence, statistical methodology has often been checked predominantly in journals. In this paper, the original research articles published in the top 10 leading medical journals published in India, and indexed by PubMed in 2003 and a decade later in 2013 have been reviewed. The present article examined—study design types, frequencies of statistical methods, errors/defects in study design and statistical analyses, and also the implementation of CONSORT checklist in RCT's (randomized clinical trials). Chi-square and Fisher's exact tests, (wherever appropriate) were used to compare any proportional differences in use of statistical methods and various errors/defects during the 10 years in the study period. All the tests used in this article were conducted at 5% level of significance and were 2-tailed. Chi-square tests were conducted at one degree of freedom. Using arbitrary  $p$  thresholds (like  $p < 0.01$ ) instead of reporting exact  $p$ -values, reporting  $p$  value without test statistics, insufficient (or inappropriate) description of methods and  $p$  values without confidence intervals were most common of the errors present. In conclusion, this study indicates that Indian medical research seems to have made some progress regarding study design defects, but there remains ample room for improvement regarding statistical analyses. Most published studies continue to use a retrospective clinical design, with randomized clinical trials being quite rare. Those RCTs that are published often have serious methodological problems, including absence of sample size estimation and power calculations, as well as failure in (or in the reporting of) randomization.

**Keywords:** journals, methodology, PubMed, randomized clinical trials, systematic review

## Introduction

Good quality medical research generally requires not only an expertise in the chosen medical field of interest but also a sound knowledge of statistical methodology. Since, medical practitioners and policy makers often depend on the correct conclusions obtained from these research papers, conducting the correct statistical analysis is of paramount importance. Hence, statistical methodology has often been checked for predominantly in Western journals. From 1970s, Altman et al. (1993) have continued to study errors/defects in several aspects of study design and statistical methodologies used in medical journals. Afterwards, a series of publication guidelines have been proposed, including CONSORT and STROBE statements. These guidelines have in general led to

some improvement in the quality of publications in clinical research worldwide.

But such checking for statistical inconsistencies is quite rare and is indeed absent in Indian journals. Previously, Goyal et. al. (2010) reported in an analysis of 46 clinical trials published in three journals—Indian Pediatrics, Indian Journal of Pharmacology, Journal of Postgraduate Medicine within the period of 2007–2008, that majority of papers had errors during the interpretation of results of these trials. This was found due to multiple treatment groups, repeated measurements of endpoints, multiple tests of significance, over reliability on P value and less use of confidence interval. Another paper by Lang et. al. (2014) analysed papers published in an 8 year span. Out of 196 papers which were analysed 150 papers had inappropriate descriptive methods, with the most common error being that Mean $\pm$ SEM (standard error of mean) was used in the place of “mean (SD)” or “mean  $\pm$  SD.” SEM generally quantifies uncertainty in estimate of the descriptive measure (like mean) whereas SD indicates dispersion of the data from the central tendency (like mean). As readers are generally interested in knowing the variability within sample, descriptive data should be precisely summarized with SD. About 600 biomedical and life science journals are published every year in India. But if we take a look at the indexing statistics it presents a bleak picture. To take an example, out of the 5500 plus journals within PubMed database, only 39 (0.71%) are from India. Similarly, in another database—EMBASE, number of Indian journals indexed in it is 128 (1.71%).

As more than 600 medical journals are being used within India by a large section of clinicians and readers, it has become increasingly more important to know and understand whether the statistical methodology used in Indian medical journals are suitable and to ascertain what all changes should be implemented for their improvement. In this article, original research articles published in 10 leading Indian medical journals based on 2013 Impact Factors published by Thomson Reuter have been reviewed. The study determined the frequency of study designs, statistical methods and results presentation/ interpretation used and the most prevalent statistical errors in these research articles. The study also examined all the RCT's (randomized clinical trials) published in these journals according to the CONSORT statement. In addition, we also discuss certain remedies which can lead to improvement of the publishing quality.

## **Materials and Methods**

### **Journals**

For this article, 10 leading medical journals published in India, and indexed by PubMed have been selected. The 10 journals are: Indian Journal of Medical Research, Indian Journal of Medical Sciences, Indian Journal of Otolaryngology & Head and Neck Surgery, The Indian Journal of Chest Diseases & Allied Sciences, Journal of Postgraduate Medicine, Indian Journal of Dermatology, Venereology, Leprology, Indian Journal of Cancer, Neurology India, Indian Pediatrics and Indian Journal of Ophthalmology.

