Making multiple claims from one data set requires some care with the statistical analysis; otherwise, scientists making the claims and their readers might easily be fooled by randomness. In the Women’s Health Initiative Dietary Modification Trial, two primary questions related to a low-fat vs a normal diet and cancer risk were examined (1–2). These hypotheses were motivated by extensive epidemiology work and were expected to be true based on considerable evidence. In fact, neither of these claims was validated in the randomized trial. That these claims were not validated surprised some people, but the results should have been expected; Ioannidis (3) pointed out that claims from nonrandomized medical studies fail to replicate approximately 80% of the time. One reason (4) that claims coming from epidemiology studies fail to replicate is that epidemiologists routinely do not correct their statistical analysis for multiple testing (5).

One of the simplest ways to correct for multiple testing is to use the Bonferroni correction, whereby the number of questions under consideration is counted and the unadjusted \( P \) values are multiplied by the number of questions. More sophisticated methods are available [6] among many others. Of course, the data analysis strategy and the counting should be specified before the data are examined.

Counting the questions Prentice et al. (7) have under consideration is a bit elusive. There are two named primary questions, breast and colorectal cancers. A normal clinical trial strategy would be to multiply each by two to give adjusted \( P \) values; any possible effects on secondary endpoints would need to be verified in a new study. In fact, five types of cancer are under consideration: breast, colon, rectum, ovary, and endometrium, as are three time periods: the first 4 years, the next 4.1 years, and total, giving a Bonferroni multiplier of 15. There are more than 60 \( P \) values reported in tables 1–5, and none are adjusted for multiple comparisons. It is possible that many additional \( P \) values were computed but are not reported. The three smallest \( P \) values are .03, .05, and .05. Any reasonable adjustment renders these \( P \) values non–statistically significant; seeing a \( P \) value as small as .03 is quite likely by chance alone. In clinical trials, results must be statistically significant in two studies to make a claim, and the usual convention of science is that anyone wishing to assert a claim should have strong supporting evidence. Although technically correct, the summary statement “A low-fat dietary pattern may reduce ...” could easily be misinterpreted. More accurate would be the statement, “Epidemiology studies suggested that five types of cancer might be reduced by adhering to a low-fat diet. A well-conducted, large, randomized trial gives no statistical support to any of these claims.”

S. STANLEY YOUNG

References

Notes
Affiliation of author: National Institute of Statistical Sciences, Research Triangle Park, NC.
Correspondence to: S. Stanley Young, PhD, National Institute of Statistical Sciences, PO Box 14006, Research Triangle Park, NC 27709 (e-mail: young@niss.org).
DOI: 10.1093/jnci/djm309 © The Author 2008. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oxfordjournals.org.

Response
We thank Dr Young for drawing additional attention to the important multiple testing issue, both in the context of our Women’s Health Initiative (WHI) report and in epidemiologic research more generally. However, our report did explicitly address this issue, and we believe that our conclusion that “a low-fat dietary pattern may reduce the incidence of ovarian cancer among postmenopausal women” is both technically correct and an accurate summary of these valuable data.

The WHI randomized controlled Dietary Modification Trial examined the potential of a low-fat dietary pattern to reduce invasive cancer risk. The protocol prospectively designated breast and colorectal cancer as primary outcomes, whereas breast, colon, rectum, ovary, and endometrium were listed as “diet-related outcomes” that may benefit from the intervention tested. A weighted log-rank test was specified to compare intervention and comparison groups, with weights increasing linearly from zero at randomization to a plateau of 1.0 at 10 years after randomization.

The weighted log-rank statistical significance level was \( P = .03 \) for ovarian cancer. In interpreting this result, we noted that a statistical significance level as extreme as \( P = .03 \) could occur by chance alone with a probability as large as 15% when five comparisons are conducted, using a conservative Bonferroni correction. However, we also offered additional evidence in support of an ovarian cancer risk reduction in the intervention group. These results included a hazard ratio trend test with time from randomization for which the statistical significance level \((P = .01)\) is not readily attributable.
to chance and lower ovarian cancer hazard ratios ($P = .05$) among women whose baseline dietary percentage energy from fat was relatively high (and who make comparatively larger reductions in the fat content of their diets). Statistical significance levels were described as “nominal” to alert the reader to the lack of multiple testing considerations in these supplementary analyses.

Tests were also presented for a list of other cancer sites because we anticipated that these would be of interest to the cancer and public health research communities. We included the proviso that these additional tests need to “be interpreted in the context of the entire set of approximately 25 site-specific comparisons.” None of these additional analyses provided evidence for an intervention effect in the light of these multiple testing issues, and we described these results as “readily attributable to chance.”

Given the trial’s cancer focus, we also reported a nominal statistical significance level for total invasive cancer ($P = .10$). We described this $P$ value as a “suggestion” and did not draw a related conclusion.

Dr Young’s perspective that outcomes beyond the primary (breast and colorectal cancer) endpoints “would need to be verified in a new study” is not practical given the cost and logistical requirements of a dietary intervention trial on this scale. This, and the fact that virtually all other available data on the association of dietary fat to cancer risk are observational, argues for making the fullest defensible use of data from the present trial.