Screening Trials Are Even More Difficult Than We Thought They Were

Helen G. Juffs, Ian F. Tannock

In this issue of the Journal, Black et al. (1) present an important analysis of methodologic pitfalls associated with randomized studies of screening interventions. The authors compare disease-specific and all-cause mortality from the 12 published randomized trials of cancer screening for which these end points were available. In seven of the 12 studies, major inconsistencies were detected in the direction or magnitude of these two outcomes. Black et al. propose that the use of disease-specific mortality as the primary end point renders screening trials subject to at least two forms of serious bias:

1) Sticky-diagnosis bias—where deaths from other causes in the screened group are wrongly attributed to the target cancer or deaths in the control group are wrongly attributed to other causes. This bias favors the control group.

2) Slippery-linkage bias—where deaths due to diagnostic or therapeutic interventions that are stimulated by the screening test are not included in disease-specific mortality. This bias favors the screened group.

Because both of these sources of bias arise from determination of cause of death, they do not affect all-cause mortality. This provides a strong argument for the use of all-cause mortality as the primary endpoint in screening trials.

The comparison of all-cause mortality and disease-specific mortality in the article by Black et al. (1) suggests that the net effect of these biases has favored screening and that slippery linkage is more important than sticky diagnosis. Slippery linkage is also inherently more important because it recognizes that screening has the capacity to cause death as well as to prevent it. In any screening study (or in the population to which its results are applied), the risk of the screened population dying of the target disease is low. Even if screening is “effective” in detecting disease in a phase at which it is still curable, a large number of subjects must be evaluated, and a substantial number of those must be treated to save one life from cancer. Although modern diagnostic tests and even cancer operations appear to be remarkably safe, rare fatal complications do occur, and more subtle effects to hasten death through cardiovascular or other causes may be missed completely. Thus, there may be a fine balance between benefit and harm from screening.

Disease-specific mortality rather than all-cause mortality has been the accepted end point of screening studies because fewer patients are required to provide adequate power. It has been assumed that disease-specific mortality is a good surrogate end point for all-cause mortality, but the current study raises serious doubts about the validity of this assumption. Whether measured directly or not, a decrease in all-cause mortality should be the ultimate aim of screening programs. A death from a nonmalignant cause is just as important as a cancer-related death. What is the feasibility of all-cause mortality as a primary end point? Seven of the 12 studies assessed in the article by Black et al. (1) were of screening mammography for breast cancer. Because screening mammography is widely available in Western countries, performing a randomized study without substantial contamination in the observed arm may now be impossible. Therefore, careful attention needs to be paid to the available data. Some caution should be exercised in interpreting all-cause mortality in studies that were not prospectively designed to address this end point. Thus, as acknowledged by the authors, some of the inconsistencies observed may be the result of chance alone. Pooling of the available data in a meta-analysis may overcome this problem. However, in the controversial article published by Gotzsche and Olsen (2), only two of eight published randomized trials of screening mammography were thought to be of sufficient quality to include in a meta-analysis. Neither of these studies demonstrated a benefit for screening mammography, leading the authors to conclude that there is no evidence that breast cancer screening reduces mortality. This article has sparked intense debate both in the lay press and in scientific literature. One of the criticisms was the authors’ emphasis on all-cause mortality as an end point (3), but the article by Black et al. (1) adds substantial credence to that approach.

The methodologic lessons provided by the analysis of Black et al. (1) should be applied to future trials. One recommendation of the authors is that future screening studies concentrate on high-risk groups. This recommendation has the advantage of requiring a smaller number of patients to detect statistically significant differences between study groups. Unfortunately, studies of high-risk patients may not be applicable to the general population because of differences in patient characteristics, the biology of disease, and the level of risk. For example, women at high risk of breast cancer, such as known carriers of BRCA1 and BRCA2 gene mutations, are at risk of developing the disease from a young age. The reduced sensitivity of mammography in premenopausal women is a well-documented problem (4,5), and the behavior of these cancers may be different. Therefore, even a well-conducted randomized study of screening for breast cancer in such high-risk women with all-cause mortality as an end point will not answer the question of whether screening is beneficial to the general population. Screening programs should be implemented only in the populations in which they have shown benefit in a clinical trial.

The concept of slippery-linkage bias applies equally to pre-
vention trials in which mortality related to unexpected complications of the intervention may go unrecognized if all-cause mortality is not reported. For example, the National Surgical Adjuvant Breast and Bowel Project (NSABP) P-1 study has led to the widespread introduction of the use of tamoxifen for prevention of breast cancer in the United States, based on a demonstrated reduction in the incidence of breast cancer (6). This event is an even more proximal surrogate end point than disease-specific mortality, which was not different between the two arms of the NSABP study. Proponents argue that a decrease in incidence will translate into a difference in disease-specific survival, which in turn will translate into a difference in all-cause survival. However, tamoxifen can cause harm (a low incidence of endometrial cancer and thromboembolic disease) as well as benefit, and the above reasoning seems to be a tenuous justification for its use in well women. Both the Study of Tamoxifen and Raloxifene for the Prevention of Breast Cancer and the Randomized Study of Selenium and Vitamin E for the Prevention of Prostate Cancer have incidence rates of the respective cancers as their primary endpoints. Even if these large and expensive trials demonstrate a decrease in cancer incidence, this finding is not necessarily proof of lives saved.

The biases demonstrated in the article by Black et al. (1) may make it almost impossible to demonstrate (or rule out) the value of screening for prostate cancer. The potential for misdiagnosis of the cause of death and mortality as a direct or indirect result of procedures resulting from screening will be particularly high in a screened population of elderly men. Because of competing causes of death, the influence of cancer screening on all-cause mortality will be difficult to detect, whereas cause-specific mortality will be subject to these types of bias. The International Prostate Screening Trials Evaluation Group is a collaboration of multiple centers conducting randomized trials of prostate cancer screening. It is hoped that the pooled sample will be about 200,000 men, with a follow-up period of at least 10 years. Even with this large sample and long follow-up, the primary end point is mortality from prostate cancer.

Population-based screening trials that are designed with improvement in all-cause mortality as the primary end point will require very large numbers of patients, lengthy follow-up, and great expense. However, we cannot justify implementation of screening programs that are costly to the individual and to the community if we are uncertain of their true benefit.

REFERENCES