Research Design and Statistical Methods in Chinese Medical Journals

Qian Wang, MS; Boheng Zhang, MS

Context.—Study design and statistical analyses have improved in journals published in Western countries, but the type of research designs and statistical methods used in medical journals outside Western countries has not been assessed.

Objectives.—To determine the frequency of research designs and statistical techniques used in Chinese medical journals, types of statistical errors present, and trends over a 10-year period.

Design.—Evaluation of all original articles published in 5 leading journals in 1985 (N = 640) and in 1995 (N = 954).

Main Outcome Measures.—Research designs and statistical methods.

Results.—Compared with 1985, significant improvement was seen in 1995: the percentages of original articles reporting clinical trials, prospective studies, or basic science research increased from 18% to 31% (P < .001), the proportion of papers using statistical tests increased from 40% to 60% (P < .001), and of those articles using statistics, the proportion using appropriate methods increased from 22% to 46% (P < .001). In both years, the most commonly used statistical methods were t tests and contingency tables. The most common errors were presentation of P values without specifying the test used, use of multiple t tests instead of analysis of variance, and use of unpaired t tests when paired tests were required.

Conclusions.—The use of statistical methods in Chinese medical journals research is improving, and by 1995, the frequency of using statistical methods in published articles was similar to the results determined in previous studies of journals. However, the lack of or inappropriate use of statistics remains a serious problem.

STANDARDIZATION OF STATISTICAL METHODS used in biomedical research articles have been increasingly scrutinized, especially in Western journals.1-10 However, they have not been studied in Chinese medical journals. With more than 400 medical journals being used in China by large numbers of clinicians and readers, it is important to know if the statistical methods used in Chinese medical journals are appropriate and to determine what improvements should be made in the tradition of self-evaluation and improvement. We reviewed original research articles published in 5 leading Chinese medical journals in 1985 and 1995. We determined the frequency of research designs and statistical techniques used and the common statistical errors in the research reports. We also compared the articles published in 1985 and 1995 to describe changes over time.

METHODS

The 5 leading medical journals publishing clinical research in China were selected: Chinese Journal of Internal Medicine, Chinese Journal of Surgery, Chinese Journal of Obstetrics & Gynecology, Chinese Journal of Pediatrics, and Chinese Journal of Cardiology. These journals, sponsored by the Chinese Medical Association, represent the highest academic level in Chinese medical journals. All use peer reviewers for original articles. Two journals routinely involve a statistical reviewer for each original research article; the others involve a statistical reviewer when considering articles containing sophisticated statistical methods. We selected papers identified as original research articles in each journal’s table of contents. All 954 original articles published in 1995 and all 640 in 1985 were reviewed.

For each article, we recorded the research design and type of statistical methods used. We then judged whether statistical tests were necessary and, if used, whether appropriate. A structured form with a checklist was used.

We categorized statistical methods by a modification of the method used by Emerson and Colditz,1 omitting cost-benefit analysis and sensitivity analysis because no articles used them. Few papers used the categories of discriminant analysis (1 paper); adjustment and standardization (2 papers); power and sample size calculations (1 paper); receiver operating characteristic curve analysis (2 papers); or mathematical modeling (2 papers), so we classified them as “others.” If more than 1 statistical method was used in a paper, all were recorded; if the same statistical method was applied in the same paper repeatedly, it was recorded only once. We also recorded whether the authors specified which statistical methods they had used.

All articles were reviewed by both authors. Discrepancies were resolved by discussion. Both authors assessed 100 articles independently; agreement on whether the statistics methods used were appropriate was 91% (κ = 0.89).

The frequency of each category of statistical test used in an article was calculated in 2 ways: as a simple percentage of the total number of articles and as a percentage of articles that used statistical
methods. \( \chi^2 \) Tests were used to compare differences in use of statistical methods during the 10 years.

**RESULTS**

Compared with 1985, the original research articles in 1995 consisted of more clinical trials, retrospective studies, and basic science studies (31.1% in 1995 and 17.6% in 1985) \( (\chi^2 = 36.4; P < .001) \) and fewer case reports and retrospective studies (Table 1). For all study designs (except case reports), a higher percentage of articles contained statistical methods in 1995. Both in 1985 and 1995, clinical trials were most likely to use statistical tests, and case reports were least likely.

Of 954 original research articles published in 1995, 573 (60.1%) used statistical tests (Table 2). Of the 381 articles without statistical analysis, 93 articles needed but omitted statistical analysis, 71 were case reports, and 217 were descriptive studies using only descriptive statistics such as means, SDs, or percentages. Of 640 articles published in 1985, the proportion of the articles using statistical methods was 40.2%, significantly lower than in 1995 \( (\chi^2 = 60.9; P < .001) \). Of the 383 articles without statistical analysis, 60 articles needed but omitted statistical methods.

