

Drs. Dupont and Baines each read the commentary of Drs. Kopans, Halpern, and Hulka before writing their own commentaries; thereafter, Drs. Kopans, Halpern, and Hulka read the commentaries and responded to Drs. Dupont and Baines.

Statistical Power in Breast Cancer Screening Trials and Mortality Reduction among Women 40–49 Years of Age with Particular Emphasis on the National Breast Screening Study of Canada

Daniel B. Kopans, M.D., Elkan Halpern, Ph.D., and Carol A. Hulka, M.D.

Breast cancer remains the leading cause of nonpreventable cancer death among women in the United States. The mortality (deaths per 100,000 women in the population) from breast cancer has remained relatively unchanged for the last 50 years¹ despite advances in treatment, such as chemotherapy and hormonal manipulation.² Studies have shown that screening asymptomatic women for breast cancer can reduce the size and stage of these malignancies,³ and this has been shown to result in prolongation of survival. Most experts will not accept survival studies as proof of the benefits of screening because of the possible biases associated with survival analysis. A longer apparent survival could result from earlier detection with no ultimate effect on death. This awareness of the presence of the disease over a longer period of time with no effect on mortality is termed lead-time bias. Periodic screening is more likely to detect slow growing cancers, and this is termed length-biased sampling. Selection bias can occur for various reasons. Among them is the fact that women using health care services tend to be more healthy. Another selection bias may come from doctors who tend to

select what they think is the "best" treatment for each patient.

It is generally believed that proof of benefit can only come from randomized, controlled clinical trials (RCTs) in which death from breast cancer is the measured endpoint. Although selection bias can compromise the generalizability of the results, the aforementioned biases can be eliminated by randomly assigning women to be screened or to act as unscreened control subjects, and using death from breast cancer as the outcome for comparison. The assignment process should be blinded to avoid inadvertent compromise of the random allocation.

Randomized, controlled clinical trials have shown a clear benefit for screening women 50 years of age and older that is almost universally accepted.^{4,5} Controversy has arisen when separate proof has been sought for a screening benefit for women aged 40–49 years. The National Breast Screening Study of Canada (NBSS) is the only trial that was prospectively intended to evaluate women in this decade of life. All the other data have come from RCT's that were designed to evaluate a larger range of ages and have been retrospectively stratified and analyzed by age. Because of weaknesses in the statistical power of this retrospective subgroup analysis and the recognition that the NBSS was not properly designed and performed, controversy has arisen concerning the interpretation of the results.^{6–8}

One of the key elements in assessing the validity of the results from RCTs is related to the power of the trial

From the Department of Radiology, Massachusetts General Hospital, and the Harvard Medical School, Boston, Massachusetts.

Address for reprints: Daniel B. Kopans, M.D., Department of Radiology, Harvard Medical School, Massachusetts General Hospital, Boston, MA 02114.

Received January 26, 1994; revisions received March 8, 1994 and April 26, 1994; accepted April 29, 1994.

to demonstrate a statistically significant result. In addition to the magnitude of the true difference in treatment effects, power is directly associated with the size of the study (sample size). A trial cannot be expected to produce statistically significant results if the numbers of women involved cannot produce sufficient numbers of "events" in the control group (deaths from breast cancer) for the screen to show a benefit. As stated by Lachin in 1981, "Clinical trials with inadequate sample size are thus doomed to failure before they begin and serve only to confuse the issue of determining the most effective therapy for a given condition."⁹

The mere fact that a trial is randomized with control subjects does not guarantee its validity. To be valid measures of the benefit of screening, RCTs must be carefully designed to answer specific questions. The test parameters and performance must be appropriate for the group being studied. If subgroups are to be analyzed, this must be anticipated and accounted for at the start to ensure that there are sufficient numbers of women in the subgroup to provide sufficient power. There are several ways that an error can occur. An apparent benefit may appear through statistical fluctuation when there is actually no benefit (false-positive), or a clinically significant benefit (mortality reduction) may be overlooked (false-negative). If the study is too small, the chance of overlooking a clinically significant benefit becomes appreciable (i.e., the power is low).