All the original research articles published in these journals were peer reviewed. As majority of guidelines have come up during the past decade or so after the millennium, we intended to check how the Indian medical journals fare just before the introduction of many such reporting guidelines (for example, STROBE, STARD, SAMPL, TREND etc) and a decade after. This duration of a decade has seen the release of many such guidelines and hence many journals would adopt these guidelines leading to a significant change in the reporting errors. Hence, the paper examined all the original research articles published in 2003 and a decade later in 2013, as we would have access to all the articles in both these years. The present article included—study design types, frequencies of statistical methods, errors/defects in study design and statistical analyses, and also the implementation of CONSORT checklist in RCT's (randomized clinical trials).

### **Quality control**

A strict and high quality control was maintained throughout the entire study. For the quality standards of the different study designs were used such as the checklists—CONSORT, STROBE, STARD, SAMPL, STREGA, PRISMA, SPIRIT and TREND. In the early 1990's Iain Chalmers said that Failure to publish an adequate account of a well-designed clinical trial is a form of scientific misconduct that can lead those caring for patients to make inappropriate treatment decisions. The guidelines and checklists for reporting of scientific research are important documents which are meant to increase the transparency of reported research work and hence practically are meant to improve upon the conduct of research itself. Ever since the CONSORT (Consolidated Standards of Reporting Trials) (Moher, 2001) as a set of guidelines for basic reporting of clinical trials, these guidelines have significantly improved and several extensions have been proposed for various other forms of clinical trials. RCTs are considered to be in the top of evidence chain but it's not always ethical to conduct them in every setting. Over these years many other guidelines or checklists have been coming up to reflect the other kinds of research studies like epidemiological (STROBE) (Von et al, 2013), diagnostic (STARD) (Rennie, 2003), meta-analysis (PRISMA) (Moher et al, 2009), statistical methodology (SAMPL) (Lang et al, 2014).

The STROBE (Strengthening of Reporting of Observational Studies in Epidemiology) checklist consists of 22 items that should be included three kinds of analytic epidemiology studies: cross-sectional cohort and case-control studies. The aim of the checklist is to help researchers report the observational studies well. CONSORT (Consolidated Standards of Reporting Trials) statement is a 25 item checklist which consists of recommendations for reporting all kinds of randomized trials. It gives the authors a standard method to prepare reports of randomised trials in a transparent manner which would then lead to a critical evaluation and interpretation.

STARD (Standards for the reporting of Diagnostic Accuracy Studies) consist of a checklist of 25 items which aims to improve the completeness of diagnostic accuracy studies which will allow the readers to assess any biases present in the study and to evaluate how it is generalizable

when compared to other studies. SAMPL (Statistical Analyses and Methods in the Published Literature) is a checklist used for assessing the completeness of basic statistical reporting for articles published in any biomedical journals. Two other checklists and after consulting various other publishing standards as mentioned above. The preliminary checklist was then subjected to discussion rounds and was later validated by three independent statistical researchers to come up with a well-defined checklist.

The study also examined the RCT's (randomized clinical trials) published in the journals via the CONSORT statement. Ten researchers participated in this study: 3 professors, 1 lecturer, and 6 postgraduates. All of them have had formal training in biostatistics and some of them have been in long-term teaching and research roles. Each article used in the study was reviewed by 3 participants independently, which included a professor with two postgraduates. Any discrepancies were noted and later solved.

## **Data analysis**

Since the researcher rigorously reviewed the articles, a simple multi-rate Kappa test after Fleiss was conducted and was found to be ( $k = 0.84$ ) which indicated almost perfect agreement. An a-priori power analyses conducted for 5% level of significance, 80% power for a medium effect size ( $\Phi = 0.3$ ) found a sample size of 82 for each year we selected for our study. For a small effect size ( $\Phi = 0.1$ ) at 5% level of significance with 80% power a sample size of 600 per year was obtained. Hence, we used no sampling procedure in our study but included all the original articles published by the top 10 Indian medical journals in 2003 and 2013. However, a post-hoc power analysis was not conducted since post-hoc power analyses is increasingly considered to be unnecessary. Such analyses are not required at all since if no significance is found, then there wasn't sufficient power to detect the difference.

