

Misuse of Statistical Methods in 10 Leading Chinese Medical Journals in 1998 and 2008

Shunquan Wu,¹ Zhichao Jin,¹ Xin Wei,² Qingbin Gao,¹ Jian Lu,¹ Xiuqiang Ma,¹
Cheng Wu,¹ Qian He,¹ Meijing Wu,¹ Rui Wang,¹ Jinfang Xu,¹ and Jia He¹

¹Department of Health Statistics, Second Military Medical University,
Shanghai 200433, China

²Clinical Medicine, College of Basic Medicine, Shanghai Jiaotong University School of
Medicine, Shanghai 200025, China

Received 17 May 2011; Accepted 14 October 2011

Academic Editor: Graham D. Dunn

Statistical methods are vital to biomedical research. Our aim was to find out whether progress has been made in the last decade in the use of statistical methods in Chinese medical research. We reviewed 10 leading Chinese medical journals published in 1998 and in 2008. Regarding statistical methods, using a multiple t -test for multiple group comparison was the most common error in the t -test in both years, which significantly decreased in 2008. In contingency tables, no significant level adjustment for multiple comparison significantly decreased in 2008. In ANOVA, over a quarter of articles misused the method of multiple pair-wise comparison in both years, and no significant difference was seen between the two years. In the rank transformation nonparametric test, the error of using multiple pair-wise comparison for multiple group comparison became less common. Many mistakes were found in the randomised controlled trial (56.3% in 1998; 67.9% in 2008), non-randomised clinical trial (57.3%; 58.6%), basic science study (72.9%; 65.5%), case study or case series study (48.4%; 47.2%), and cross-sectional study (57.1%; 44.2%). Progress has been made in the use of statistical methods in Chinese medical journals, but much is yet to be done.

KEYWORDS: statistical methods, misuse, errors, medical journals

1. INTRODUCTION

Statistics play a key role in biomedical research [1–6]; their correct use is thus essential to a high-quality study. The misuse or inaccurate use of statistical methods may point the research in the wrong direction and produce incorrect study results.

China produces a large number of biomedical articles. According to the database of the Institute for Scientific Information (ISI), there has been a significant increase in the quantity and quality of Chinese biomedical publications in the last two decades, especially in the last decade [7]. However, it is common to find inappropriate statistical methods in Chinese medical journals. He et al. reported in 2009 that many more statistical errors existed in Chinese medical journals than in international journals [8].

Our previous study compared the research design, statistical analyses, and presentation and interpretation of results of 10 leading Chinese medical journals published in 1998 and 2008 in Chinese [9]. The main results we obtained were the frequencies of different types of study design, defective proportions in design and statistical analyses, and the inappropriate presentation and interpretation of results. Further, we mentioned that the most frequently used statistical methods were still the simple tests, although more sophisticated statistical methods were already being applied in 2008. As for the study design, our focus was primarily on retrospective studies, with clinical trials receiving relatively little attention.

In this research, we again used the 10 leading Chinese medical journals published in 1998 and 2008 and extracted new data on the misuse and inaccuracy of each statistical method as an extension. We listed and compared the most common errors of each method that appeared in medical articles in 1998 and 2008. We also compared the proportions of the incorrect use of statistical methods in different study designs between the two years; in our previous study [9], we had compared the proportions of design defects in various study designs. All statistical procedures and methods in each article were reviewed, and trends in the misuse of statistical methods were reported. We summarised the progress that had been made during the past 10 years and discussed the current concerns about Chinese medical journals. We analysed the possible reasons of the main errors and suggested some improvements on the quality of Chinese medical journals.

2. METHODS

2.1. The 10 Chinese Medical Journals Used

The 10 leading Chinese medical journals that we examined were *Chinese Journal of Internal Medicine*, *Chinese Journal of Surgery*, *Chinese Journal of Pediatrics*, *Chinese Journal of Obstetrics and Gynecology*, *Chinese Journal of Ophthalmology*, *Chinese Journal of Hematology*, *Chinese Journal of Stomatology*, *Chinese Journal of Cardiology*, *Chinese Journal of Oncology*, and *Chinese Journal of Tuberculosis and Respiratory Diseases*. These journals were peer-reviewed, sponsored by the Chinese Medical Association, and indexed by Medline; however, they were only published in Chinese. All the original articles published—1,335 in 1998 and 1,578 in 2008—were reviewed [9].