The RCTs were not properly designed or executed to evaluate women aged 40–49 years as a separate subgroup. Through retrospective subgroup analysis, the trial data are being used to answer questions that the trials were not designed to answer. It is reasonable to perform subgroup analyses in an effort to identify future areas for research, but the use of statistically limited data for defining health care guidelines is questionable.¹⁰ The results of any scientific investigation cannot be evaluated out of context and without regard for the design and performance of the trials whose data are being evaluated. It is not merely the fact that a trial is randomized with control subjects that legitimizes its results. The design and performance of the trial have a significant impact on the validity of the results, and these cannot be overlooked.

Any benefit from screening is ultimately measured statistically. Factors that influence the statistical validity of the trial must be accounted for in the design and performance of the trial. The allocation of individuals to be screened or to act as control subjects must be done in a blinded fashion so that there is no possibility for selection bias to enter into the trial. The number of women in the trial is crucial for the results to have statistical validity. In a trial of breast cancer screening where mortality is the measure, there must be sufficient numbers

of women in the arms of the trial to produce a sufficient number of deaths among the control group to satisfy the statistical requirements of significance for an anticipated benefit.

To better understand the usefulness of the data from the RCTs that have been performed to date and to better assess the ability of these trials to provide data on the benefit from screening women 40–49 years of age, we calculated the sample size that would be needed to provide the ability to demonstrate a statistically significant benefit for women at these ages. In an effort to place the available data in the proper perspective, we performed the power calculation that is at the heart of every controlled clinical trial.

The Power Calculation

The power in a screening trial is the likelihood that for a given sample size, the study will be sufficiently large to be able to detect an anticipated difference in the outcomes between the screens and control subjects such that the difference will be statistically significant. Because these are statistical estimates, the power calculation can be made using varying combinations of parameters describing the study design and magnitude of the expected difference. Power is usually expressed as the trial having an 80% chance of revealing the expected difference at a statistical level of significance of 0.05 after a predefined period of follow-up. If the difference is greater than what was expected, the power is even greater. If the real difference is less than what was expected, then the power is less than 80%. The calculation is used to ascertain the total number of women needed to participate in the trial for the study to have an 80% power to demonstrate the expected benefit.

Several assumptions must be made in performing the power calculation for cancer screening studies.

1. What is the expected prevalence and incidence of cancer in the population to be studied?
2. How many deaths from these cancers are expected in the control (unscreened) group against which the screened group is to be compared?
3. What is the true benefit of screening in terms of the percentage reduction in deaths due to breast cancer? The lower the benefit, the greater the number of women needed in the trial to show that this benefit is statistically significant. In general, the sample size determined by the power calculation provides the number of women needed to demonstrate the anticipated benefit. A trial is considered to be adequate only if the real benefit is equal to or larger than the one specified in the power calculation. Conversely, trials with sample sizes appreciably

smaller than that determined by the power calculation are not considered to be adequate trials of the anticipated benefit.

4. What percentage of women assigned to the study group refuse to be screened, violate the protocol, or are lost to follow-up (noncompliers)? The higher the noncompliance, the greater the number of women needed in the trial.
5. What percentage of the unscreened control subjects obtain mammography outside the trial ("cross-over" that produces contamination of the control group)? The higher the rate of contamination, the greater the number of women needed in the trial to compensate for the loss of statistical power.

Each of these factors influences the number of women needed in the trial to have the conventional 80% chance of demonstrating the expected benefit. Problems with any of the factors can compromise the trial and reduce the power to demonstrate a real benefit.

There is a sixth factor which must be considered in assessing the validity of the trial.

6. What portion of the population do the studied women represent, and how "transferable" will the results of the trial be to the general population.

This does not affect the power of the trial but, rather, its external validity.

Sample Size Calculation

We chose to calculate the sample size needed to have an 80% chance (power) of detecting a 25% mortality reduction among women 40–49 years of age at the 5% level of significance using a two-tailed test. The two-tailed test was chosen because some, such as the investigators in the NBSS have raised concern that screening these women may result in harm. This can only be ascertained using the two-tailed test.