Epidata3.1 was used for double data entry and management. SPSS 16 and GraphPad were used for subsequent analysis, as appropriate. Chi-square and Fisher's exact tests, (wherever appropriate) were used to compare any proportional differences in use of statistical methods and various errors/defects during the 10 years in the study period. All the tests used in this article were conducted at 5% level of significance and were 2-tailed. Chi-square tests were conducted at one degree of freedom.

## **Results**

### **Statistical Analysis**

As in many articles using statistical methods had increased markedly in ( $\chi^2 = 26.96$ ,  $\Phi = 0.14$ ,  $p < 0.0001$ ) from 42.5% (250/588) to 56.7% (439/774). In 2003, 59.01% (347/588) articles

had no statistical analyses, in which 13.4% (79/588) needed statistical analyses but were omitted. In 2013, 51.03% (395/774) articles had no statistical analyses, in which 14.34% (111/774) needed but statistical analyses were omitted. The most used statistical methods remained the simple tests (i.e., t-tests, contingency tables and ANOVA). However, some more sophisticated statistical methods, such as repeated-measures ANOVA, logistic regression and survival analysis were used in 2013. The proportion of erroneous statistical analyses had not decreased at all ( $\chi^2 = 0.592$ ,  $\Phi = 0.027$ ,  $p = 0.4418$ ), 25% (80/320) in 2003 compared to 22.6% (111/490) in 2013. Statistical methods which had been misused frequently were not only seen in more complex tests but also observed in these simple tests. Compared with 2003, the error proportions of t-test were seen to decrease (t-test: 44.80% (26/58) compared to 24.5% (39/159),  $\chi^2 = 8.346$ ,  $\Phi = 0.21$ ,  $p = 0.0039$ ) and for regression analyses (69.2% (9/13) in 2003 as compared to 33.33% (15/45) in 2013,  $\chi^2 = 5.358$ ,  $\Phi = 0.31$ ,  $p = 0.0206$ ) but not for contingency tables (28.9% (24/83) compared to 27.9% (52/186),  $\chi^2 = 0.026$ ,  $\Phi = 0.0098$ ,  $p = 0.8719$ ). The most common mistakes for these methods were using multiple t-tests for multiple group comparisons, no significant level adjustment for multiple comparisons in association analysis (contingency tables), ignoring or misusing the method of multiple pair-wise comparisons in ANOVA, not reporting regression equation, not mentioning the beta-coefficients and not mentioning whether the basic assumptions were met.

