Clinton Presidential Records

Graphic/Offensive Material Marker

This is not a presidential record. This is used as an administrative marker by the William J. Clinton Presidential Library Staff to identify graphic and/or offensive material that is not suitable for underage researchers. Record(s) located in this folder may be considered offensive by researchers.

PLEASE READ BEFORE PROCEEDING INTO FOLDER

HCFA

Department of Health & Human Services

FACT SHEET

Health Care Financing Administration

October 1994

MEDICARE COVERAGE OF MAMMOGRAMS

Screening Mammograms:

Since January 1, 1991, Medicare has been helping to pay for mammography screening for the early detection of breast cancer. **Screening mammograms**, together with a physician's interpretation of the results, are covered by Medicare once every two years for women aged 65 and over. For other women covered by Medicare, the test is covered by Medicare at varying frequency, according to the age and risk category of the patient. Medicare pays 80 percent of the lowest of the following for mammograms:

- The physician's actual charge,
- The amount calculated in the Medicare fee schedule for radiological services, or
- A fee limit. (The 1994 fee limit is set at $59.63, which means Medicare will pay a maximum of 80 percent of $59.63 or $47.70.)

Payment for Screening Mammograms:

Medicare will pay for screening mammograms for Medicare beneficiaries (women over age 65 and certain disabled women under age 65) as follows:

- Medicare will not pay for mammography screening procedures performed on women under age 35.
- For women over age 34 but under 40, Medicare will pay for only one screening mammogram.
- For women over age 39 but under 50 determined to be at high risk of developing breast cancer, and for women over age 49 but under 65, Medicare will pay for an annual mammography screening, i.e. payment may not be made for a screening mammogram performed within one year of a previous mammography screening.
- For women over age 39 but under 50 determined not to be at high risk and for women over aged 65, Medicare will pay for one screening mammogram every two years, i.e. payment may not be made for a mammography screening performed within two years of a previous mammography screening.

MORE

Payment for Diagnostic Mammograms:

Medicare also helps pay for **diagnostic mammograms** when the patient is being treated for cancer or shows symptoms of some abnormality (i.e., lumps, pain, or discoloration). Medicare will pay for an unlimited number of these medically necessary diagnostic mammograms.

Quality Standards:

The Mammography Quality Standards Act of 1992 went into effect on October 1, 1994 replacing previous rules which required facilities seeking payment for mammography screening to meet quality standards defined in Medicare regulations. To lawfully perform either screening or diagnostic mammography under the Mammography Quality Standards Act, facilities must obtain a certificate from the Food and Drug Administration by meeting quality standards for personnel, equipment, quality assurance and reporting established by the FDA. A facility must also pass a review of clinical images to be accredited by an FDA-approved body and successfully complete an annual inspection by the FDA or state personnel to retain its FDA certificate. HCFA has agreed to minimize duplication by accepting the FDA certificate as evidence that the facility is qualified for Medicare reimbursement for mammography.

Statistics:

- HCFA estimates that approximately 4.9 million Medicare beneficiaries are expected to receive screening and diagnostic mammograms in fiscal year 1994 at a cost to Medicare of about $240 million. In fiscal year 1995, about 5 million Medicare beneficiaries are expected to receive mammograms at a cost of about $260 million. In fiscal year 1996, about 5 million Medicare beneficiaries are expected to receive mammograms at a cost of about $270 million.

- The American Cancer Society estimates there were 182,000 new cases of invasive breast cancer in 1994 and that 90,100 of these cases were in women age 65 and older.

- The American Cancer Society estimates there will be 46,000 deaths in females from breast cancer in 1994 and that 25,700 of these deaths were in women age 65 and older.

For More Information:

For more information about breast cancer detection and mammography, call the toll-free telephone number of the National Cancer Institute, 1-800-4-CANCER.

For Medicare information, write to:

Office of Public Affairs
Health Care Financing Administration
Room 428-H Humphrey Building
200 Independence Avenue, S.W.
Washington, D.C. 20201

Clinton Presidential Records
Digital Records Marker

This is not a presidential record. This is used as an administrative marker by the William J. Clinton Presidential Library Staff.