A mortality reduction of 25% is likely in this age group, based on the results of the Health Insurance Plan of New York (HIP) study.⁴ This differs from the 40% used in the NBSS calculation, which was chosen because the investigators believed they could not mount a larger trial that would have permitted the demonstration of a smaller benefit. The NBSS investigators have stated that their "sample size [was] chosen on the basis of feasibility."¹¹ They could not have mounted a larger trial. Their small sample size diluted the power of the trial. In our model, we assumed that each of the study women were screened each year for 5 years and that we would want to see a benefit when we analyzed the results having followed each woman for 5 years after

her last screen or from the time of breast cancer detection. This also differs from the NBSS, which appears to have followed women only until the end of the 10th year of the study. As a result, women who were enrolled after the first year had less than 5 years' follow-up from their last screen. Based on the results of the other trials, 5 years after the first screen is likely too soon to expect a benefit to appear for these women, but we chose to do the analysis using 5 years' of follow-up because this is the time that was chosen for analysis of the NBSS¹² (although they delayed their analysis because of the lower-than-expected number of deaths among the control women).

Based on data from the National Cancer Institute's Surveillance, Epidemiology, and End Results program,¹³ we assumed that the unscreened control women, who had breast cancer, would have an 80% survival at 5 years after detection (for purposes of a sensitivity analysis, we let the assumed mortality rate range from 10 to 30%). In addition, we assumed that 20% of the control women would have mammograms (contamination) despite agreeing to participate (in reality, this figure has ranged as high as 35% in the Malmö trial¹⁴ and was 26% in the NBSS). This was used to reduce the assumed mortality in the control group. We did not adjust [increase] the rate in the screened group for either imperfect compliance (women who don't return for all the screens) or refusal to be screened (noncompliance). We also assumed that 85% (70–100% for sensitivity) of the women invited for screening would participate and not be lost to follow-up. Although, in conventional trials, patients are not enrolled until they agree to participate, such as in the NBSS, we included the possibility of these losses, because in the HIP and Swedish trials women were randomized and then invited to participate.

Using the "Chi-Square with Yates Correction" in a power analysis program¹⁵ and the above assumptions, we estimate that 1440 cases of breast cancer would be required in each arm of the study to have an 80% power to detect a 25% reduction in the number of deaths due to breast cancer within the period of initiation of screening to the end of follow-up (5 years after the last screen). The number required per arm in the sensitivity analysis ranges from a low of 699 (30% mortality among unscreened women with only a 10% contamination) to a high of 4280 (10% mortality among unscreened women and 39% contamination). The sample size is more sensitive to the assumed mortality than to the extent of contamination.

To determine the total number of women who must be entered into the study, we assumed that there would be 7.2 cases of cancer (5.4–9.0 for sensitivity) over the 5 years of screening per 1000 women completing the

study (prevalence plus 4 years of incremental incidence).

By simple extrapolation, to attain 1440 women with breast cancer in each arm to have an 80% power to detect a 25% mortality benefit at the 5% level would require entering 235,294 women in each arm (479,588 women in all with 200,000 women in each arm completing the study). The sensitivity analysis reveals that this could be as low as 155,334 per arm if the mortality of the control subjects at 5 years was 30% with only a 10% contamination, and 9.0 cancers per 1000 women and no women lost to follow-up. The upper limit would be 2,264,550.

Discussion

The publication of the early results of the Canadian National Breast Screening Study (NBSS) has focused attention on whether or not there is a benefit from using mammography to screen women who are 40–49 years of age for breast cancer. Opponents of screening these younger women, even those who support screening women 50 and over, suggest that multiple studies have failed to prove a benefit among women 40–49 years of age.^{16,17} As a consequence, the National Cancer Institute has withdrawn support for screening these women. What has been ignored and what our analysis shows is that there have been no studies that have been designed with the statistical power to be likely to show a statistically significant benefit. Even if all the eight trials are added together, the total number of women who participated, younger than 50 years of age (screens and control subjects), is less than 200,000. When contamination (up to 35%) and noncompliance rates (up to 30%) are factored in, the effective number of women is even smaller.

