Common pitfall of statistical analysis I: Alpha inflation

Chung-Hong CHAN 陳仲康 and **Daniel Kwok-Keung NG** 吳國強
Department of Paediatrics, Kwong Wah Hospital, Hong Kong

In clinical study, one accept a pre-selected Type I error rate (alpha), i.e. false positive, of 5%. This error may be inflated (alpha inflation) by analyzing the same datum more than once. Alpha inflation refers to the phenomenon that the more statistical tests, e.g. three chi-square test for three pairs of results derived from the same pool, the more likely one could find a significant result when it is actually not. The problem of alpha inflation is a frequently ignored problem in medical statistics. When an individual decides to compare the three groups by three chi-square tests, the true alpha or Alpha' can be estimated by the following equation:
Alpha' =1-(1-Alpha) $^{\text{no. of test}}$

An example is given in Table 1.

Table 1. Academic performance in relation to daytime sleepiness

	Excess daytime sleepiness: Yes	Excessive daytime sleepiness: No	p
Good academic performance	83 (5.7%)	1378 (94.3%)	0.019
Average academic performance	87 (6.6%)	1238 (93.4%)	
Poor academic performance	19 (11.2%)	150 (88.8%)	

For this 3x2 table, chi-square test was used and only one p-value is reported. In this example, the test statistics and p-value are 7.94 (df=2) and p=0.019 respectively. You conclude that the difference in the frequency distribution of academic performance in excessive daytime sleepiness (EDS) subjects and non-EDS subjects are statistically significant. You go on to ask which of the above three groups has a higher proportion of subjects with EDS. You proceed to draw three 2x2 tables and analyse by chi-square test for three times. You will then conclude that two significant differences are obtained, i.e. good vs bad and average vs bad, when the truth is only one significant difference, i.e. good vs bad (Table 2). The reason for the difference is attributed to the fact that each group are compared twice in these three comparisons in the uncorrected case. The type I error rate in the uncorrected case is $1-(1-0.05)^3$ = 14.26% and NOT the preset 5%.

Table 2. Interpretation of Chi-square tests in uncorrected alpha and Bonferroni adjusted alpha

Comparisons	p-value from Chi-square test	Using conventional alpha of 0.05, i.e. alpha inflation	Using Bonferroni adjusted alpha of 0.016
Good vs Bad	0.008	Reject null hypothesis	Reject null hypothesis
Average vs Bad	0.038	Reject null hypothesis	Accept null hypothesis
Good vs Average	0.371	Accept null hypothesis	Accept null hypothesis

The common way to deal with this kind of problem is simple: to adjust the Type I error rate lower than 0.05 by Bonferroni adjustment. The Bonferroni adjusted level of alpha can be easily calculated by the following equation: Adjusted alpha=alpha/no. of test. Using the same example, we carry out three chi-square test for three tables. We will choose an adjusted alpha of 0.05/3=0.016. Therefore, we will only reject the null hypothesis when p<0.016, instead of 0.05. It can ensure the overall chance of making Type I error to be less than 0.05.

Although Bonferroni adjustment is a simple technique to address the problem of alpha inflation, it is not without problem. Bonferroni adjustment is known for being too conservative and may increase the chance of higher than intended Type II error, i.e. false negative. The problems of Bonferroni adjustment were discussed in the papers by Perneger and Rothman.

References:

1. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995;310:170.
2. Perneger TV. What's wrong with Bonferroni adjustments. BMJ 1998;316:1236-8.
3. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990;1:43-6.