Endnotes for Online Supplement

[i] This is not to say that two-sample parametric tests of location do not rely on shape restrictions. A classic example is the Behrens-Fisher problem of testing for a difference in population means in the presence of unequal variances (Kim and Cohen 1988). The historical distinction being, however, that parametric tests of location such as the *t*-test, were *designed* specifically to test the shift hypothesis whereas two-sample linear rank tests were designed to test for differences in distribution (e.g., Mann and Whitney 1947).

[ii] Instead we could have employed either full-enumeration permutation testing or sampled from the permutation distribution (pp. 14-15 in Manly 1997). However, this would have made Monte Carlo simulation run times impracticably long.

[iii] Alternatively, one could take a data-adaptive approach and apply standard chi-square testing when cell expectations are not small.

Kim S and Cohen AS. On the Behrens-Fisher problem: a review. *J Educ Behav Stat* 1998; 23: 356-377.

Mann HB and Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 1947; 18: 50-60.

Manly BFJ. *Randomization, bootstrap and Monte Carlo methods in biology*, 2nd ed. London: Chapman & Hall, 1997.