To demonstrate that an apparent benefit is not due to chance or to reduce the possibility of overlooking a true benefit, a sufficient number of individuals must participate, and there must be a sufficient number of deaths to have the statistical power to determine that a benefit, when present, is significant. It requires an even larger study to “prove” that there is no benefit.

When an RCT is designed, an estimate is made as to the expected number of women who will develop breast cancer within the period of the study, and, because mortality reduction is the goal, the number of anticipated deaths must be factored into the calculation. The anticipated ability of mammography to reduce deaths must be predicted, as well as the expected death rate among the unscreened control group. The number of unscreened control women who “cross over” and obtain screening must be estimated, and the sample size adjusted accordingly. Similarly, in trials such as the HIP

and those in Sweden, where populations were randomized before invitation, noncompliance must be accounted for, and the sample size increased accordingly. In the HIP, 30% of the women who were offered screening refused. Contamination and noncompliance dilute the statistical power of the trial. Because these women are counted as having participated in the groups to which they were randomized, they dilute the effect of screening, and the sample size must be increased to compensate for this dilution of the statistical power of the trial. We accounted for contamination and noncompliance by adjusting the anticipated incidence, prevalence, and detection probabilities, but the required sample size increase can be calculated by multiplying the base sample size (assuming no contamination and perfect compliance) by the factor $1/(1-r-c)^2$ where r is the rate of contamination, and c is the rate of noncompliance.⁹ All these dilutional effects must be considered in determining how many women are needed in the screened and control groups for the study to have the statistical power to demonstrate the expected benefit. If the sample size is too small, then only a larger-than-anticipated benefit can be shown.

The only RCT trial, thus far, that was intended at the outset to evaluate the efficacy of screening women in the specific age group of 40–49 years is the NBSS. The majority of the RCTs have been designed to evaluate screening benefit for a broader range of ages. For example, the HIP was designed to evaluate the benefit of screening women aged 40–64 years. The size of the population involved in the trial was predicated on the expected results, among this entire group, to prove a screening benefit for this range of ages. Analysts have tried to separately evaluate the benefit for women 40–49 years of age in the RCTs, through retrospective segregation by age. Because the studies were not designed for this, there are insufficient numbers of women in the younger populations to have the statistical power to prove the expected benefit even when it appears. As a consequence, although fewer women have died among the screened populations relative to the control subjects for these ages, the observed differences have not achieved the 5% level of significance. Statisticians argue that without sufficient study patients, it is not certain whether this apparent benefit is due to chance alone.¹⁸ All of the studies have had insufficient statistical power to be able to prove an expected benefit of approximately 25–30%.

The results of the NBSS have been used as an example of a large trial that fails to show any advantage,² but the power calculation in that trial was based on a 40% benefit.¹¹ This large a benefit, although possible, has never been suggested as being likely. The NBSS is further compromised by the fact that the quality of the

test being measured, mammography, has been shown to be poor. An independent review of the mammography showed that for the first 4 to 5 years, more than 50% of the mammograms were judged by independent review to be poor or unacceptable.^{19,20} In addition to the fact that two successive advisors to the program resigned because the quality issue was not being addressed during the trial,⁶ the NBSS's own physicist has acknowledged that the quality of the mammography was poor.²¹

There are additional design problems and statistical weaknesses in the NBSS. The study was designed without taking into account the 26% of unscreened control subjects who crossed over and had mammograms outside the program. These have been dismissed as being "diagnostic" studies. This explanation ignores the basic nature of mammography. In fact, all mammograms are screening studies because they offer the opportunity to discover clinically occult cancers. Contamination of the control group must be accounted for in the power estimates. Contamination of the control group is a problem that has occurred in most of the other RCTs. In the Malmö trial, 35% of the younger women in the control group had mammograms outside the trial.¹⁴ Proper study design requires that additional women be added to make up for the expected dilutional effect of this contamination or that the follow-up period be extended. If the NBSS had wished to maintain the same statistical power to be able to show a 40% or greater benefit, they would have needed a trial with almost 90,000 women aged 40–49 years, rather than the 50,000 women who actually participated. A recent review by Canadian epidemiologists agreed that the numerous problems with the trial should prevent it from being used to change the present screening guidelines.⁸