This marker identifies the place of a publication.

Publications have not been scanned in their entirety for the purpose of digitization. To see the full publication please search online or visit the Clinton Presidential Library's Research Room.

The National Breast and Cervical Cancer Early Detection Program

AT-A-GLANCE

1993–1994

"There are pockets of disaster all over this country where poor, uninsured, and underinsured Americans are dying of cancer and other illnesses because of lack of information about and access to quality health-care services. Action must be taken immediately to correct the appalling health conditions reflected in these communities."

Dr. Harold P. Freeman, Chairman, President's Cancer Panel, Director of Surgery, Harlem Hospital Center, testifying before a congressional subcommittee, April 1990.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service

CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

PRINCIPLES & PRACTICE OF ONCOLOGY

PPO updates Volume 8

Daniel B. Kopans, M.D.

September 1994, Number 9

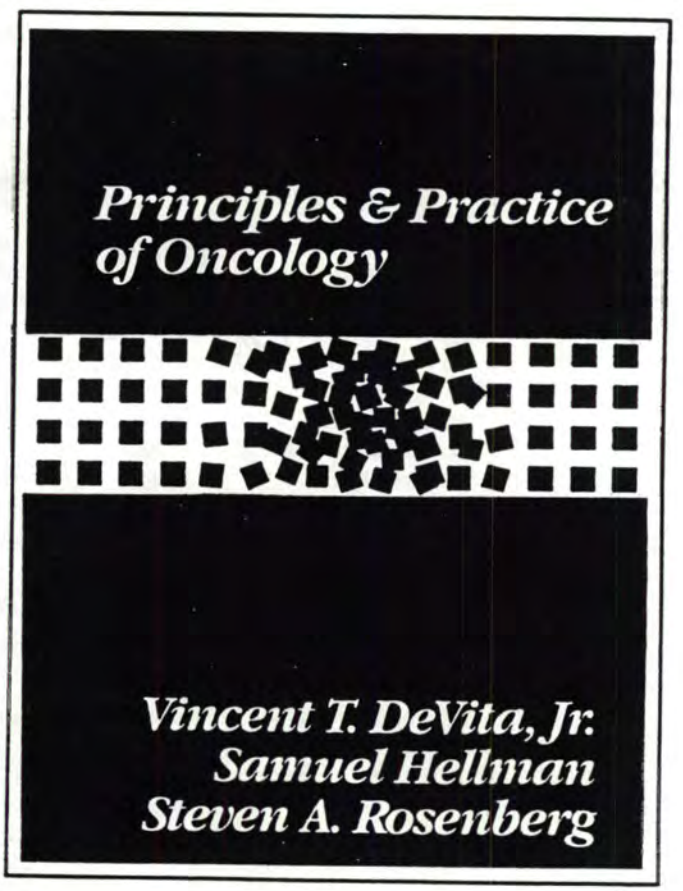
Breast Cancer Screening: Women 40 to 49 Years of Age

On the basis of data from the National Cancer Institute Surveillance, Epidemiology and End Results (SEER) program, the incidence of breast cancer has increased by approximately 1% each year since 1940.[1] In that year, approximately 60 women in every 100,000 were diagnosed with breast cancer. By 1983 the incidence had risen to approximately 95 women for every 100,000. During the same time period, the death rate from breast cancer has remained virtually unchanged at approximately 27 breast cancer deaths per 100,000 each year. Not only is the steady rise in breast cancer incidence over the past 40 to 50 years unexplained, but it is unclear whether the stable mortality in the face of the rising incidence reflects earlier detection and better treatment (lives saved), or merely the fact that more nonlethal cancers are being diagnosed each year. The inability to determine which cancers have lethal potential and which cancers may never cause death lies at the heart of the breast cancer screening debate.

Daniel B. Kopans, M.D. is an Associate Professor of Radiology at Harvard Medical School, Cambridge, Massachusetts, and Director of Breast Imaging, Massachusetts General Hospital, Boston, Massachusetts.