2.2. Data Collection and Processing

We recorded four types of the most commonly used statistical methods in medical journals: *t*-test, contingency tables, analysis of variance (ANOVA), and rank transformation nonparametric test. Some rarely used methods, such as mixed/multilevel model, analysis of covariance, and survival analysis were not analysed in this study. We listed the main errors that appeared in the four types of methods. The frequencies and percentages were adopted to describe the occurrences of misuse of statistical methods. The quantities and proportions of the incorrect use of statistical methods in the different study designs were estimated. The details of quality control have been described in our previous study [9] SAS 9.1.3 was used, and a Chi-square test was conducted to assess the trends. A *P* value of less than 0.05 was considered to be statistically significant. The odds ratio (OR) and 95% confidence interval (CI) were also estimated.

3. RESULTS

3.1. Errors in the Different Statistical Methods

As described in our previous study, 492 and 570 articles used the *t*-test in 1998 and 2008; 319 and 523 used contingency tables; 202 and 446 used ANOVA; 67 and 187 used the rank transformation nonparametric test [9]. The specific errors that occurred in the four types of statistical methods of both years are listed in Table 1.

The proportion of the misuse of the *t*-test decreased significantly in 2008. The prevalence of using multiple *t*-tests for multiple group comparisons (using multiple *t*-tests to compare the means of more than two groups) was particularly disconcerting: 31.1% (153/492) in 1998 and 22.6% (129/570) in 2008 ($\chi^2 = 9.71$, $P = 0.002$, OR = 1.54, 95% CI: 1.17 to 2.03). Two other major errors were the use of the *t*-test under a nonparametric setting (from 18.1% (89/492) in 1998 to 10.5% (60/570) in 2008; $\chi^2 = 12.52$, $P < 0.001$, OR = 1.88, 95% CI: 1.32 to 2.67) and the use of the *t*-test to conduct repeated-measure data analysis (from 14.8% (73/492) in 1998 to 10.5% (60/570) in 2008; $\chi^2 = 4.48$, $P = 0.034$, OR = 1.48, 95% CI: 1.03 to 2.13).

A declining trend was also observed in the inaccurate use of contingency tables. The two most common mistakes were no significant level adjustment for multiple comparison (from 25.7% (82/319) in 1998 to 14.2% (74/523) in 2008; $\chi^2 = 17.53$, $P < 0.001$, OR = 2.10, 95% CI: 1.48 to 2.98) and no continuity correction or Fisher exact test if needed (from 16.3% (52/319) in 1998 and to 10.1% (53/523) in 2008; $\chi^2 = 6.90$, $P = 0.009$, OR = 1.73, 95% CI: 1.15 to 2.61). The third most common mistake was using the Chi-square test for ranked data, where no decline was seen ($\chi^2 = 3.00$, $P = 0.083$, OR = 1.59, 95% CI: 0.94 to 2.69).

Unfortunately, the incorrect use of ANOVA remained high in 2008. Over a quarter of the articles misused the multiple pair-wise comparison of post hoc ANOVA in both years ($\chi^2 = 1.30$, $P = 0.225$, OR = 0.80, 95% CI: 0.55 to 1.17). However, two errors decreased in 2008: using ANOVA to analyse repeated-measures data and not using a multiple pair-wise comparison of ANOVA when needed; both reached a level of statistical significance ($\chi^2 = 6.65$, $P = 0.010$, OR = 1.74, 95% CI: 1.14 to 2.67 and $\chi^2 = 6.89$, $P = 0.009$, OR = 2.11, 95% CI: 1.20 to 3.72, resp.).

There were two main errors in the rank transformation nonparametric test. One was the use of multiple pair-wise comparison for multiple groups ($\chi^2 = 4.43$, $P = 0.035$, OR = 2.21, 95% CI: 1.04 to 4.47), although fewer errors of the sort were found in 2008. The other one, wherein the wrong type of rank sum test was used for different study types, did not show a significant difference ($\chi^2 = 2.07$, $P = 0.150$, OR = 3.89, 95% CI: 0.85 to 17.88).