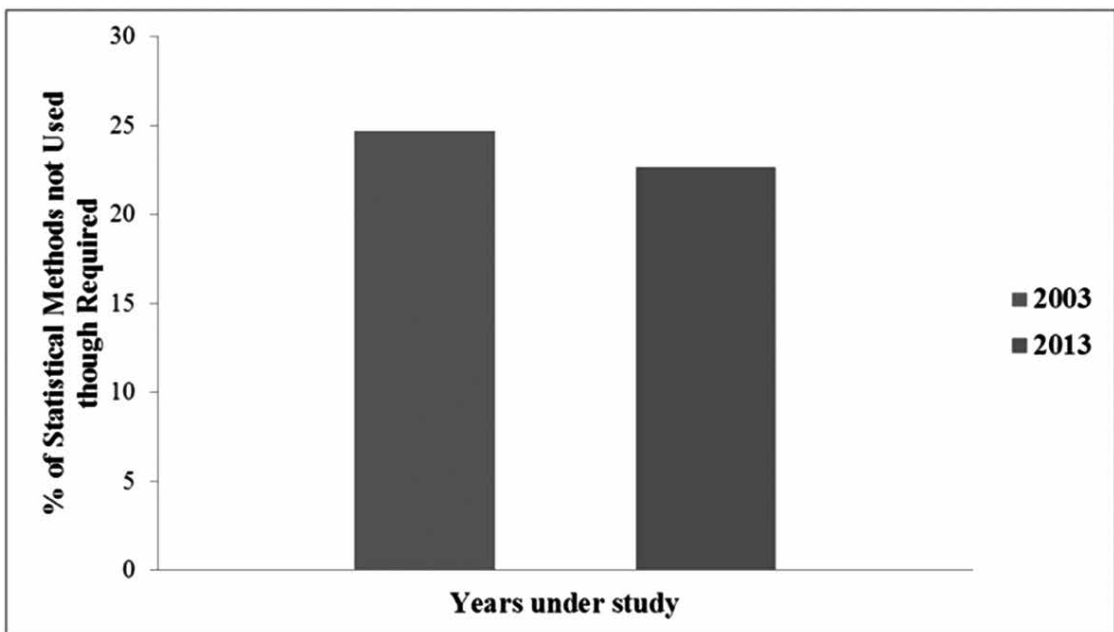


Figure 1. Percentage of statistical methods not used though required in the years 2003 and 2013.

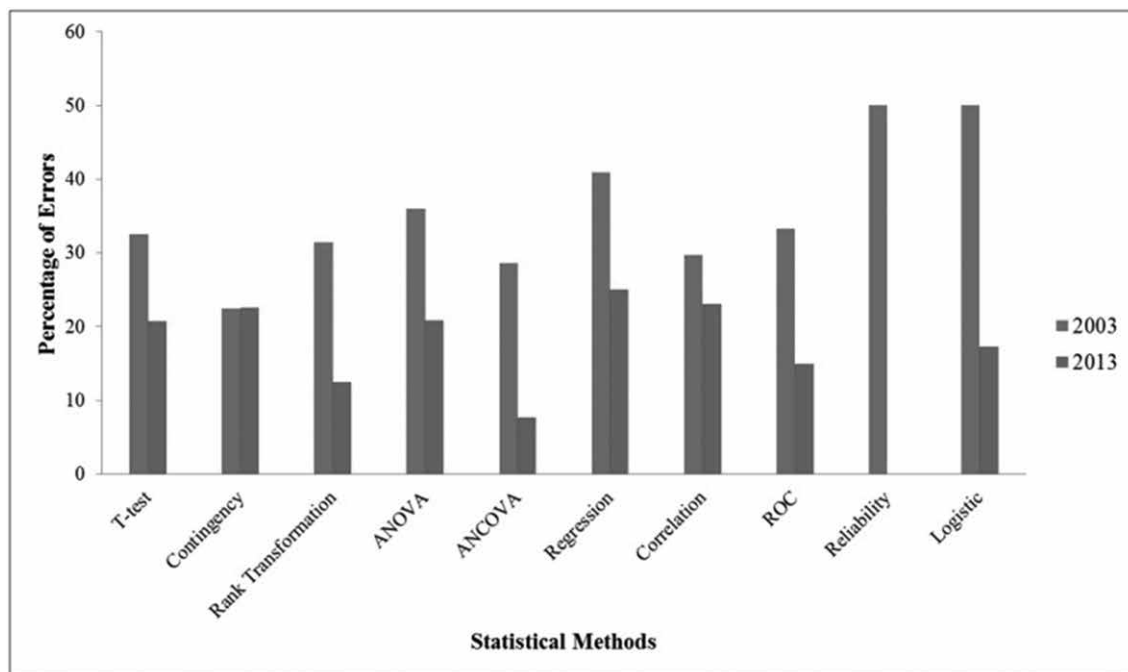


Figure 2. Percentage of errors present in major statistical methods in the years 2003 and 2013.

## Results, interpretation and presentation

Only study design, presentation and interpretation of results were seen to be improved, but serious errors/defects have remained. Inappropriate presentation of statistical results was the most common defect seen in both the observed years. Using arbitrary  $p$  thresholds (like  $p < 0.01$ ) instead of reporting exact  $p$ -values, reporting  $p$  value without test statistics, insufficient (or inappropriate) description of methods and  $p$  values without confidence intervals were most common of the errors present. The overall proportion of inappropriate presentation of results decreased significantly ( $\chi^2 = 24.477$ ,  $\Phi = 0.178$ ,  $p < 0.0001$ ), 82.2% (263/320) in 2003 compared to 66.3% (325/490) in 2013; Using arbitrary  $p$  thresholds instead of reporting exact  $p$  values had not decreased ( $\chi^2 = 1.05$ ,  $\Phi = 0.036$ ,  $p = 0.3055$ ) from 2013 [33.7% (165/490)] as compared to [37.1% (119/320)] in 2003

## Discussion

Currently, RCT's (Randomized clinical trial) is considered as the "gold standard" for evidence synthesis in clinical trials. But surprisingly, even in 2013, after 10 years' of steady progress, randomized clinical trials published in Indian medical journals remained quite low (around 5%).