To a certain extent, following the population for a longer period of time (woman years) may compensate for an inadequate sample size, but increasing the number of years of follow-up by increasing the length of follow-up for each woman is not as effective as increasing the total by increasing the number of women enrolled. For example, the benefit of screening will be higher for 10,000 women screened each year for 5 years and followed for 5 years than for 5000 women screened for 5 years and followed for 10 years. The first group will have had 50,000 woman/years of screening, whereas the second will have had only 25,000 woman/years of screening. The main advantage of longer follow-up is to permit deaths from moderate growth tumors to occur in the control group to demonstrate the benefit for the screened group.

An additional problem with the statistical power of the NBSS that has been given little attention is that the NBSS calculations were based on an expected mortality

of 212 per 100,000 in the control group. In fact, the death rate appears to have been much lower than this figure, with only 111 deaths per 100,000 actually occurring in the control group. The 5-year survival rate for women 40–49 years of age in the NBSS control group was greater than 90%, a much higher figure than was expected. Fewer deaths among the control subjects further reduces the actual power of the study from what had been planned. The fact that the mortality rate for women 40–49 years of age in the unscreened control group was so low would have required almost no deaths in the screened population to prove a benefit.

The excellent survival among the control subjects raises, among others, questions about the generalizability of the results to the Canadian population. Women in Canada, aged 40–49 years, with breast cancer in 1980 had a 75–80% 5-year survival. It has been demonstrated that people who volunteer for a trial frequently have a better outcome than the general population (self-selection bias). It may be that the results of the NBSS only apply to women who volunteer to participate in an RCT. It is uncertain how applicable are the results for these self-selected volunteers when they are applied to the population of women who are referred for mammography by their physicians, as is the general situation in the United States.²² The better-than-expected survival among the control women may have been due to a transfer of women with advanced cancers from the control group to the screened group at the time of the unblinded allocation.

Evaluation of results after too short a follow-up period may also produce a false impression. Contrary to "conventional wisdom," women younger than 50 years of age actually live longer with breast cancer than do women older than 50 years of age.²³ For there to be a demonstrable benefit from screening, women with breast cancer in the control group must die during the period of study. Thus, the excellent survival among the control subjects may result in delay in the appearance of benefit for the screened women. This is likely the case in the HIP study, in which the benefit for women 40–49 years of age did not begin to appear until 7–8 years after screening began. A similar result has appeared in the Swedish screening programs, which, when evaluated together, revealed a 13% mortality reduction that began to appear at 8 years.²⁴

Evaluation of results too soon after the end of screening, such as appears to be the case in the NBSS, also diminishes the benefit of screening in two additional ways. It reduces the effect of benefit for cancers incident during the period of screening while emphasizing the outcome from those prevalent at initiation of screening. It also reduces the period of follow-up from the standard 5 years for incident cancers. If, for exam-

ple, it is assumed that all women were screened each year during a 5-year study (five screens in all), 5 years of follow-up after initial screening does not mean that 5-year survival statistics are applicable to all cancers in the studies. Many of the study cancers will have a much shorter period of risk. For example, suppose that there was a prevalence of 3.2 cancers per 1000 women at the first screen and an incidence of 1.0 cancer per 1000 women in each successive year. If benefit, in terms of reduced mortality, were to be evaluated 5 years after the initial screening of 100,000 women, there would be 720 cases of breast cancers in the study, but only some of them would be subject to a full 5 years of follow-up. There would be 100 incident cancers subject to the risk of death within 1 year of diagnosis, 100 with 2 years of risk, 100 with 3 years of risk, 100 with 4 years of risk, and only the 320 prevalent cancers would be subject to the usual standard 5 years of survival. Prevalent cancers bias the evaluation because they are less likely to benefit from discovery (because of the advanced nature of many prevalent cancers) than incident cancers. The effect is that the apparent benefit of screening is reduced by mixing prevalent cases with incident cases.