This update was co-edited by Dr. Rosenberg in his private capacity. No official support or endorsement by the National Institutes of Health is intended or should be inferred.

The opinions and/or clinical experiences outlined herein are those of the authors and do not necessarily represent the views of the sponsor.



SPONSORED BY AN EDUCATIONAL GRANT FROM
BRISTOL-MYERS ONCOLOGY DIVISION

DETERMINING BENEFIT

There is no question that screening, especially using mammography, can detect breast cancers at a smaller size and earlier stage than the usual medical care.[2,3] Fortunately, most women who are diagnosed with breast cancer do not die from it, and therefore the significance of earlier diagnosis has been questioned. The major difficulty lies in defining a method of analysis that can determine whether screening is truly beneficial. The most common way of evaluating a cancer intervention is to review the results of the patients' survival. Although this seems, superficially, to be straightforward, there are significant problems with merely analyzing survival data.

LEAD TIME BIAS

Some have argued that screening merely reveals the presence of a cancer earlier in its development but does not affect the ultimate day of death. Because "survival" is measured from the day of diagnosis to the day of death, survival can appear to be prolonged, when it is merely the date of detection that is pushed back, with the date of death unchanged. This "lead time" bias is the main reason why survival data are not accepted as "proof" of benefit for screening.

INDIVIDUAL SELECTION BIAS

Other factors can bias screening data. Women who voluntarily participate in a screening program may inadvertently select themselves, or be selected by their physicians, and have better-than-expected breast cancer experiences than the general population. An apparent, but spurious prolongation of survival can come from "selection" bias.

LENGTH BIAS SAMPLING

Periodic screening is more likely to detect slow-growing, presumably more indolent cancers than the faster-growing cancers, which are more likely to be lethal. If the time between screens is long (screening interval), faster-growing cancers are less likely to be detected at a screen (diagnosed earlier) but will "rise to the surface" between screens, and mortality will not be affected. If only the cancers detected by the screen are analyzed, "length bias sampling" might make the benefit appear greater than it is.

INTERMEDIATE MEASURES OF BENEFIT

Many have hoped that benefit could be determined through an analysis of "intermediate endpoints" and not require death from breast cancer as the only measure. Because cancer size and lymph node positivity have, repeatedly, been shown to be the best prognostic indicators,[3] it has been argued that benefit is shown by the fact that screening detects cancers at a smaller size and earlier stage.[4] Unfortunately, because there is not a perfect one-to-one correspondence between breast cancer deaths and these intermediate measures, or any other measurable factors, intermediate endpoints have not been accepted as proof of benefit. Most analysts will only accept a true alteration in the death rate from breast cancer as an acceptable measure.

RANDOMIZED, CONTROLLED TRIALS

Because of lead time bias, length bias sampling, and selection biases, survival data and intermediate endpoints have not been accepted as proof of the benefit of screening. The only way to eliminate these biases and measure benefit directly is through the randomized, controlled trial (RCT), which uses death from breast cancer as the measured endpoint. Because of the laws of probability, and chance fluctuations, benefit is measured statistically.

Randomized, controlled trials work by randomly dividing large numbers of women into two groups to try to produce two identical populations. If the numbers are large enough, and allocation is truly random, then the groups, overall, will have identical demography, will have the same number of women who develop breast cancer over time, and will have the same number of women who die of breast cancer over time. The benefit of screening can then be measured directly by offering one group screening and comparing subsequent deaths from breast cancer to those of the other group, which acts as an unscreened control. If there is no benefit from screening, the number of breast cancer deaths among the screened women will not differ from those among the control group each year. If screening actually saves some women from dying from breast cancer, there will be fewer deaths among the screened group.

Because chance may still play a role, it is commonly agreed that the difference in cancer deaths between the screened women and the controls must be statistically significant at the 95% level of confidence (the benefit is not due to chance) before most would accept the fact that screening reduces the rate of death from breast cancer.