3.2. Misuse of Statistical Methods in Different Study Designs

Despite the significant growth in use of statistical methods, substantive errors still existed in different study designs. Table 2 shows the quantities and proportions of the statistical methods used and the errors that were found. Errors mainly occurred in clinical trials, basic science study, and retrospective study.

In the clinical trials, over half of the articles with statistical methods had mistakes in both years. No statistical significance was seen in the clinical trials during the last 10 years for randomised controlled trials ($\chi^2 = 1.70$, $P = 0.192$, OR = 0.61, 95% CI: 0.29 to 1.29) and nonrandomised clinical trials ($\chi^2 = 0.02$, $P = 0.878$, OR = 0.95, 95% CI: 0.48 to 1.87). A mass of statistical errors existed in basic science study, which was used frequently in both years. The proportions of errors were 72.9% (175/240) in 1998 and 65.5% (268/409) in 2008 ($\chi^2 = 3.81$, $P = 0.051$, OR = 1.42, 95% CI: 1.00 to 2.01). The situation was equally worrisome in retrospective study, case-control study, and case study or case-series study. Although a downward trend in mistakes was seen in case-control study ($\chi^2 = 7.05$, $P = 0.008$, OR = 1.59, 95% CI: 1.13 to 2.24), there was no significant improvement in case study or case-series study ($\chi^2 = 0.04$, $P = 0.837$, OR = 1.05, 95% CI: 0.68 to 1.62). It was gratifying to see a significant drop in the proportion

TABLE 1: Errors/defects in statistical methods.

Types of errors	Incorrect use in 1998 <i>n</i> (%)	Incorrect use in 2008 <i>n</i> (%)	χ^2	<i>P</i> value	OR* (1998/2008)	95% CI**
<i>t</i> -test	305 (62.0%)	253 (44.4%)	32.83	<0.001	2.04	(1.60,2.61)
(1) Using multiple <i>t</i> -test for multiple group comparison	153 (31.1%)	129 (22.6)	9.71	0.002	1.54	(1.17,2.03)
(2) Using paired <i>t</i> -test for unpaired data or vice versa	40 (8.1%)	34 (6.0%)	1.91	0.167	1.40	(0.87,2.24)
(3) Using <i>t</i> -test under nonparametric setting	89 (18.1%)	60 (10.5%)	12.52	<0.001	1.88	(1.32,2.67)
(4) Using <i>t</i> -test without considering the baseline	52 (10.6%)	33 (5.8%)	8.19	0.004	1.92	(1.22,3.03)
(5) Using <i>t</i> -test to conduct repeated-measure data	73 (14.8%)	60 (10.5%)	4.48	0.034	1.48	(1.03,2.13)
Others	28 (5.7%)	8 (1.4%)	14.82	<0.001	4.24	(1.91,9.39)
Contingency tables	154 (48.3%)	169 (32.3%)	21.35	<0.001	1.96	(1.47,2.60)
(1) No continuity correction or Fisher exact test if needed	52 (16.3%)	53 (10.1%)	6.90	0.009	1.73	(1.15,2.61)
(2) No significant level adjustment for multiple comparison	82 (25.7%)	74 (14.2%)	17.53	<0.001	2.10	(1.48,2.98)
(3) Misusing Chi-square test for paired fourfold table	10 (3.1%)	12 (2.3%)	0.55	0.458	1.38	(0.59,3.23)
(4) Using Chi-square test for ranked data	29 (9.1%)	31 (5.9%)	3.00	0.083	1.59	(0.94,2.69)
(5) Ignorance of stratification factors	12 (3.8%)	12 (2.3%)	1.54	0.215	1.66	(0.74,3.75)
(6) Using <i>P</i> value of Chi-square test instead of contingency coefficient to describe the correlation of two variables	8 (2.5%)	4 (0.8%)	3.13	0.077	3.34	(0.98,11.18)
Others	21 (6.6%)	19 (3.6%)	3.81	0.051	1.87	(0.99,3.53)
ANOVA***	128 (63.4%)	263 (59.0%)	1.12	0.289	1.20	(0.85,1.70)
(1) Using one-factorial ANOVA to analyse data from multifactorial designs	10 (5.0%)	31 (7.0%)	0.94	0.333	0.70	(0.34,1.45)
(2) Ignoring the setting of ANOVA for completely random design data	25 (12.4%)	53 (11.9%)	0.03	0.858	1.05	(0.63,1.74)
(3) No multiple pair-wise comparison of ANOVA when needed	25 (12.4%)	28 (6.3%)	6.89	0.009	2.11	(1.20,3.72)
(4) Misusing the method of multiple pair-wise comparison of ANOVA	51 (25.3%)	132 (29.6%)	1.30	0.255	0.80	(0.55,1.17)

