not endorse such exemptions. Researchers should instead consider the context-specific consequences of type I and type II errors in the given situation.

Note that the most extreme arguments against multiplicity adjustment that were cited by Streiner (1) come from unempirical articles that seem to fundamentally misunderstand the topic. For example, Schulz and Grimes (9) depicted Bonferroni adjustments as omnibus tests of the "universal null hypothesis" that do not allow significance to be localized to individual tests—a known misconception (2).

It is true that "the decision regarding whether or not to correct for multiple testing is a philosophical one" (1), and many other decisions are similarly subjective, such as what level is appropriate, which error rate (e.g., FWER, FWER, or FDR) is most relevant, and which tests to include in the family. But the potential for type I error inflation caused by disregarding multiplicity is a mathematical reality that is not up for debate.

The author declared no conflicts of interest.

Andrew V Frane

From the Department of Psychology, University of California, Los Angeles, Los Angeles, CA (E-mail: avfrane@ucla.edu).

REFERENCES


Reply to AV Frane

Dear Editor:

I thank Frane for his thoughtful comment. He raises a number of important points, but I am afraid the commentary is also based in part on a misreading of the article, and statements of "facts" are, at best, subject to debate. To begin with, he disagrees with the statement that the Bonferroni, Holm, and Hochberg procedures "[assume] that the null hypothesis is true for all of the tests" and states that, "the Bonferroni procedure accommodates any proportion of the null hypotheses being true, not just the 'worst case' scenario that all null hypotheses are true." Perneger (1), however, states that "The Bonferroni method is concerned with the general null hypothesis (that all null hypotheses are true simultaneously)" (p. 1236). Similarly, Frane also objects to the statement that the Bonferroni and Holm procedures assume the independence of the tests (there is no argument that the Hochberg approach does require it). Abdi (2), on the other hand, states that the Šidák–Bonferroni correction, \[1 - (1 - P_{	ext{FWER}})^{1/k}\] (where FWER indicates the familywise type I error rate), is derived assuming the independence of tests. He then says that the Bonferroni correction is a simpler approximation of this, introduced because of the difficulty involved in working with fractional powers. This would therefore imply that the same assumption, the independence of tests, would also apply in the latter case, a position echoed by McDonald (3). Although others [e.g., Castellada et al. (4) and Frane] would disagree, I do not think this issue is as black and white as Frane suggests.

A misreading of the article is exemplified by the fact that the Šidák–Bonferroni procedure is not described in the article as a multistep method; this adjective was applied only to the Holm and Hochberg procedures, which Frane states were incorrectly described. Most helpfully, he does not elaborate in what way these latter procedures have been misrepresented in the article, making any response to this statement impossible. Furthermore, the statement that all null hypotheses in a family are true is "unreasonable" was made within the context of adjusting for multiplicity after a significant omnibus F test. If the overall test is significant, then, by definition, at least one of the null hypotheses must be false. A further misreading of the article occurs in his first footnote. He is correct that the Holm procedure was not designed with the false discovery rate (FDR) or positive false discovery rate (pFDR) in mind, but I did not say it was; the article explicitly states that although "it was developed before the term FDR was introduced, it can be seen as the first—and now perhaps best known—of the techniques to control the pFDR." Finally, Frane states that my claim that simultaneous significances in correlated outcomes reinforce each other and strengthen confidence in the results is "misguided," because "when outcomes are positively correlated, outcomes reinforce each other and strengthen confidence". However, this overlooks the converse argument: If 2 (moderately) correlated outcomes both support rejection of the null, then this should give further credence to the effectiveness of the intervention, not weaken it, as a Bonferroni correction would.

The author declared no conflicts of interest.

David L Streiner

From the Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario (e-mail: streiner@mcmaster.ca).

REFERENCES