Randomized, controlled trials eliminate the aforementioned biases. Lead time bias, for example, is eliminated in RCTs. Survival data are not a factor with death as the measured endpoint. If screening detects only cancers that are not destined to be lethal, then the death rate among the screened women will not be altered, and no benefit will be found, so the results of the RCTs are also independent of length bias sampling.

POPULATION SELECTION BIAS AND GENERALIZABILITY OF THE RESULTS

In order to be valid, screening trials must be designed and performed properly. The population of women chosen to participate must be considered to determine how the results relate to the general population. If the population is too

Published by J. B. Lippincott Company, 227 East Washington Square, Philadelphia, PA 19106, phone (215) 238-4398; fax (215) 238-4203. *Copyright © 1994 by J.B. Lippincott Company. Printed in the U.S.A. All rights reserved. Reproduction without specific permission is prohibited.* *ISSN 0892-0567.*

selected, then the results may not be applicable to a wider population of women. The ideal screening programs study a population of women that is representative of the entire population at risk. Most of the RCTs have first targeted populations without preconditions, randomly allocated the women in the population to one of the two groups, and then invited the women allocated to the screening group to be screened. This method provides the most representative study but runs the risk that women invited for screening may not wish to participate. This will dilute the power of the trial to show a benefit, because once an individual is allocated to a group, she is counted as being in that group whether or not she actually participated.

To ensure participation, a trial may ask for volunteers before allocation. This was the method used by the National Breast Screening Study (NBSS) of Canada, which recruited volunteers prior to randomization. This increased the participation, but the results of studies of volunteers may not reflect what would be expected in the general population. Randomization will avoid individual selection bias within the trial, but there may be selection bias in terms of the overall population.

THE RANDOMIZED, CONTROLLED TRIALS OF BREAST CANCER SCREENING

Eight randomized, controlled trials of breast cancer screening have been performed in three countries around the world since the early 1960s.

The Health Insurance Plan (HIP) of New York Trial

The Health Insurance Plan (HIP) of New York study was initiated in 1963[5] and randomly assigned a population of 62,000 women, ages 40 to 64, to two groups. Women in the study group were offered mammographic and physical examination screening every year for 4 years, whereas the control group was not even informed of the study and received only routine medical care. Only 65% of the women offered screening were screened at least once, and 88% of these attended at least two sessions.

THE SWEDISH SCREENING TRIALS

Five screening studies have been undertaken in Sweden.[6] The Swedish trials involved the use of mammography alone, without the benefit of physical examination. In three of the trials, only single-view mediolateral oblique mammography was used. The first two trials are known as the Two-County Trial. They were performed in the counties of Kopparberg and Ostergotland. Although they are actually separate trials, their data are frequently reported together.[7] Although the trials began as RCTs, when the data began to show a benefit from screening, the control groups were offered screening. Because the control groups are now being screened also, any benefit from screening will be difficult to discern from follow-up data.

The Swedish Two-County Trial

THE KOPPARBERG TRIAL. The Kopparberg trial began in 1977.[8] There were 38,562 women ages 40 to 74 who were randomly assigned to the screened group, and 18,478 who were allocated to the control group. Rather than randomizing women individually, randomization was accomplished by communities or "clusters" in which the women lived. Mammography was offered to the screened women ages 40 to 49 every 2 years, and to women ages 50 to 74 every 33 months. Among the women offered screening, 89% complied for at least the first screen. Single-view mammography was used.

THE OSTERGOTLAND TRIAL. The Ostergotland trial began in 1978 as the companion trial to Kopparberg. The trial design was similar, with women ages 40 to 74 randomized by community clusters. There were 38,405 women who were offered screening, and 37,145 who were chosen to act as unscreened controls. Single-view mammography was used every 2 years for women ages 40 to 49 and every 33 months for women ages 50 to 74, and there was 89% compliance with the first screening invitation.