TABLE 1: Continued.

Types of errors	Incorrect use in 1998 <i>n</i> (%)	Incorrect use in 2008 <i>n</i> (%)	χ^2	<i>P</i> value	OR* (1998/2008)	95% CI**
(5) Using ANOVA to analyse repeated-measures data	45 (22.3%)	63 (14.1%)	6.65	0.010	1.74	(1.14,2.67)
Others	16 (7.9%)	10 (2.2%)	11.64	0.001	3.75	(1.67,8.42)
Rank transformation nonparametric test	29 (43.3%)	33 (17.7%)	17.57	<0.001	3.56	(1.93,6.57)
(1) Using multiple pair-wise comparison for multiple group comparison	14 (20.9%)	20 (10.7%)	4.43	0.035	2.21	(1.04,4.47)
(2) Using wrong type of rank sum test for different study types	4 (6.0%)	3 (1.6%)	2.07	0.150	3.89	(0.85,17.88)
Others	20 (29.9%)	6 (3.2%)	38.11	<0.001	12.84	(4.88,33.77)

*OR: odds ratio; **CI: confidence interval; ***ANOVA: analysis of variance.

Incorrect use of *n* (%): for each statistical method, *n* is the number of articles using this statistical methods incorrectly and the percentage = *n*/the number of papers using this statistical methods \times 100%; for each error under certain statistical methods, *n* is the number of articles with this mistake and the percentage = *n*/ the number of papers using these statistical methods \times 100%.

TABLE 2: Statistical methods under different study designs.

Types of study design	1998		2008	
	Articles that used statistical methods <i>n</i> (%)	Incorrect use of statistical methods <i>n</i> (%)	Articles that used statistical methods <i>n</i> (%)	Incorrect use of statistical methods <i>n</i> (%)
Systematic review	0	0	6 (100.0%)	0 (0%)
RCT*	64 (98.0%)	36 (56.3%)	56 (93.3%)	38 (67.9%)
Non-RCT**	82 (91.1%)	47 (57.3%)	58 (95.1%)	34 (58.6%)
Cohort study	47 (79.7%)	28 (59.6%)	80 (92.0%)	17 (21.3%)
Case-control study	254 (92.4%)	148 (58.3%)	276 (97.2%)	129 (46.7%)
Cross-sectional study	56 (74.7%)	32 (57.1%)	52 (88.1%)	23 (44.2%)
Case study or case series study	122 (31.9%)	59 (48.4%)	233 (48.9%)	110 (47.2%)
Diagnostic test	47 (74.6%)	20 (42.6%)	63 (82.9%)	25(39.7%)
Basic science study	240 (74.1%)	175 (72.9%)	409 (87.4%)	268 (65.5%)
Total	912 (68.3%)	545 (59.8%)	1233 (78.1%)	644 (52.2%)

* RCT: randomised controlled trial.

**Non-RCT: nonrandomised clinical trial.

N = total articles (1998:1335; 2008:1578).

All articles *n* (%): *n* is the number of articles of each type of study design and percentage = *n*/*N* \times 100%.

Articles that used statistical methods *n* (%): *n* is the number of articles using statistical methods in each type of study design and the percentage = *n*/the number of articles of each type of study design, incorrect use of statistical methods *n* (%): *n* is the number of articles using statistical methods incorrectly and percentage = *n*/the number of articles using statistical methods in each type of study design.

of errors in cohort study ($\chi^2 = 19.01$, $P < 0.001$, OR = 5.46, 95% CI: 2.48 to 12.05), but no improvement was observed in cross-sectional study ($\chi^2 = 1.80$, $P = 0.180$, OR = 1.68, 95% CI: 0.79 to 3.60).

