Many eyes will be drawn to the article reported in this issue of the Journal by Veronesi et al. (1), who tested the effect of the vitamin A analogue fenretinide on the prevention of second breast malignancies in women with early breast cancer. Veronesi et al. randomly assigned nearly 3000 women to receive either fenretinide orally or no treatment (control group) for 5 years and followed them for a median of 8 years. There were totals of 136 contralateral and 221 ipsilateral second malignancies that appear equally likely to occur in the fenretinide or control group. In addition to this overall finding, the article highlights a possible treatment–covariate interaction, wherein fenretinide appears to reduce the risk of second cancers in premenopausal women but appears to increase the risk in postmenopausal women.

The trial appears to be well designed and conducted. There is no obvious source of bias or error that would explain the observed findings. The magnitude of the interaction, a risk increase or a risk reduction of 20%–30%, the marginal statistical significance level of .05, and the fact that the interaction test was not specified before the data were in hand seem tailor-made to stress our inferential abilities. The study illustrates important issues in the design, analysis, reporting, and interpretation of clinical trials.

The potential for observing either a quantitative interaction (magnitude of the treatment effect) or a qualitative interaction (direction of treatment effect) was not part of the study design. Some investigators might be tempted to disregard the observed interaction because it was discovered in a post hoc analysis. There is a statistical perspective that encourages such reasoning, and it is sometimes carried to the extreme in a regulatory environment when only prespecified analyses are permitted.1 Our approach here should be more reasoned.

One important issue in evaluating the observed interaction is the role of chance. Was the effect discovered by an exhaustive search through dozens of possible interactions, or was it found as a result of a thoughtful, parsimonious, biologic question appropriately applied to existing data? Which came first—the question or the analysis? Details regarding these matters help to assess the play of chance, which is incompletely characterized by the “statistical significance level.” Veronesi et al. do not provide enough information for us to be completely comfortable about these questions.

Dismissing the observed interaction because the type I error is higher than the nominal value is not appropriate. Doing so requires excessive emphasis on significance testing (P values) and “protection” of the overall type I error. Suppose, for example, that no post hoc analyses were performed until an investigator hypothesized on a biologic basis that an interaction might be expected. Then the test was performed. Such an analysis would be “prespecified” with respect to the statistical test, but it would be “post hoc” with respect to the data collection. Although some might assess the role of chance differently in this case, the evidence would remain the same.

The substantive question with respect to interpretation is whether or not corroborating biologic knowledge exists regarding the observed interaction. If there were none, we would require additional evidence before believing the observation. On this point, Veronesi et al. provide a rationale for the differential effect of fenretinide based on modulation of circulating insulin-like growth factor-I, for which evidence became available after this study was initiated. Such an effect could be consistent with the reciprocal effects in premenopausal and postmenopausal women. Even so, this is a biologic hypothesis rather than a proven mechanism.

It appears that we can neither dismiss fenretinide as producing no effect nor can we endorse it for use in premenopausal women solely on the basis of findings from the trial. Additional empirical evidence must be sought. It is unfortunate that such a large and lengthy clinical trial has left us with unresolved fundamental questions. But at least the new questions can be more sharply focused.

**Reference**


© Oxford University Press