Searching for a Defensible Application of Alpha-Adjustment Tools

DANIEL J. O'KEEFE
University of Illinois at Urbana-Champaign

Any principle offered to underwrite the imposition of alpha-adjustment procedures must be applied consistently. Hewes (2003) and Tutzauer (2003) propose that alpha adjustment can appropriately be mandated when the set of tests concerns the key claims of a single theory. But consistent application of this reasoning though, commits one to bizarre beliefs and research practices—which suggests that this reasoning does not provide a satisfactory rationale for imposing alpha adjustments.

Despite the thoughtful commentaries of Hewes (2003) and Tutzauer (2003), I continue to believe that there is not yet a well-articulated principled basis for imposing alpha-adjustment procedures. To be sure, the parties to this discussion apparently agree on quite a few things. We all believe that more reflective use of methodological tools is very much to be desired. I think we agree that the best protection against Type I error is replication. And we surely all agree (to use Tutzauer's language) that alpha adjustment should be undertaken "only when it makes sense to do so." What separates us is whether we think there are in fact circumstances in which it "makes sense" to impose such adjustments.

My starting point—and I believe this premise also to be shared by Hewes and Tutzauer—is the belief that the mandated application of alpha-adjustment procedures requires a principled basis. That is, any claim that alpha adjustment is to be mandated in a specified circumstance requires a general justification that is applied consistently: Whenever the specified conditions are met for the mandated application of alpha-adjustment procedures, then such procedures have to be invoked. (If one's rationale for imposing alpha adjustments in one case involves conditions X, Y, and Z being met, then one is committed to believing that whenever those conditions are met, alpha adjustment is required. Otherwise, one's rationale is not actually a principled one.)

Daniel J. O'Keefe (Ph.D., University of Illinois, 1976) is a professor in the Department of Speech Communication at the University of Illinois at Urbana-Champaign. Correspondence concerning this article should be addressed to Daniel J. O'Keefe, Department of Speech Communication, 702 S. Wright St., 244 Lincoln Hall, University of Illinois at Urbana-Champaign, Urbana, IL 61801-3694, USA; email: dokeefe@uiuc.edu.

In this regard, the logic of some of my earlier arguments was apparently not as clear as it might have been. In a number of ways I tried to show that no sensible person should accept the consequences of consistent application of the principle commonly invoked to justify alpha adjustment. That principle (the usual justification for alpha adjustment) is based on the sheer number of tests conducted. I argued that if one really believes that the sheer number of tests mandates alpha adjustment, then one would necessarily be committed to all sorts of palpably bizarre research practices—committed (for instance) to believing that my reanalysis of Professor P's data was unproblematic, to thinking that careerwise alpha adjustment would be a good idea, and so forth. Any methodological principle whose consistent application has the consequence of committing one to such absurd practices is plainly a defective principle.

Unfortunately, this reasoning was apparently not as transparent as it should have been, leading to some misapprehension of the point of some of my examples. For instance, concerning the example of my reanalysis of Professor P's data: The point is not that I want some statistical method to detect my unethical behavior (the current editorial system should detect it well enough). The point instead is that, according to the logic of the customary rationale for imposing alpha-adjustment procedures, there is nothing wrong with my hypothetical underhanded conduct, which suggests that something is amiss with that rationale. That rationale would not merely permit but actually require alpha adjustment when I conducted those additional tests, but allowing, much less insisting on, such adjustment would obviously be foolish. The point is not that we need some way to stop my hypothetical conduct; the point is that there is something wrong with a methodological rationale that underwrites such conduct. Indeed, that was the general logic of several of the arguments in the original essay: Consistent application of the usual justification for alpha adjustment would lead us to endorse plainly defective research practices, which means there must be something wrong with that justification.

So the question is: Is there some way of mandating application of alpha-adjustment procedures that does not have the consequence of committing one to obviously defective research practices? Put somewhat differently: Can one identify a principled basis on which to mark out a collection of tests as one over which alpha adjustment is required, without having consistent application of that principle lead to the endorsement of unacceptable research practices?

Hewes and Tutzauer both offer one such possible principled basis connected to the testing of a theory's set of claims. Tutzauer (2003) would mandate alpha adjustment in a circumstance involving testing "a strong theory making a web of explicit claims, [where] the invalidation of any single claim (i.e., the truthfulness of any single null)[invalidates] the theory

as a whole" (p. 457). Hewes (2003) suggests that "if the set of hypotheses all need to be supported for a theory to be plausible, then the reasons for the success (or failure) of the set as a whole need to be assessed" and hence the imposition of alpha-adjustment procedures is appropriate (p. 452). (Hewes's and Tutzauer's views are not quite identical, but for present purposes their differences can be put aside.) Broadly put, then, the proposed principle is that alpha-adjustment procedures are to be mandated in circumstances in which the researcher wants to consider the impact of data on a theory as a whole. So, for instance, when the tests reported in a single article all bear on key claims of a given theory, experimentwise alpha adjustment is said to be essential. In this circumstance, the researcher is interested in the results of the tests as a set—a set bearing on the theory and thus a coherent whole about which conclusions are wanted—and hence minimizing the chances of error anywhere in the set is important.