The Malmö Trial

The Malmö trial was undertaken in Sweden beginning in 1976 and ending in 1986.[9] Approximately 42,000 women ages 45 to 69 were randomly divided, as individuals, into a group of 20,695 offered screening and a control group of 20,783. The interval between screens for the study group was 18 to 24 months (average, 21 months). For the first two screens, both mediolateral oblique and craniocaudal views were obtained. If the breasts were not radiographically dense, the subsequent screens included only a single-view oblique. Compliance with screening was 74% at the first round.

The Stockholm Trial

The Stockholm trial began in 1981.[10] The trial randomized 38,525 women ages 40 to 65 who were offered screening; 20,651 women in the same age range were randomized as unscreened controls. Randomization was by individual and based on a system using birth dates. Of the women invited to be screened, 82% complied with the first screen. The women were offered screening every 2 years with single-view oblique mammography. In the first screen, 53% of the cancers were palpable. Because the Two-County Trial was already showing a benefit, the program expected to screen all women in subsequent screens. This will dilute the benefit of the screen.

The Gothenburg Trial

The Gothenburg trial in Sweden was begun in 1982. The trial randomized 52,000 women ages 40 to 59. There were 20,724 women randomized and invited to be screened and

28,909 women randomized to be unscreened controls. Screening, using two-view mammography, was performed every 18 months, and 84% of the women invited to the first screen participated.

The Edinburgh Trial

With physicians' practices used as the basis for randomization, 45,130 women were divided into two groups—23,226 in the study arm and 21,904 in the control arm. The women involved were ages 45 to 64 at the time of entry into the program. Those in the screened group were offered mammography (two views) and a clinical breast examination (CBE) at the first screen. Mammography was performed every 2 years. Clinical examination was used as the screen in years 2, 4, and 6, and CBE was combined with single-view mammography (the mediolateral oblique) in years 3, 5, and 7.[11] The fact that only 60% of the women complied, and that the randomization by physician practice introduced socioeconomic differences, somewhat compromises the interpretation of the results of the trial.

The National Breast Screening Study(NBSS) of Canada

The National Breast Screening Study (NBSS) of Canada differs in several ways from the other trials. It was actually two separate trials organized by age. The women were all volunteers. All were given a CBE, and then they were randomized to have annual mammography and a CBE, or to act as controls. The randomization was not blinded in that all of the participants were given a CBE prior to allocation. The trial was such that all women (screens and controls) had at least one "screening" CBE. Furthermore, the fact that women who had symptoms of breast cancer were permitted in the study offered the opportunity to imbalance the trial by placing a larger percentage of women with advanced cancers in the screened group.[12]

One of the Canadian trials involved 50,430 women ages 40 to 49 who were randomized into two groups. There were 26,214 women who were screened with a CBE and mammography in the first year of the screening trial (diminishing to 14,146 by the fifth screen). These have been compared with 25,216 women who had just a CBE and then acted as "unscreened" controls over the 5-year trial. Screening consisted of two-view mammography (the straight lateral mammogram was used for much of the trial) and a CBE every year for 4 or 5 years.

The second part of the Canadian trial involved 39,405 women ages 50 to 59 who were divided into two groups. There were 19,711 women who participated in the first screen (dropping to 9576 by the fifth screen), and they were compared with 19,694 women in the control group. As in the under-50 trial, the screened women had two-view mammograms and a CBE every year. The control group, however, had a CBE each year.

THE RESULTS OF THE RANDOMIZED, CONTROLLED TRIALS

There has been considerable debate over how the data from the RCTs should be analyzed. For practical reasons (cost and logistics), none of the trials has been as large as would be desirable to accurately test the benefits of screening (see Statistical Power, below). Most have had marginal power to provide statistically significant results for mortality reductions of 30% or less for the populations the trial designs involved. This power has been further reduced by retrospective division of the data into subgroups that segregate the participants by age. It is this retrospective, unplanned subgroup analysis (producing groups with smaller numbers) that has caused much of the disagreement and confusion over the value of screening for women at different ages. Most would agree that unplanned "subgroup" analysis is useful for determining the next research question, but when the numbers become so small that the results cannot be statistically significant, these analyses should not be used to determine medical advice. When the trials were analyzed as they were designed, the following results were observed (Table 1).