4. DISCUSSION

4.1. Possible Reasons for the Occurrence of Errors

Among the errors, the biggest problem was the inappropriate choice of statistical methods. The possible reason for this was that not much attention was paid to the distributional characteristics of the variables and the nature of the data. Apparently, due to the researchers' lack of basic knowledge of statistics, they ignored the application condition of a certain method. When the quantitative data did not meet the prerequisites for parametric tests, they blindly applied the tests. Many researchers mistakenly believed that the Chi-square test was a universal tool for dealing with contingency tables, and they used it on data without taking the data characteristics into consideration. Some multifactorial experimental studies were split into a series of single-factor studies, which disassembled the intrinsic link or interactions among factors and led to one-sided or even wrong conclusions. Park et al. stated that the selection of the correct statistical method depends on the data structure and underlying statistical assumptions [10]. However, some errors were very common among articles, and they were wrongly cited or used by others, resulting in a vicious circle. As Altman DG said, "once incorrect procedures become common, it can be hard to stop them from spreading through the medical literature like a genetic mutation" [11].

4.2. Correct Methods Should Be Used in These Situations

Regarding the *t*-test, the most frequent error was using multiple *t*-tests for multiple group comparison, which may increase the probability of making a Type 1 error. There are several methods for multiple comparison, such as the Bonferroni method, Scheffé method, Tukey method, Newman-Keuls method, and Duncan method [12]. Around 4.95% and 6.95% of the articles which used ANOVA in 1998 and 2008 employed one-factorial ANOVA to analyse data from multifactorial designs. One-factorial ANOVA is used when there is only one experimental factor; when two or more experimental factors are involved, multifactorial ANOVA should be used [13, 14]. The *t*-test and standard ANOVA require independent data that have no correlation with each other. Repeated-measure data do not meet this requirement; instead, repeated-measures ANOVAs or mixed-effects models should be used. Mixed-effects models are recommended, as they have greater flexibility to model time effects and can handle missing data more appropriately [15]. A common error encountered in contingency tables in both years was that there did not exist continuity correction or Fisher exact test if needed. It is considered incorrect to use the Chi-square test directly in contingency table analysis if the total sample size is not more than 20, or if more than 20% of the expected frequencies are less than five; Fisher's exact test should be applied in both cases [16]. Nonparametric tests are often used in place of parametric tests when the assumptions of the parametric test have been grossly violated (e.g., if the distributions are too severely skewed.) Nonparametric tests are also recommended for small sample sizes or data sets with many ties. The error proportions of using the Chi-square test for ranked data were 9.09% and 5.93% in the two years. Instead of a Chi-square test, a rank transformation nonparametric test should be used on ranked data. For study designs, the more complex the study design was, the more mistakes in statistical methods were likely to appear.

4.3. Progress and Worries

In general, progress has been made in the statistical methods of Chinese medical journals in the last decade. The percentage of articles using statistical methods has increased, and the proportion of errors has significantly decreased in most of the statistical methods and study designs. This conclusion was consistent with what Wang et al. reported in 1998 that the proportion of papers in Chinese medical journals using

appropriate statistical methods had increased in 1995 compared with 1985 [17]. From this point, we can see that Chinese medical researchers have made great efforts to employ statistical methods in their studies. However, we cannot be overoptimistic because the situation is very far from satisfactory. Although statistical errors also exist in the medical journals of western countries, the proportion is smaller. McGuigan reviewed all papers published in the *British Journal of Psychiatry* in 1993 and found that 40% of the papers contained statistical errors [18]. Welch and Gabbe reviewed 145 clinical articles published in *American Journal of Obstetrics and Gynecology* in 1994 and pointed out 46 articles (31.7%) that were deemed to have applied statistics inappropriately [19]. Kurichi and Sonnad reported that only 27% of the studies in five selected surgical journals of America in 2003 included incorrect selection or reporting of statistical methods [20]. Another study, conducted by Neville et al., assessed the frequency of statistical errors in dermatological literature. The study revealed that only 14% of the articles with statistical analysis contained errors in the methods; 26.5%, in the presentation of the results; 2.6%, in both [21].