In both 1985 and 1995, the most commonly used statistical methods were \( t \) tests and contingency tables (Table 2). More sophisticated statistical methods were used in 1995 than in 1985. In 1995, 20 articles used multiple regression, while in 1985 only 5 articles used multiple regression. When presenting survival analysis, in 1995, 3 articles used the Cox regression model, while in 1985 none did and only 3 articles used simple survival analysis. The proportion of articles using more than 1 statistical method was higher in 1995, 260 (27.3%) of the 954 articles, than in 1985, 63 (9.8%) of the 640 articles \( (\chi^2 = 71.9; P < .001) \).

More articles that used statistical methods used them appropriately in 1995 (46%) than in 1985 (22%) \( (\chi^2 = 45.3; P < .001) \) (Table 3). For inappropriate analyses, the most common error was presenting \( P \) values without specifying the statistical methods the authors used, especially in 1985. Because more statistical tests were used in 1995, the percentage of articles using incorrect methods actually increased, from 24% in 1985 to 36% in 1995. The most common errors were using \( t \) tests instead of analysis of variance (ANOVA) to compare means of more than 2 groups and using unpaired \( t \) tests instead of paired \( t \) tests.

In 1995, the percentage of articles using statistical methods varied across the 5 journals from 39% to 70%; in 4 journals, the percentage was more than 60%. The proportion of articles using inappropriate statistical methods varied from 39% to 71%, with 2 of the journals at about 70% and the other 3 journals at about 40%.

**COMMENT**

We found that 60% of original research articles published in 5 major Chinese medical journals in 1995 used statistical tests, a significant increase over the 40% in 1985. In both years, \( t \) test and contingency tables were used most frequently, followed by ANOVA and Pearson correlation. The study by Emerson and Colditz\(^2\) of 332 original research articles published in 1979 in The New England Journal of Medicine found that 73% used statistical methods. If we exclude the 72 case reports in 1995 in our study, the proportion of articles using statistical methods would be 65%, slightly lower than the
study by Emerson and Colditz. Morris\(^5\) found that only 66 (23%) of 103 original articles published in an orthopedics journal in 1984 used correct statistical methods. Of 592 articles published in ophthalmic journals in 1990, 66% contained statistical techniques. Kanter and Taylor\(^6\) found that statistical methods were used in 48% of original research articles published in 1993 in *Transfusion*. Thus, use of statistical methods in leading Chinese medical journals appear to be at least as frequent as most journals around the world for which data are available.

The rank order of frequency of specific statistical methods used in our study was similar to other studies.\(^1,2,5,11\) However, articles in Chinese medical journals tended to use *t* tests and contingency table analysis more often and sophisticated methods less often. We found substantial improvement in the use of statistical methods over time. Rosenfeld and Rockette\(^11\) and Juzych et al\(^12\) also found improvement when comparing articles in the 1990s with those in the 1970s and 1980s.

Despite the improvement between 1985 and 1995, the proportion of articles using inappropriate statistical methods in 1995 was still high, and less than half, only 46%, used statistical methods correctly. Morris\(^5\) found that 46% of articles in the surgery journal published in 1984 used correct methods. The most common error in our study, presenting *P* values without specifying the test used, was also noted in other studies.\(^2,5,10\) In our study, the proportion of articles using nonparametric statistics and data transformation were only 1.7% and 0.8%, respectively, in 1995, much lower than other reports.\(^1,2,11\) Rosenfeld and Rockette\(^11\) documented increasing use of nonparametric statistics in articles from 0% in 1967 to 7.3% in 1989. The low proportion of articles with nonparametric statistics in Chinese medical journals may relate to the authors ignoring the assumptions underlying parametric statistics.

The frequency of certain study designs changed over the decade, with clinical trials, prospective studies, and basic science studies becoming more frequent. Even so, in 1995 these types of designs accounted for 31% of all articles, and clinical trials made up only 11%. This finding contrasts with that of McDermott et al\(^12\) who found that in 1991 clinical trials made up 35% of published articles in *JAMA, The Lancet*, and *The New England Journal of Medicine*.

This study covered all original research articles published in 2 different years over a decade, and the results give an estimate of the progress occurring in Chinese medical journals. However, since we studied only the 5 leading Chinese medical journals, it is likely that our results are better than the average level in Chinese medical journals.

The quality of using statistical methods in Chinese medical research is gradually improving, and the frequency of using statistical tests in Chinese medical journals appears comparable to that in other parts of the world. However, the lack or inappropriate use of statistics remains a problem. The type of problems we found, such as omitting descriptions of statistical methods or giving *P* values without specifying the statistical method, suggests that educating medical journal editors in research, design, and biostatistics might improve the quality of journal articles published in Chinese medical journals.

We gratefully acknowledge the support provided by Suzanne Fletcher, MD, and Robert Fletcher, MD, in reviewing the manuscript, making suggestions for revision, and improving our written English. We also thank Michael Phillips, MD, for his helpful advice and enthusiastic encouragement.

References