This proposed rationale for alpha adjustment is crucially different from the usual justification. The common rationale for imposing alpha-adjustment procedures is Type I error risk inflation per se, that is, the risk occasioned simply by having a large number of tests, no matter the basis for marking out the set of tests. This new suggested rationale invokes a concern about errors in theory acceptance and rejection and hence proposes that the unit of collection (the abstract category under which the tests are to be collected) is "tests relevant to a given theory." Because the rationales are different, this new approach does indeed avoid many of the absurd consequences of adopting the customary rationale. For instance, this principle would not endorse alpha adjustment simply because one investigator conducted 20 tests but would require alpha adjustment if those 20 tests were all tests of the same theory.

But in the end, I believe that this way of marking out a collection of tests for mandatory alpha adjustment has consequences just as unpalatable as those associated with the usual rationale. Consider a concrete case, with alternative scenarios. Suppose I test 20 hypotheses (each of which is important to the plausibility of the theory) concerning a given theory. In Scenario A, I report these 20 tests in a single experimental report. The proposed principle requires alpha adjustment; with a simple Bonferroni correction, for instance, the nominal alpha would be .0025 (.05/20).

In Scenario B, I report these 20 tests in two different articles, 10 tests in each. If the proposed principle requires alpha adjustment only within each article, then a different alpha level will be employed than in scenario A. (Using a Bonferroni adjustment, alpha will be .005 in each of the two articles.) But surely the proposed principle requires transcending the level of the particular article; surely the principle could not be subverted simply through choice of publication format (spreading the tests across multiple publications). In short, the proposed principle must

require that all the relevant tests of this theory should be treated as a collection requiring alpha adjustment—not just those tests that happen to be reported in a particular article. So, for example, if the tests of a given theory appear in a number of different articles by a number of different investigators, then all those tests should, according to the proposed principle, have alpha-adjustment procedures applied to them as a set.

Thus the proposed principle would, if consistently applied, have two consequences. First, each newly reported test of a theory would be required to adjust its alpha level so as to take into account the number of tests reported previously. So, for example, if 20 tests of a theory had previously been performed, and I conduct 5 new tests, the alpha level for my new tests would need to reflect the number of previously conducted tests. Applying a Bonferroni correction, the nominal alpha level for my 5 new tests would be .002 (.05/25). It is left as an exercise for the reader to estimate the nominal alpha level required by the proposed principle for any new test of some social-scientific warhorse such as cognitive dissonance theory. (My own guesstimate puts it somewhere south of .00005.)

Second, whenever new tests of a theory were reported, investigators would be required to revisit earlier-reported findings and make corresponding alpha adjustments. So, to continue the example, my having conducted 5 new tests of that theory would mean that the 20 previously reported tests (presumably conducted with an alpha level that reflected the performance of 20 tests) would now have to be reassessed using a more stringent alpha level, reflecting the performance of a total of 25 tests of the theory.

No sensible person would be willing to accept these consequences. That is to say, no sensible person would be willing to endorse or adopt the practices that the proposed principle requires. And these consequences are indeed unavoidable, given consistent application of the proposed principle. To avoid these consequences, one would have to believe that "the alpha level must be adjusted for tests of a theory when those tests are all reported in one article simultaneously, but alpha adjustment is not required when those tests are reported in different articles or at different times." Obviously such a belief abandons the proposed principle.

To come at the same point from a different angle: Anyone who worries about overall error when 20 tests of a given theory are reported in a single article must, to be consistent, have identical worries whenever 20 tests of a given theory are reported—in one publication or many, by one investigator or many, simultaneously or spread over time. And so such a person must be committed to correspondingly identical application of alpha-adjustment procedures across all those circumstances, with the absurd consequences just noted.

In short, I believe that "all the tests of a given theory" is no more suitable as a unit of collection than "all of the tests that appear in a given article" or "all the tests conducted by a given investigator" or "all the tests that appear in a given journal issue." The proposed principle, if applied in a genuinely principled (consistent) manner, would commit the research community to practices fully as bizarre as those compelled by the more usual justifications for alpha-adjustment procedures. We should not endorse any principle that commits us to such practices.

It may well be that some more suitable principled justification can still be devised. Even if one believes, as I do, that no adequate rationale is yet in hand, the possibility of some future satisfactory principle cannot be ruled out. But in the absence of a well-articulated defensible rationale for the imposition of alpha-adjustment procedures, I continue to believe that the practice of requiring or employing such adjustments should be abandoned.

REFERENCES

Hewes, , D. E. (2003). Methods as tools: A response to O'Keefe. *Human Communication Research, 29,* 448–454.

Tutzauer, F. (2003). On the sensible application of familywise alpha adjustment. *Human Communication Research, 29,* 455–463.