In China, the situation is quite depressing, as the error rate of statistical methods remains high. In ANOVA, the total error rate hit approximately 60% in 1998 and 2008. Many mistakes were made even in the most basic aspects. For instance, 31.10% of articles used *t*-test for multiple group comparison in 1998 and 22.63% in 2008. In clinical trials, over half of the articles had statistical errors. The proportion of errors was extremely high in basic science study and retrospective study, even if these were frequently used.

In addition, many sophisticated statistical methods, such as analysis of covariance, repeated-measures analysis, logistic regression, and survival analysis were seldom used in Chinese medical journals—an observation that was also made by Wang and Zhang in their study [17]. This suggests that a large amount of data is not being efficiently analysed, so that much of the information is wasted. Considering the high incidence of errors in the simple statistical methods, it is not hard to imagine how bad the situation is with regard to sophisticated statistical methods. Moreover, since we studied only the 10 leading medical journals, it is likely that our results were above the actual average of Chinese medical journals. It must be noted, though, that some Chinese research papers published in international medical journals, whose statistical methodology might be of better quality, were not included in this study. Thus, as our next step, we intend to conduct a survey on Chinese clinical studies that have been published in international journals.

5. RECOMMENDATIONS

The 10 medical journals we selected are representative of excellent Chinese medical journals. Nevertheless, there is still a wide gap between them and the international journals with respect to statistics. Some measures are needed to decrease the errors in the statistical methods and improve the quality of articles.

Firstly, clinicians and medical researchers should correct their attitude about writing. Their purpose of publication should be to make their results known to their colleagues and raise the level of medical science of mankind; it should not be personal aggrandisement. Only correct medical outcomes can benefit people. Secondly, statistical education should be enhanced among clinicians and researchers; they should have a basic concept of statistics and study design. An integrated and detailed protocol should be made beforehand. And when performing and analyzing RCTs, CONSORT statement is recommended as a guideline, which is accepted internationally now. Thirdly, statisticians should assume an important role in the research; in other words, a research group should include a statistician as a consultant. Finally, statistical reviewers should be included in the editorial boards of the journals. Some journals merely intend to make profits through page charges, publishing random articles without taking quality into account. Measures must be implemented to prevent such practices.

ACKNOWLEDGMENTS

This study was supported by the “Strategic Priority Research Program” of the Chinese Academy of Sciences Grant no. XDA01040100, the Ministry of Science and Technology of China (2008ZX09312-007,

2009ZX09312-025, 2008ZX10002-018). The authors thank the following members for data collection and group discussion: Xiaofei Ye, Xiaojing Guo, Xinji Zhang, and Jiajie Zang. Shunquan Wu and Zhichao Jin contributed equally to this study.

