The Natural History of Breast Cancer

Despite the appeal of early detection of breast cancer, uncertainty about the value of mammography continues. In this issue of the Archives, Zahl et al use a clever study design in an attempt to estimate the value of screening. The initiation of a public screening program in Norway allowed them to compare biopsy-confirmed invasive tumors in women receiving a single mammogram between the ages of 50 and 64 years, with the cumulative number of tumors in a group of women aged 50 to 64 years who had been screened on 3 occasions. Because a variety of risk factors were similar for the 2 groups, the cumulative tumor rate in the multiple screen group was expected to be the same as the rate in the age-matched single screen group. However, the rate in the single screen group was about 22% lower.

See also page 2311

How can we explain this 22% difference? The authors argue persuasively that the results could not be accounted for by ascertainment bias, differential risk, changes in the sensitivity of mammography, or historical changes in breast cancer incidence. They also argued that the large difference they observed was not likely to have resulted from changes in the use of hormone therapy. Their favored explanation is that some tumors that may have been detected at earlier ages in the single screen group spontaneously regressed prior to the screen at age 50 to 64 years. The suggestion is quite provocative, but is it credible?

Others have described cases of spontaneous remission of breast cancer, but in most reports such cases are dismissed as extremely rare. However, it is well known that the course of malignant diseases may vary considerably. The morphologic diagnosis is a poor predictor of the course of disease. Prostate cancer is a well-known example. Most prostate tumors grow slowly, do not produce metastases, and do not affect the life of their hosts. Some cases of prostate cancer grow fast, metastasize in several organs, and impair the patient’s quality and length of life. Similar biological variation has been shown for other types of cancer such as breast cancer, B-cell malignant neoplasms, and cancer of the urinary bladder. Not many years ago it was commonly believed that most patients with untreated breast cancer eventually die of breast cancer. Today there is increasing evidence that a considerable proportion of breast cancer cases are not life threatening to their hosts.

Because the study by Zahl et al was not a randomized clinical trial, methodological concerns may lead to other explanations for these findings. One possibility is that the larger number of mammograms in the multiple screen group could account for the differences. We know, for example, that between 20% and 30% of visible lesions are overlooked. Studies suggest that detection rates are higher if films are reviewed by multiple radiologists. One study showed that among 108 radiologists, there was a range of 40% in the sensitivity for detecting breast lesions. Women who have had 3 consecutive mammograms may be up to 20% more likely to have a positive result on 1 of the 3 tests. However, the study by Zahl et al included 1 additional screen in each group. If the multiple vs single screen explanation is correct, we should have seen a narrowing of the difference between the 2 groups following the additional screen. That did not occur. Furthermore, if we assume that tumors missed in early screens continue to progress, they should have showed up in the tumor registry. They did not. The design of the study has many imperfections, but we should not overlook its strengths. It was population based, it had very high participation, and the outcomes were well documented in an independent tumor registry. Considering the strengths and weaknesses of the methodology, the findings should not be dismissed.

Another reason not to disregard the findings of Zahl et al is that they are consistent with several observations that have troubled investigators for years. For example, randomized clinical trials rarely show the benefits of screening, particularly for women younger than 50 years. The Canadian National Breast Screening Trial hinted of spontaneous regression, and the Wisconsin Breast Cancer Epidemiology Simulation Model required a concept like spontaneous regression to account for observational data. Although the findings of Zahl et al seem counterintuitive, the spontaneous regression hypothesis is difficult to rule out.

Clearly, these findings require further investigation. If the spontaneous regression hypothesis is correct, we would expect no difference in age-adjusted breast cancer mortality between the 2 groups. The hypothesis suggests that about 20% of the women in the multiple screen group received unnecessary treatment because their tumors would have disappeared on their own. The alternative hypothesis is that some women in the single screen group missed the opportunity for early intervention. Although mortality data were not offered in the article by Zahl et al, a finding of no difference might support their suggestion, whereas a difference favoring higher sur-
vival in the multiple screen group might refute it. We are not aware of other data suggesting a survival advantage as high as 20% that is attributable to different screening strategies. Thoughtful investigators will find other ways to evaluate hypotheses derived from the observations.

Perhaps the most important concern raised by the study by Zahl et al\(^2\) is that it highlights how surprisingly little we know about what happens to untreated patients with breast cancer. In addition to not knowing the natural history of breast cancer for younger women, we also know very little about the natural history for older women. We know from autopsy studies that a significant number of women die without knowing that they had breast cancer (including ductal carcinoma in situ).\(^3\) The observation of a historical trend toward improved survival\(^4\) does not necessarily support the benefit of treatment. Improved survival is equally well explained by lead-time bias.\(^5\) Widely cited studies describing the dire consequences of refusing treatment may not be entirely credible. In the best known of these, Verkooijen and colleagues\(^2\) have shown great creativity, and we hope other investigators will find other ways to assess the hypothesis will be difficult. Many will evaluate the study by Zahl et al\(^2\) using criteria applied to randomized clinical trials, and they will find features of the design not to their liking. The best scientific study of the natural history of breast cancer would require that some women go without treatment. Patient and physician preference assure that most women with positive mammograms will accept treatment.\(^1\) However, the groups were in no way comparable. Those who refused were 10 years older, more likely to be of lower income, single, and at more advanced stages of disease. Larsen and Rose\(^6\) reviewed the literature going back to the first part of the last century and concluded that spontaneous remission of breast cancer occurs but only rarely.

If the spontaneous remission hypothesis is credible, it should cause a major reevaluation in the approach to breast cancer research and treatment. Certainly it is worthy of further evaluation. But, finding better data to assess the hypothesis will be difficult. Many will evaluate the study by Zahl et al\(^2\) using criteria applied to randomized clinical trials, and they will find features of the design not to their liking. The best scientific study of the natural history of breast cancer would require that some women go without treatment. Patient and physician preference assure that most women with positive mammograms will accept treatment.

Although scientifically necessary, a placebo-controlled randomized clinical trial is unlikely. At present, most would consider a trial unethical. However, the data of Zahl et al\(^2\) have raised an important challenge. Ethical concerns about age-based criteria for screening are frequently based on preferences, but not necessarily on evidence.\(^7\) Preferences influence the acceptance of randomization and have therefore a considerable impact on what we call “state of the art” (Porzsolt et al, unpublished data, 2008).\(^8\) We must also consider the ethical concerns associated with overdagnosis and overtreatment. To address these issues, we need to encourage thoughtful alternative experimental designs. Zahl and colleagues\(^2\) have shown great creativity, and we hope others will follow their lead. Although mammography screening is common in the United States, other countries are only now phasing in national programs. The availability of massive data from electronic medical records may also offer possibilities for other noneperimental comparisons. Even without a randomized controlled trial, funding agencies should encourage creative quasi-experiments that can help evaluate the spontaneous remission hypothesis.

Robert M. Kaplan, PhD
Franz Porzsolt, MD, PhD

Correspondence: Dr Kaplan, Health Services, School of Public Health, UCLA, CHS 31-293-C, Los Angeles, CA 90095 (rmkkaplan@ucla.edu). Dr Porzsolt, Clinical Economics University of Ulm, Frauensteige 6, D-89073 Ulm, Germany (franz.porzsolt@uniklinik-ulm.de)

Financial Disclosure: None reported.

REFERENCES