A delay in benefit may also occur in relationship to the interval between screens and the growth characteristics of cancers in the population. The longer the interval between screens, the greater the percentage of women whose cancers will grow to a level of lethality, at which point the tumor is incurable. Similarly, the faster the tumor growth, the more likely that short term cancer mortality will not be affected (the tumors reach the level of incurability before they are detected). If death from cancers with moderate growth rates can be affected, evidence of a benefit will be delayed because these cancers will not be as rapidly lethal among the control women as the higher growth rate lesions. Younger women appear to have cancers that have a shorter lead time for mammographic detection,^{25,26} and this may also account for delayed benefit.

It has been suggested that the trials consistently show a benefit for women 50 years of age and older but do not show a benefit for those 49 years of age and younger. What is absent from this analysis is the fact that the trials contained two to four times as many women 50 years of age and older as compared with the younger aged women. Moreover, the incidence of breast cancer is higher among older women, and the death rate is higher. Assuming the benefit is the same for the younger women as for the older women, the trials, a priori, would require many more younger women than older women to achieve the same level of significance. To the contrary, because the trials were not designed to evaluate younger women as a separate group,

the number of women in the younger age groups is markedly smaller than the older women.

Among the other problems that can occur in RCTs is the process of randomization. The remarkable survival experience of the control group in the NBSS (over 90% at 5 years) not only reduces the ability to demonstrate a benefit for screening but also brings into question the technique of randomization. There were far fewer deaths than expected among the control subjects (10.4 deaths/10,000 observed versus 13.7 expected deaths).

Women with advanced breast cancer, whose outcome cannot be influenced by screening, were permitted to participate in the NBSS. The fact that there was an excess of women with advanced cancers who were allocated to the screened group (prevalence year) and that this excess did not decrease over the period of screening, but actually increased, suggests the possibility that there was a shift of some lethal cancers from the control group to the screened group. Such a shift is possible because contrary to proper study design, the randomization was not blinded. Women who volunteered to participate in the NBSS were first given a clinical breast examination and then were randomized from an open available list. Although the investigators have been unable to find proof that the randomization was compromised, the fact that women with clinical symptoms of breast cancer were not excluded from the trials, and that all women were examined before randomization offered the opportunity to compromise the process.

It would be difficult to prove such a compromise because a shift of 10 to even 100 women from one group to the other would not affect the overall distribution of demographic characteristics, which appear, overall, to be equally distributed. Nevertheless, there were 33 women with lymph node-positive (advanced) breast cancer who were allocated to the screening group compared with only 21 among the control subjects. Thirty of the 33 women had palpable cancers at the time of allocation. Among these women, 19 had four or more positive axillary nodes (poor prognosis breast cancer), whereas only 5 women with four or more positive nodes were allocated to the control group. Among the 19 women with poor prognosis breast cancer allocated to screening, 17 of the 19 cancers were palpable at the time of allocation.

This allocation imbalance alone likely accounts for the fact that a benefit cannot be demonstrated in the early follow-up period. The fact that reviewers of these trials, who are relying on the results alone, have not reviewed the underpinnings of the trials is evident in the fact that Elwood et al., in their meta-analysis of the eight controlled clinical trials,¹⁰ clearly state in their publication, that women with known clinical symptoms

of breast cancer should be excluded from screening trials, but fail to acknowledge later in their analysis that women with clinical signs of breast cancer were permitted in the NBSS.

Our power calculations assume that the screening is done properly. If the screen is ineffective because of poor quality mammography, as in the NBSS, because there is too long an interval between screens as in the Swedish Two-County Trial, or because there is too high a diagnostic threshold for intervention, as in the Malmö study, then it may be impossible to detect a benefit, even by increasing the number of women in the trial or by following them for a longer period of time.