REFERENCES

- [1] S. J. Schwartz, M. Sturr, and G. Goldberg, “Statistical methods in rehabilitation literature: a survey of recent publications,” *Archives of Physical Medicine and Rehabilitation*, vol. 77, no. 5, pp. 497–500, 1996.
- [2] C. D. Scales Jr., R. D. Norris, G. M. Preminger, J. Vieweg, B. L. Peterson, and P. Dahm, “Evaluating the evidence: statistical methods in randomized controlled trials in the urological literature,” *Journal of Urology*, vol. 180, no. 4, pp. 1463–1467, 2008.
- [3] S. V. Law, D. N. Chudasama, and D. J. Rinchuse, “Evidence-based orthodontics: current statistical trends in published articles in one Journal,” *Angle Orthodontist*, vol. 80, no. 5, pp. 952–956, 2010.
- [4] W. Huang, J. M. LaBerge, Y. Lu, and D. V. Glidden, “Research publications in vascular and interventional radiology: research topics, study designs, and statistical methods,” *Journal of Vascular and Interventional Radiology*, vol. 13, no. 3, pp. 247–255, 2002.
- [5] J. Goldin, W. Zhu, and J. W. Sayre, “A review of the statistical analysis used in papers published in clinical radiology and british journal of radiology,” *Clinical Radiology*, vol. 51, no. 1, pp. 47–50, 1996.
- [6] S. Al-Benna, Y. Al-Ajam, B. Way, and L. Steinstraesser, “Descriptive and inferential statistical methods used in burns research,” *Burns*, vol. 36, no. 3, pp. 343–346, 2010.
- [7] G. C. Makris, A. Spanos, P. I. Rafailidis, and M. E. Falagas, “Increasing contribution of China in modern biomedical research. Statistical data from ISI Web of Knowledge,” *Medical Science Monitor*, vol. 15, no. 12, pp. SR15–SR21, 2009.
- [8] J. He, Z. Jin, and D. Yu, “Statistical reporting in Chinese biomedical journals,” *The Lancet*, vol. 373, no. 9681, pp. 2091–2093, 2009.
- [9] Z. Jin, D. Yu, L. Zhang et al., “A retrospective survey of research design and statistical analyses in selected chinese medical journals in 1998 and 2008,” *PLoS One*, vol. 5, no. 5, Article ID e10822, 2010.
- [10] E. Park, M. Cho, and C. S. Ki, “Correct use of repeated measures analysis of variance,” *Korean Journal of Laboratory Medicine*, vol. 29, no. 1, pp. 1–9, 2009.
- [11] D. G. Altman, “Poor-quality medical research: what can journals do?” *Journal of the American Medical Association*, vol. 287, no. 21, pp. 2765–2767, 2002.
- [12] L. S. Kao and C. E. Green, “Analysis of variance: is there a difference in means and what does it mean?” *Journal of Surgical Research*, vol. 144, no. 1, pp. 158–170, 2008.
- [13] M. L. Gaddis, “Statistical methodology: IV. Analysis of variance, analysis of covariance, and multivariate analysis of variance,” *Academic Emergency Medicine*, vol. 5, no. 3, pp. 258–265, 1998.
- [14] K. J. Ottenbacher, “Interpretation of interaction in factorial analysis of variance design,” *Statistics in Medicine*, vol. 10, no. 10, pp. 1565–1571, 1991.
- [15] R. Gueorguieva and J. H. Krystal, “Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry,” *Archives of General Psychiatry*, vol. 61, no. 3, pp. 310–317, 2004.
- [16] K. H. Yim, F. S. Nahm, K. A. Han, and S. Y. Park, “Analysis of statistical methods and errors in the articles published in the Korean Journal of Pain,” *Korean Journal of Pain*, vol. 23, pp. 35–41, 2010.
- [17] Q. Wang and B. Zhang, “Research design and statistical methods in Chinese medical journals,” *Journal of the American Medical Association*, vol. 280, no. 3, pp. 283–285, 1998.
- [18] S. M. McGuigan, “The use of statistics in the British Journal of Psychiatry,” *British Journal of Psychiatry*, vol. 167, pp. 683–688, 1995.
- [19] G. E. Welch II and S. G. Gabbe, “Review of statistics usage in the American Journal of Obstetrics and Gynecology,” *American Journal of Obstetrics and Gynecology*, vol. 175, no. 5, pp. 1138–1141, 1996.
- [20] J. E. Kurichi and S. S. Sonnad, “Statistical methods in the surgical literature,” *Journal of the American College of Surgeons*, vol. 202, no. 3, pp. 476–484, 2006.

- [21] J. A. Neville, W. Lang, and A. B. Fleischer Jr., “Errors in the Archives of Dermatology and the Journal of the American Academy of Dermatology from January Through December 2003,” *Archives of Dermatology*, vol. 142, no. 6, pp. 737–740, 2006.

This article should be cited as follows:

Shunquan Wu, Zhichao Jin, Xin Wei, Qingbin Gao, Jian Lu, Xiuqiang Ma, Cheng Wu, Qian He, Meijing Wu, Rui Wang, Jinfang Xu, and Jia He, “Misuse of Statistical Methods in 10 Leading Chinese Medical Journals in 1998 and 2008,” *TheScientificWorldJOURNAL*, vol. 11, pp. 2106–2114, 2011.
