The fields of medicine and public health have improved markedly over the last century owing in no small part to the rigorous application of statistical design and analysis to questions of human health. As Edwin B. Wilson wrote in 1923, “To a certain extent, we are all necessarily statisticians, whether doctors or not” (1). By the mid-20th century, the efforts of Ronald A. Fisher and A. Bradford Hill had finally succeeded in establishing the value of randomized clinical trials for reliably answering biomedical questions (2). For cancer researchers, not only has the use of randomized trials accelerated the identification of new, more effective treatment for diagnosed cancer and led to markedly improved survival even for some late-stage cancer patients, but using this design the subfield of cancer control and prevention has made huge strides in identifying methods for early detection and prevention of highly lethal cancers, to the extent that these methods were recently credited with the lion’s share of the cancer incidence and mortality reduction observed in the United States over the past few years (3). For example, both screening for breast cancer by mammography and screening for colorectal cancer by fecal occult blood testing have been shown by several randomized trials to reduce mortality from these two cancers (4–7).

The principle of a basic randomized trial is concurrently simple and powerful: the causal effect of an intervention can be directly estimated by creating two groups of individuals who do not differ systematically but only by sampling error and then intervening in one group and not the other. Only two hypotheses compete to explain differences after the intervention: the intervention itself or sampling error. The latter is controlled by replication (ie, a large enough sample size) and by statistical analysis. What remain are the intervention’s effect and bounded uncertainty about its size. Extending this idea to interventions on groups of individuals rather than one at a time, as when a worksite is offered screening, is also simple but deceptively so. The power of a trial randomizing individuals is still there, but controlling and quantifying the uncertainty is no longer a simple matter. Because individuals within the unit of randomization may have correlated outcomes, calculations based on sampling variability must take this correlation into account. Ignoring this correlation will fool the researcher into believing more certainty results than is justifiable. It has taken the methodological community some time to come to grips with this challenge, and it appears that researchers still lag behind.

Murray et al. (8) have provided the cancer research community—especially those engaged in screening for or preventing cancer—an excellent but somewhat disheartening summary of the state of the literature in regard to the use of what is an invaluable tool in the researchers’ arsenal, the group-randomized trial. Murray et al. (8) reviewed a fairly representative sample of papers featuring group-randomized trials that address cancer and were published from 2002 to 2006. They found the majority to be lacking in rigorous application of statistical methods in both their design and analysis. Not only were they unable to find any mention of sample size calculations in nearly half the papers, but less than a quarter of them gave an appropriate sample size calculation. Further, when evaluating the analytic methods, the authors found that more than half used invalid methods of analysis, primarily methods that understate variability and thus overstate statistical significance. Whereas flaws in design can lead to underpowered studies and perhaps point to gaps in the knowledge of those who review grant proposals, flaws in the analysis, as Murray et al. (8) point out, can lead to false findings of efficacy. The appropriate analysis of an underpowered study will at least reveal that deficit; an inappropriate analysis of a trial hides the true meaning, whether it is adequately designed or not.

Separation of Health and Statistics

Timothy R. Church

Copyright 2008. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oxfordjournals.org.
It is, perhaps, not too surprising that a relatively new tool is so often misused, however. It is certainly consistent with what has been observed with other statistical applications in the health literature. Much has been written, for example, about the shortcomings of observational study analysis with regard to the assessment of systematic biases, which are often ignored in quantifying the uncertainty of estimated associations. Although methods have been developed to account for them directly (9), measurement error, unmeasured confounders, and selection bias have all been routinely given short shrift in the analysis and reporting of observational studies, and a consequence of this fact is that findings from such studies too often turn out to be spurious. This unreliability may contribute to the public’s growing disaffection with the health recommendations coming out of such research. Many statistical applications, such as proportional hazards regression, often are misused to overstate certainty, and some that go unused, such as sensitivity analyses, should be used more often to avoid mistaken inferences.

This disjunction of health research and statistical practice has consequences that go far beyond academic niceties. The initiative to improve the health of all individuals depends heavily on the appropriate interpretation of well-designed studies, an activity that depends in turn heavily on the close marriage of researchers and statisticians. The separation of questions of health from those regarding statistical inference inevitably leads to poorer answers. Papers like the review of Murray et al. (8) help to make more researchers aware of the need to collaborate with statisticians to apply appropriate statistical design and analysis methods to their studies and so avoid the mistakes the authors identify. By assigning the correct level of uncertainty about research findings through valid design and analysis, the frequency with which conclusions are overturned will decrease and our own and the public’s confidence in our therapies and health recommendations will inevitably increase.

References