Published in final edited form as:

J Clin Epidemiol. 2009 October ; 62(10): 1112–1113. doi:10.1016/j.jclinepi.2009.04.009.

Valid, Adaptive, Data-Driven Comparisons of Binary Endpoints

Vance W. Berger, PhD

National Cancer Institute and University of Maryland Baltimore County, Biometry Research Group, National Cancer Institute, Executive Plaza North, Suite 3131, 6130 Executive Boulevard, MSC 7354, Bethesda, MD 20892-7354, (301) 435-5303 (voice), (301) 402-0816 (fax), vb78c@nih.gov

Editor: Austin and Goldwasser [1] missed a key opportunity to point out that in fact it is perfectly valid to use the data to select the cutoff for dichotomization in a 2xk contingency table as long as the proper penalty is paid for doing so. In general, this penalty tends not to be so great, and the approach saves one from having to pretend to know prospectively where the treatment effect will manifest itself maximally. In fact, an adaptive approach can be used whenever any one of several analyses may prove to be the most impressive. One simply computes each p-value, selects the lowest among these, and then uses it not as a p-value *per se* but rather as a test statistic for use with a design-based permutation test [2]. The question is how low is this lowest p-value relative not to a uniform distribution on the unit interval but rather relative to the permutation distribution of similarly computed minimum p-values. This latter reference distribution will, of course, be stochastically smaller than a uniform distribution on the unit interval, so the raw minimum p-value will be adjusted upwards by filtering it through this validation transformation. When all analyses considered are binary analyses resulting from dichotomizing a categorical endpoint all possible ways (the Lancaster decomposition), the adaptive test described reduces to the Smirnov test [3].

References

1. Austin PC, Goldwasser MA. Pisces did not have increased heart failure: data-driven comparisons of binary proportions between levels of a categorical variable can result in incorrect statistical significance levels. *Journal of Clinical Epidemiology* 2008;61:295–300. [PubMed: 18226754]
2. Berger VW, Ivanova A. Adaptive tests for ordinal data. *JMASM* 2002;1(2):269–280.
3. Permutt T, Berger VW. Rank tests in ordered 2xk contingency tables. *Communications in Statistics, Theory and Methods* 2000;29(5):989–1003.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.