Despite the publication of CONSORT and the widespread adoption in Western journals the quality of study design and statistical analyses of the RCT's merely improved marginally. We hypothesize many reasons for this. First, Indian medical researchers and clinicians, in general, do not have a solid training in study designs. Very few medical colleges in India have attached statistics departments attached to them. Second, few of the better quality research papers have mostly been published in English in international journals. Along the lines of International Clinical Trials registry, India too established a similar registry for clinical trials. The Clinical Trials Registry—India (CTRI), set up at the National Institute of Medical Statistics, ICMR, New Delhi is a free and online system for registration all clinical trials being conducted in India. Registration of clinical trials in the CTRI is now mandatory, as per notification of the Drugs Controller General (India). The study hence strongly endorses this requirement for full implementation of the clinical trial registration system to promote the quality of randomized clinical trials in India. Although a significant increase was seen in the rates of registration of clinical trials, the proportion of such trials remained low. It is found that in 2013 study design types were unchanged since 2003 Retrospective studies still remain the most published one. Prospective clinical research, which includes randomized clinical trials and non-randomized clinical trials, only accounted for 12.3% in 2013. McDermott et al (2003) in 1991 found that 35.0% papers published in JAMA, The Lancet and New England Journal of Medicine were clinical trials. This is seen in sharp contrast to the Indian condition even after 20 plus years later. Medical researchers should take the advantage of India's large population, heavy cash investment in research by Indian Council of Medical Research (ICMR), Department of Biotechnology (DBT) and other organizations like Public Health Foundation of India (PHFI), broad disease spectrum found here (both lifestyle based and infectious), and quite a low cost to conduct high quality randomized clinical trials. Health policy-makers should encourage medical research by conducting more randomized clinical trials with relevant guidance to the researchers to pay more attention to quality than quantity of their publications.

It is noticed a massive increase in the use of rank based nonparametric test in 2013, which indicates that more attention are being paid to the assumptions of parametric test. This progress is mainly attributable to the emphasis of statistical education among medical postgraduates in India. Serious problems have remained nonetheless. Simple methods like t-tests, contingency tables, and ANOVA are seen to be used incorrectly. Regarding presentation of statistical results, even prestigious journals like Nature and BMJ were found to have a defect proportion of 38.0% and 25.0%, respectively. Using arbitrary  $p$  thresholds (like  $p < 0.01$ ) instead of reporting exact  $p$ -values, reporting  $p$  value without test statistics, insufficient (or inappropriate) description of methods and  $p$  values without confidence intervals were most common of the error/defects observed in this study. Precise  $p$ -value and test statistic should be provided at the same time. We emphasize that, medical colleges should start teaching of the basic statistical concepts in a major way and hence strengthen statistical understanding among medical students. Also, in hospitals, a continuous education program in biostatistics should be encouraged among clinicians. According to the international practices followed by prestigious journals like JAMA, BMJ etc., editorial

board should have qualified statisticians who should have a major say in checking the statistical quality of published papers.

The journals selected in this present study covered the important clinical fields and represented the top academic level of India based on impact factors released by Thomson Reuters in 2013. It should be noted, however, that high quality Indian research is regularly published in high-level international (e.g. American and European) journals. But in this study, we have not included such papers, which is a potential limitation. Also, another major limitation of this study is that since it's primarily based on Indian medical journals hence, no generalization is possible with respect to other external, international journals.

In summary, this study indicates that Indian medical research seems to have made some progress regarding study design defects, but there remains ample room for improvement regarding statistical analyses. Most published studies continue to use a retrospective clinical design, with randomized clinical trials being quite rare. Those RCTs that are published often have serious methodological problems, including absence of sample size estimation and power calculations, as well as failure in (or in the reporting of) randomization. Therefore, it is urgently recommended the full implementation of the CONSORT statement and of the registration system for clinical trials. It seems that medical research published in Indian journals still has much room for improvement, when compared to Western journals, not only in clinical knowledge, but also in study design and statistical methodology.