No single study has come close to the numbers of women needed to achieve statistical validity for women 40–49 years of age. The NBSS had a total of only 50,000 women (25,000 in each arm). Their estimates of cancer death were much higher than what actually occurred in the trial, and their power calculations were based on a higher benefit than should have been anticipated. Thus, they underestimated the number of women needed in the trial.

It is deceptive to suggest that scientific study has failed to show a benefit for the mammographic screening of women aged 40–49 years when the question has never been studied scientifically. The fact that there is a mortality reduction that has appeared for women aged 40–49 years in five out of the eight RCTs, which range from 22% in Edinburgh²⁷ to 49% in the Malmö trial,¹⁷ despite the fact that the trials are too small and were not optimized for women ages 40–49, suggests that there is a likely benefit from screening. A benefit can also be inferred from other data that suggest that there should be a benefit if mammography is performed properly and at the appropriate interval between screens. The Breast Cancer Detection Demonstration Project showed that there is little difference in the detection rates for women 40–49 years of age relative to those 50 years of age and older,²⁸ and this has been confirmed in more modern screening programs.²⁹ If absolute proof of benefit is required, then the appropriately designed and executed study should be undertaken. Until such a study is undertaken, the available data suggest that screening should be available for women 40–49 years of age. The only reason to cease screening these women is economic,³⁰ and this should be stated directly so that the true issue can be debated. There are ways that screening costs can be reduced significantly,^{31,32} and before support is withdrawn for screening these women, cost reduction methods should be used.

It is fallacious to suggest that the data show no benefit for younger women. It is clear that the trial data presently available are derived from study populations that are too small. Some have argued that if such a large

trial is needed to prove a benefit, then the benefit must be inconsequential. It would be inconsistent to suggest that a 25–30% mortality reduction for women aged 50–59 years is valuable, but the same reduction for women aged 40–49 years is not. In fact, at this time, almost the same number of lives would be saved in both groups because the number of cancers diagnosed among women aged 40–49 years is almost the same as for women aged 50–59 years.³³ In 1993, as a result of the “baby boomers,” there were 29,800 cancers diagnosed among women aged 40–49 years and only 8% more (31,500) among women aged 50–59 years.

A trial that could confidently prove no benefit would have to be even larger than one that could show a significant benefit. Women and their physicians should not be misled into believing that there have been adequate trials and that there is no scientific support for screening women aged 40–49 years. From a scientific perspective, the trials have not been adequate to evaluate these women separately. Women and physicians should be aware that if the trials are analyzed as they were designed (and the Canadian data are not included), there has been a statistically significant mortality reduction for women aged 40–74.

References

1. Breast Cancer. Cancer facts and figures—American Cancer Society 1994:10.
2. Early Breast Cancer Trialists' Collaborative Group. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. *Lancet* 1992; 339:1–15, 71–85.
3. Baker LH. Breast Cancer Detection Demonstration Project: five-year summary report. *Ca Cancer J Clin* 1982; 32:194–225.
4. Shapiro S, Venet W, Strax P, Venet L. Periodic screening for breast cancer: the health insurance plan project and its sequelae, 1963–1986. Baltimore (MD): The Johns Hopkins University Press, 1988.
5. Tabar L, Fagerberg CJG, Gad A, Baldetorp L, Holmberg LH, Grontoft O, et al. Reduction in mortality from breast cancer after mass screening with mammography. *Lancet* 1985; 1:829–32.
6. Kopans DB, Feig SA. The Canadian National Breast Screening Study: a critical review. *AJR Am J Roentgenol* 1993; 161:755–60.
7. Burhenne LJW, Burhenne HJ. The Canadian National Breast Screening Study: a Canadian critique. *AJR Am J Roentgenol* 1993; 161:761–3.
8. Boyd NF, Jong RA, Yaffe MJ, Trichler D, Lockwood G, Zylak CJ. A critical appraisal of the Canadian National Breast Cancer Screening Study. *Radiology* 1993; 189:661–3.
9. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. *Controlled Clin Trials* 1981; 2:93–113.
10. Elwood JM, Cox B, Richardson AK. The effectiveness of breast cancer screening by mammography in younger women. *Online J Curr Clin Trials* 1993; 32.
11. Miller AB, Howe GR, Wall C. The National Study of Breast Cancer Screening. *Clin Invest Med* 1981; 4:227–58.
12. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 1.: breast cancer detection and death rates among women aged 40–49. *Can Med Assoc J* 1992; 147:1459–76.