Because of the small numbers of women involved in the individual trials, and the consequent wide confidence intervals, only the HIP trial and the Kopparberg trial have reached statistical significance individually. However, when all of the trials are combined in a meta-analysis (excluding the Canadian trial because of its differing study design), the overall results achieve statistical significance (at the 0.05 level) for a mortality reduction of 25% to 30%.[14] Because the trials were designed to evaluate women without segregation into age subgroups, the trials, analyzed as they were designed, demonstrate a statistically significant mortality reduction for women who begin screening by age 40.

TABLE 1. Mortality Results from the Randomized, Controlled Trials

	FOLLOW-UP PERIOD (YRS)	MORTALITY
HIP	10	29% decrease
Kopparberg	12	32% decrease
Ostergotland	12	18% decrease
Malmö	12	19% decrease
Stockholm	8	20% decrease
Gothenburg	7	14% decrease
Edinburgh	10	16% decrease
NBSS (ages 40–49)	7	36% excess
NBSS (ages 50–59)	7	3% decrease

(Adapted from Shapiro S: Screening: assessment of current studies. Cancer 74: 231, 1994.)

RANDOMIZED, CONTROLLED TRIALS ARE NOT INFALLIBLE

Although they are, currently, the most objective method for measuring benefit from an intervention, RCTs are not infallible and may be compromised in various ways. The mere fact that a trial is randomized, with controls, does not guarantee the validity of its results. The design and performance of the trial can influence its outcome.

FACTORS THAT INFLUENCE THE RESULTS OF RANDOMIZED, CONTROLLED TRIALS OF BREAST CANCER SCREENING

THE STATISTICAL POWER OF A TRIAL AND THE POWER CALCULATION

The validity of the statistical analysis of an RCT lies in its design and execution. An estimate of how successful a trial will be in demonstrating an expected benefit is termed its "statistical power." Mathematical formulas are used to determine how many women must participate in a trial, so that if there is a mortality difference, the numbers will be large enough to achieve statistical significance. For example, if there were 10 deaths in the screened group and 20 in the controls, the 50% mortality reduction might not be statistically significant, whereas 100 deaths compared with 200 (the same ratio) would be significant. The "power calculation" provides an estimate of how large a "sample size" (number of women in the trial) is needed to be able to demonstrate an expected reduction in mortality with statistical significance.

The sample size will be influenced by the number of women who will develop breast cancer over the course of the trial. If the incidence of cancer is low, the number of women who must participate must be larger than in a population that has a higher incidence of breast cancer. Because death is the point of measurement, the sample size will depend not just on the number of women who develop breast cancer but on the number of women who die of breast cancer, and, in particular, the number of women who die of breast cancer in the control group against which the screened women are compared.

Certain statistical levels have been generally accepted as proof. Designers of clinical trials usually calculate the sample size such that there will be a sufficient number of individuals with breast cancer and a sufficient number of deaths from breast cancer to have an 80% chance of showing the expected level of benefit (or higher) with a 95% likelihood that the benefit is not due to chance. This is expressed as significance at the 0.05 level.

The power calculation is based on the expected prevalence and incidence of breast cancer in the population to be screened and the mortality rate from breast cancer that is anticipated. The smaller the expected benefit, the larger the number of women needed in the trial, or, conversely, if the number of women in the trial is small, only a very large benefit can be demonstrated with statistical significance. Many trials are too small to begin with. Analysts warn:

> . . . if the statistical test fails to reach significance, the power of the test becomes a critical factor in reaching an inference. It is not widely appreciated that the failure to achieve statistical significance may often be related more to the low power of the trial than to an actual lack of difference between the competing therapies. 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.[15]

FACTORS THAT INFLUENCE THE POWER OF A TRIAL

There are numerous factors that influence the power of a trial to demonstrate, with statistical significance, an expected benefit. As already stated, the number of women involved must be sufficiently large to provide the numbers of cancers and cancer deaths needed for significance at a specified level of benefit. If the benefit is smaller than expected, the sample size must be larger to maintain the same power. Because benefit is measured as fewer deaths among the screened women when compared with the controls, if the number of cancer deaths in the control group is lower than expected, the number of women in the trial must be increased to have the same statistical power. Of course, if women in the control group do not die of breast cancer (as happened in the Canadian study), then screening cannot be shown to have any benefit.