## References

- Altman DG (1981) Statistics and ethics in medical research. VIII-Improving the quality of statistics in medical journals. *Br Med J (Clin Res Ed)* 282: 44–47. [PMC free article] [PubMed]
- Altman, DG, Gore SM, Gardner MJ, Pocock SJ (1983) Statistical guidelines for contributors to medical journals. *Br Med J (Clin Res Ed)* 286: 1489–1493. [PMC free article] [PubMed]
- Avram, MJ, Shanks CA, Dykes MHM, Ronai AK, Stiers WM (1985) Statistical methods in anesthesia articles: an evaluation of two American journals during two six-month periods. *Anesth Analg* 64: 607–611.[PubMed]
- Bailar, JC, Mosteller F. Guidelines for statistical reporting in articles for medical journal: amplifications and explanations. *AnnInternMed.* 1988; 108:266–273. [PubMed]
- Chan, AW, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. (2013) SPIRIT 2013 Statement: Defining standard protocol items for clinical trials. *Ann Intern Med.* 2013;158(3):200–207 doi: 10.7326/0003-4819-158-3-201302050-00583 [PMC free article] [PubMed]
- Des Jarlais, DC, Lyles C, Crepaz N (2004) Improving the reporting quality of nonrandomized



- evaluations of behavioral and public health interventions: the TREND statement. *Am J Pub Health* 94: 361–366. [PMC free article] [PubMed]
- Emerson, JD, Colditz GA (1983) Use of statistical analysis in the *New England Journal of Medicine*. *N Engl J Med* 309:709–713. [PubMed]
- Feinstein, AR (1974) Clinical biostatistics, XXV: survey of statistical procedures in general medical journals. *Clin Pharmacol Ther* 15:97–107. [PubMed]
- Fletcher, RH, Fletcher SW (1979) Clinical research in general medical journals: a 30-year perspective. *N Engl J Med* 301:180–183. [PubMed]
- Goyal, J., Jaykaran, Kantharia ND, Preeti Y, Bharddwaj P (2010) Reporting statistics in clinical trials published in Indian journals: a survey. *African Health Sciences* 10(2): 204–207 [PMC free article][PubMed]
- Horton, NJ, Switzer SS (2005) Statistical methods in the journal. *N Engl J Med*. 353: 1977–1979[PubMed]
- Juzych, MS, Shin DH, Seyedsadr M, Siegner SW, Juzych LA (1992) Statistical techniques in ophthalmic journals. *Arch Ophthalmol* 110:1225–1229. [PubMed]
- Kanter, MH, Taylor JR (1994) Accuracy of statistical methods in transfusion: a review of articles from July/August 1992 through June 1993. *Transfusion* 34:697–701. [PubMed]
- Lang, TA, Altman DG (2014) Statistical Analyses and Methods in the Published Literature: The SAMPL Guidelines\*, in *Guidelines for Reporting Health Research: A User's Manual* (eds D Moher, D. G. Altman, K. F Schulz, I Simera and E Wager), John Wiley & Sons, Ltd, Oxford, UK: doi: 10.1002/9781118715598.ch25
- Moher, D, Liberati A, Tetzlaff J, Altman DG (2009) The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 6(7):e1000097 doi: 10.1371/journal.pmed.1000097 [PMC free article] [PubMed]
- Moher, D, Schulz KF, Altman D (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 285: 1987–1991. [PubMed]
- Morris, RW (1988) A statistical study of papers in the journal of bone and joint surgery (BR) 1984. *J Bone Joint Surg Br* 70:242–246. [PubMed]
- Rennie, D (2003) Improving reports of studies of diagnostic tests: the STARD initiative. *JAMA* 289: 89–90. [PubMed]
- Rosenfeld, RM, Rockette HE (1991) Biostatistics in otolaryngology journals. *Arch Otolaryngol Head Neck Surg* 117:1172–1176. [PubMed]
- Schwartz SJ, Sturr M, Golberg G (1996) Statistical methods in rehabilitation literature: a survey of

recent publications. *Arch Phys Med Rehabil* 77:497–500. [PubMed]

Von, Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, et al. (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 370: 1453–1457. [PubMed]

Wainapel SF, Kayne HL (1985) Statistical methods in rehabilitation research. *Arch Phys Med Rehabil* 66:332–334. [PubMed]