13. National Cancer Institute Statistics Review 1975-1988. Bethesda (MD): National Institutes of Health. 1991 NIH Pub. No.:91-2789.
14. Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Folke L, et al. Mammographic screening and mortality from breast cancer: The Malmo Mammographic Screening Trial. *BMJ* 1988; 297:943-9.
15. Borenstein M, Cohen J. Statistical power analysis. Hillsdale (NJ): Lawrence Erlbaum Associates, Inc., 1988.
16. Fletcher SW, Fletcher RH. The breast is close to the heart. *Ann Intern Med* 1992; 117:969-71.
17. Fletcher SW, Black W, Harris R, Rimer BK, Shapiro S. Report of the International Workshop on Screening for Breast Cancer. *J Natl Cancer Inst* 1993; 85:1644-56.
18. Chu KC, Smart CR, Tarone RE. Analysis of breast cancer mortality and stage distribution by age for the Health Insurance Plan Clinical Trial. *J Natl Cancer Inst* 1988; 80:1125-32.
19. Baines CJ, Miller AB, Kopans DB, Moskowitz M, Sanders DE, Sickles EA, et al. Canadian National Breast Screening Study: assessment of technical quality by external review. *AJR Am J Roentgenol* 1990; 155:743-7.
20. Kopans DB. The Canadian Screening Program: a Different perspective. *AJR Am J Roentgenol* 1990; 155:748-9.
21. Yaffe MJ. Correction: Canada Study [letter]. *J Natl Cancer Inst* 1993; 85:94.
22. Fox S, Klos DS, Tsou CV, Baum JK. Breast cancer screening recommendations: current status of women's knowledge. *Fam Community Health* 1987; 10:39-50.
23. Adami H, Malke B, Holmberg L, Persson I, Stone B. The relation between survival and age at diagnosis in breast cancer. *N Engl J Med* 1986; 315:559-63.
24. Nystrom L, Rutqvist LE, Wall S, Lindgren A, Lindqvist M, Ryden S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet* 1993; 341:973-8.
25. Moskowitz M. Breast cancer: age-specific growth rates and screening strategies. *Radiology* 1986; 161:37-41.
26. Tabar L, Faberberg G, Day NE, Holmberg L. What is the optimum interval between screening examinations?: an analysis based on the latest results of the Swedish Two-county Breast Cancer Screening Trial. *Br J Cancer* 1987; 55:547-51.
27. Eckhardt S, Badellino F, Murphy GP. UICC Meeting on Breast-Cancer Screening in Pre-Menopausal Women in Developed Countries. *Int J Cancer* 1994; 56:1-5.
28. Feig SA, Hendrick RE. Risk, benefit, and controversies in mammographic screening. A categorical Course in physics: technical aspects of breast imaging. RSNA Publications, 1992:103-18.
29. Sickles EA, Kopans DB. Deficiencies in the analysis of breast cancer screening data. *J Natl Cancer Inst* 1993; 85:1621-4.
30. Eddy DM, Hasselblad V, McGivney W, Hendee W. The value of mammography screening in women under 50 years. *JAMA* 1988; 259:1512-19.
31. Bird RE, McLelland R. How to initiate and operate a low-cost screening mammography center. *Radiology* 1986; 161:43-7.
32. Sickles EA, Weber WN, Galvin HB, Ominsky SH, Sollitto RA. Mammographic screening: how to operate successfully at low cost. *Radiology* 1986; 160:95-7.
33. Smith RA. Epidemiology of breast cancer. A categorical course in physics: technical aspects of breast imaging. 2nd ed. Oak Brook (IL):RSNA Publications, Radiological Society of North America, 1993:21-33.