Breast Cancer Screening Comes Full Circle

Michael Baum

During the mid-1990s, I was invited to address the local chapter of the American Cancer Society in Miami, Florida, to talk about the British approach to mammographic screening. During my presentation, I described the tough issues of balancing the benefits for the few versus the harms for the many, and I suggested that maybe screening does not benefit the premenopausal woman at all. Despite my role in establishing the National Screening Programme when I was Chief of Surgery at King’s College London in 1988, my comments were not well received, and, as the audience stormed out on me in a paroxysm of pique, I learned a painful lesson that day that some topics, particularly breast cancer screening, do not lend themselves to polite and rational scientific debate.

However, I believe that a scientific debate is highly warranted because screening for any disease is notorious for the artifacts it throws up that make interpretations of apparent benefit extremely difficult. In particular, there are at least three screening artifacts or biases applying to breast cancer screening by mammography.

1) Lead-Time Bias: A time shift in the detection of a breast cancer “to the left” will mean that screen-detected cancers are smaller than clinically detected cancers. This shift will artifactually increase survival from the time of diagnosis to the time of death, without necessarily influencing the natural history of the disease. Despite awareness of lead-time bias, the results of screening programs are reported regularly as triumphant, with stunning 5-year survival data (1). The estimates of lead-time bias vary between 1 and 4 years but they can never be accurate, given the effect of length bias (2).

2) Length Bias: Whatever the period between screens, the fastest-growing tumors will surface as interval cancers (3). These interval cancers are often described as “advanced” because of size alone, but it is important to remember that size is a function of both biology and chronology. Put another way, screening is particularly good at detecting cancers with more favorable natural history. If some of these so-called “good” cancers have a natural history that is longer than the woman’s lifetime, should she die of an unrelated disease, these “good cancers” might be called “pseudo cancers” as suggested by Gilbert Welch (4). The existence of pseudo cancers would lead indirectly to a gross underestimate of lead-time bias because they represent the tail of an asymmetric distribution curve.

3) Class Bias: Women from the higher socioeconomic and educational strata are more likely to accept the invitation to screen and to participate in other health promoting patterns of behavior. Currently, it is known that social class alone, even after controlling for stage of disease, is a powerful prognostic indicator (5). No amount of public promotion to attend screening will alter this variable. Thus, the more women of lower socioeconomic and educational strata that are convinced to seek out screening, the more diluted the overall effect of screening.

Because of these biases, most methodologists accept that the only way to reliably investigate the benefits of screening is to compare breast cancer mortality in two populations randomly allocated to be screened or not to be screened. All other approaches are surrogates that fail to avoid the triple trap of the biases described above.

This long preamble leads me to the article by Taplin and his colleagues in this issue of the JNCI (6). Using data from the 3 years preceding a breast cancer diagnosis for participants in several large integrated health care plans, Taplin et al. sought to understand, in a case–control study, whether the occurrence of late-stage breast cancers was associated with a failure in mammography screening implementation. Screening implementation was divided into three groups: absence of screening, absence of detection, or potential breakdown in follow-up. The authors found that the distribution among the three implementation groups differed by case–control status and that women in the absence-of-screening group had higher odds of having advanced disease than women in the other two categories. A variety of demographic characteristics associated with lower socioeconomic status were associated with women in the absence-of-screening group. The authors conclude that, to reduce late-stage breast cancer occurrence, reaching women in the absence-of-screening group should be a top priority for screening implementation.

I would like to propose, however, another interpretation of the article by Taplin et al. (6). The authors designed their case–control study such that case subjects were women with advanced breast cancer and the matched control subjects were women with early-stage disease. However, the authors used an unusual definition of advanced breast cancer: cancer greater than 3 cm in size with or without evidence of metastases at the time of diagnosis. This definition does not conform to any staging system of which I’m aware. The case and control subjects were then allocated into one of three bins: absence of screening, absence of detection, and breakdown of follow-up after a positive test. The last category contained very few subjects, some of whom might simply have reflected procrastination on the part of physician or patient after a borderline finding, so I will not consider them further.

The first bin (absence of screening) reflects a noncompliant population who, by definition, will be subject to lead-time and class biases. It is, therefore, hardly surprising to find that this population contains more women with advanced tumors (i.e., greater than 3 cm in diameter) and more women that belong to a lower socioeconomic and educational stratum of society. These “defaulting women” will also include women over the age of 75 years—women who tend not to be invited as vigorously to participate in screening because of a lack of evidence of efficacy for this age group.

The next bin (absence of detection) reflects the compliant population who obtained screening and are thus, by definition,
susceptible to length bias. The underlying assumption for delineating “absence of detection” as a category is that the women in the category had something that could be detected. However, as far as I can determine, the women in this category are those who had normal mammograms and then went on to develop cancer before their next screening round. In my view, these cancers are interval cancers. As mentioned earlier, interval cancers are subject to length bias and thus will inevitably be more often tumors of greater than 3 cm for biological rather than chronological reasons.

Taplin et al. (6) conclude that the findings of their study should trigger amplified efforts in public education to persuade more women, especially those of lower socioeconomic, educational level, and older women, to get screened regularly. To me, this is a form of social engineering, i.e., applying public pressure to behave in a manner deemed responsible by the agents of the State or putting it another way, “nanny knows best.” Ethically, I believe that encouraging women to make informed choices is more supportable than social engineering. Indeed, one helpful way for women to arrive at an informed choice is to encourage them to read the book Should I Be Tested for Cancer? (4).

Biologically, social engineering has limitations because in the same way a leopard cannot change its spots, you cannot change the incidence of interval cancers even if all women got regular mammograms. Furthermore, the decision to attend screening does not in itself change socioeconomic status. This fact is important because the detection of more tumors that are less than 3 cm in diameter might be a necessary but not sufficient condition for the reduction in breast cancer mortality. Even if smaller tumors are detected in women from lower socioeconomic strata, these women will still be subject to the, as yet undefined, biologic disadvantage associated with their social status. Consequently, efforts directed at improving the socioeconomic status of women in lower strata might indirectly have a greater effect on reducing breast cancer mortality than efforts directed at attaining universal mammography.

So, is the article by Taplin et al. (6) one elaborate tautology proving yet again that screen-detected cancers are good, interval cancers are bad, and that “noncompliant” women tend to have bad disease? I think so but I could be wrong.

REFERENCES