Some tips about statistics on medical research

Na Qin[1]*, Jiahui Zhang[1]*, Weihong Zhang[2], Juncheng Dai[1], Wensen Chen[2]

[1]Department of Epidemiology and Biostatistics, Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, School of Public Health, Nanjing Medical University, Nanjing 210029, China; [2]Department of Infection Management, the First Affiliated Hospital of Nanjing Medical University, Nanjing 210029, China

*These authors contributed equally to this work.

Correspondence to: Juncheng Dai. Department of Epidemiology and Biostatistics, Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Cancer Center, School of Public Health, Nanjing Medical University, Nanjing 211166, China. Email: djc@njmu.edu.cn; Wensen Chen. Infection Management Office, The First Affiliated Hospital, Nanjing Medical University, 300 Guangzhou Rd, Nanjing 210029, China. Email: metrischen@126.com.

Since the word "statist" first mentioned by William Shakespeare in Hamlet, statistics has kept its mystery from us for hundreds of years. Until late 1940s, a clinical report was published *British Medical Journal* (*BMJ*), which evaluated the effect of streptomycin treatment of tuberculosis with randomized experiments, the application of statistics in medical researches has gradually gained its popularity (1,2). An editor in *JAMA* even compared statistics to a pillar of medicine, highlighting the importance of medical statistics in medicine (3).

Nevertheless, we have seen universal misuse along with the wide use of statistics, especially in medical journals (4), which is immoral and can lead to serious clinical consequences. Given that, many medical journal editors have made great efforts to drawing detailed guidelines as Bertolaccini *et al.* mentioned (4,5), in order to reinforce the credibility of the results. However, little evidence shows that the quality of statistics has improved (6). Commonly, these typical errors often appear in the following several parts.

In research design phase, it is crucial to make sure of the aims of the study, outcome variables and end-points. What's more, surgeons should always keep principles of randomization, control and binding in mind. Although these principles may not be the golden standard, they have long been regarded as the best methods to avoid statistical bias (7). Moreover, one should calculate the sample size according to the formula accurately under different types of research, so as not to omit the estimation of sample size.

For analysis, surgeons should thrash out whether the data is continuous or categorical at first, which ought to be analyzed with different methods. In short, for normally distributed data, student's *t*-test and χ^2 test are used for continuous variables and categorical variables respectively (8). While the data is not in normal distribution, nonparametric tests might be a better choice. If a study contains multiple comparison tests, it is necessary to apply multiple comparison correction, which can greatly reduce the false positive rate. Before making a decision, it is also noticeable that whether there are confounding factors need to be adjusted with multivariate techniques, as confounding factors may bring a false association. Of course, this standard is not applicable for all types of studies; it should be selected according to the specific situation.

When it comes to the interpretative analysis, more attention should be put on this phase. In most cases, it's better to provide P value (or corrected P values) and the test statistics at the same time (9). Significance test is the most commonly means to investigate whether there are differences between different groups under the null hypothesis. Nevertheless, P value cannot be the only criterion in most clinical researches. Conventionally, if a P value is greater than 5%, it is thought to be "not significant" (9), and the results are "negative". However, this doesn't

actually mean that there is no difference, the "negative" result may just be due to the absence of adequate power to detect such a difference, such as lack of enough sample sizes or overcorrection of the P value. Actually, even if you get a "positive" result, this doesn't necessarily represent clinical relevance, either. Usually, it is preferable to declare a significant statistic difference at the 5% level. In other words, this means that one false positive out of twenty tests is permitted, while sometimes this false positive is not welcomed in clinical medicine. Moreover, as the example presented in Bertolaccini *et al.*, even though you find a significant difference between two similar clinical indexes, this may not mean clinical significance. As you can see, no more statistical methods is required in this phase, and all the interpretative analysis are based on the result of statistical analysis, yet one should still be careful when claiming the significance of effects.

Clinical research is a major event related to human health, so it is imperative to improve its credibility. Besides the sufficient professional knowledge of surgeons, three other indispensable parts are also needed to conduct a high quality clinical research, including well research design, precise statistical analyses and methodical interpretation (9). In view of these, surgeons should think clearly at the very beginning. Although these principles seem to be simple, it is of great importance to implement statistics carefully and accurately throughout the process, so that the results will be more credible and meaningful.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Altman DG. Statistics in medical journals: developments in the 1980s. Stat Med 1991;10:1897-913.
2. Nelder JA. Statistics in medical journals: some recent trends. Stat Med 2001;20:2205.
3. A Pillar of Medicine. JAMA 1966;195:1145.
4. Bertolaccini L, Viti A, Terzi A. The chicken-and-egg debate about statistics and research. J Thorac Dis 2014;6:1349-50.
5. Goodman SN, Altman DG, George SL. Statistical reviewing policies of medical journals: caveat lector? J Gen Intern Med 1998;13:753-6.
6. Cooper RJ, Schriger DL, Close RJ. Graphical literacy: the quality of graphs in a large-circulation journal. Ann Emerg Med 2002;40:317-22.
7. Strasak AM, Zaman Q, Pfeiffer KP, et al. Statistical errors in medical research--a review of common pitfalls. Swiss Med Wkly 2007;137:44-9.
8. Brook AD, Ahrens TS, Schaiff R, et al. Effect of a nursing-implemented sedation protocol on the duration of mechanical ventilation. Crit Care Med 1999;27:2609-15.
9. Jin Z, Yu D, Zhang L, et al. A retrospective survey of research design and statistical analyses in selected Chinese medical journals in 1998 and 2008. PLoS One 2010;5:e10822.