If the mortality reduction actually proves to be larger than expected, then fewer women will be needed to achieve significant numbers. Similarly, if there are more deaths than expected among the control women, the trial will need fewer women.

NONCOMPLIANCE AND CONTAMINATION

The results of screening and the power of the trial to demonstrate a benefit are also influenced by the number of women who actually participate as they are allocated. Women who are randomized to the screened group but refuse screening (noncompliance) are still counted as having been screened in the subsequent analyses. Similarly, women allocated to the control group cannot be prevented from obtaining screening outside the trial, yet proper trial conduct requires that they be counted as unscreened controls despite the fact that they "contaminate" the purity of the control group. Women who are counted as being screened, but were not, and women who are counted as unscreened controls, but were screened, dilute the power of the trial to show its expected benefit.

In theory, proper trial design requires that the sample size be increased to take into account noncompliance and contamination. The sample size should be increased by the following factor:

> The planned sample size *multiplied by the factor* 1 / (1 − the rate of contamination − the rate of noncompliance)[2]

For example, the Canadian trial of screening women ages 40 to 49 involved 50,000 women in this age group. They cal-

culated their sample size to be sufficient to demonstrate a 40% or larger benefit. That trial had a 26% rate of contamination (women in the control group who had mammography). To maintain its power, the trial would have had to increase the sample size to approximately 89,000.

In practice, because of the fact that large trials are expensive and difficult to mount, these compensations are not made. Those analyzing the data must recognize that if the number of women is too small to begin with, or dilutional effects from control group contamination and screened group noncompliance are not accounted for in planning the trial by increasing the sample size, then the original ability of the trial to demonstrate a benefit is compromised. As stated by Moher and colleagues,

> Essential to the planning of an RCT is estimation of the required sample size. The investigator should ensure that there is sufficient power to detect, as statistically significant, a treatment effect of an a priori specified size. The opposite perspective of conceiving the problem is that the investigator should ensure that there is a low ... probability of making a type II error, or concluding that the result is not statistically significant, when the observed effect is clinically meaningful.
>
> If a trial with negative results has insufficient power, a clinically important but statistically nonsignificant effect is usually ignored, or worse, is taken to mean that the treatment under study made no difference.[16]

THE RANDOMIZATION PROCESS

The validity of RCT requires that the two groups being compared be as identical as possible. To ensure that the comparison groups are comparable, randomly allocating women to be screened or to act as controls must be performed scrupulously. Random assignment should be a blinded process so that there is no chance of inadvertently compromising the study. If the prospective number of cancer deaths is not equally divided between the two groups with the same distribution of potentially lethal cancers in both screen and control groups, the trial may be imbalanced from the beginning and may provide spurious results.

DESIGN AND PERFORMANCE CHARACTERISTICS INFLUENCE RCT

Randomized, controlled trials require a great deal of planning. In addition to the basic statistical questions and appropriate choice of the population to be included, major decisions have to be made concerning the type of screening, the interval between screens, the quality of the screen, and other, often uncontrollable factors, such as possible differences in therapy between the two groups. Many of the trials have not controlled very well for these confounding factors.

SELECTION OF A POPULATION

As noted above, the selection of a population to be studied influences the results of the trial and determines whether or not the results will be generalizable. The Two-County Trial in Sweden, for example, targeted women in several Swedish counties, randomly assigned women in a county prospectively, and then offered screening to those who were allocated to that arm, with women in an adjacent county acting as uninvited controls. Because the allocation was prospective, by county, and there was no possibility of selection bias, the results are likely to be generalizable to the entire population. The NBSS of Canada, on the other hand, randomly assigned women who volunteered for the program. It may not be legitimate to extrapolate from that trial to the general population because the results may reflect the effect of studying volunteers. The fact that the volunteers may not mirror the general population is manifest in the fact that the background 5-year survival for women ages 40 to 49 in Canada was 75%. Among the control women in the NBSS, who should reflect the average Canadian woman, the 5-year survival was over 90%. This suggests that the women who volunteered for the NBSS were not representative of the general population of Canadian women.

SHOULD WOMEN WITH CLINICAL SIGNS OF CANCER BE INCLUDED IN A SCREENING TRIAL?

Because most of the trials randomly chose women from a larger population, women with clinical signs and symptoms of breast cancer were invariably included in the screen. Epidemiologists have cautioned against the inclusion of these women.[17] Women with clinical signs of breast cancer frequently have more advanced disease and will, obviously, not benefit from screening. If these women are randomly assigned, they will only serve to dilute the benefit of screening. If they are not allocated equally to both groups, they will imbalance the trial from the outset by the overweighting of inevitable deaths in the group with the excess of these cancers. It has been suggested that such an imbalance happened in the NBSS.[12]

WHAT IS THE APPROPRIATE INTERVAL BETWEEN SCREENS?

The amount of time between screens can have a very important impact on the success of screening. If it is assumed that, at some point in the development of a breast cancer, a level is reached where cure is no longer possible (the level of lethality), the goal of screening is to interrupt the "natural history" of the tumor before it reaches the "level of lethality." This is likely the point in time when the cancer successfully metastasizes. For some cancers this point may be reached very early in the tumor's natural history, even before it can be detected by screening. For other cancers, the level may never be reached, and the woman may die of some other cause before her cancer becomes lethal. To be effective, screening must interrupt a significant number of cancers that, if left to the usual level of detection, would ultimately be lethal.

The success of the RCTs demonstrates that, for some cancers, the natural progression can be interrupted. To accomplish this, screening must be performed with sufficient frequency so that if the cancer is not visible at one screen it will grow enough over the interval between screens to become detectable, but screening must not be performed so infre-

quently that the cancer reaches the "level of lethality" before the next screen. To intercept faster-growing tumors, the interval between screens must be shortened. The longer the interval between screens, the greater the percentage of women whose cancers will grow to a level of lethality before the cancer is detected. Similarly, the faster the tumor grows, the more likely it is that short-term cancer mortality will not be affected if the interval between screens is too long (the tumors reach the level of incurability before they are detected).

The avoidance of death among screened women from cancers with moderate growth rates will not be apparent immediately because the screened woman's counterpart in the control group must die for a benefit to appear. Younger women may have cancers with faster growth rates,[18,19] and this may also account for delayed benefit.

It is likely that the mammography screening interval of 24 months or more, for women ages 40 to 49, in many of the trials has reduced the ability of those trials to demonstrate a benefit for women in this age group. The shorter lead time for mammographic detection in younger women is the reason why they should be screened every year instead of every 1 to 2 years, which is the present recommendation. Screening these women every 2 years will detect their cancers at about the time they or their doctors would have detected the cancer without screening[14,15]; thus, the impact of earlier detection by screening will be diminished. In many of the trials, women under age 50 were screened every 2 years. This is likely too long an interval. The trials that screened these women every year demonstrated greater mortality reduction than those that screened every 2 years.[20]

HOW MANY MAMMOGRAPHIC PROJECTIONS SHOULD BE USED?

There has been considerable debate concerning the number of mammographic projections that should be used in screening. Half of the trials used single-view, lateral, or mediolateral oblique mammograms. This was done to minimize x-ray exposure and to reduce the cost of screening. Unfortunately, it has been shown that at least 10% of cancers will be missed by screening with single-view mammography.[21,22] The trials that used two-view mammography have had a greater mortality reduction than those that used single-view mammography.[20]

WHAT IS THE QUALITY OF THE MAMMOGRAPHY?

One of the many issues that has not been carefully addressed in the screening trials is the quality of the mammography. Poorly performed mammography, including poor positioning of the breast in the machine, improper exposure, improper film processing, and poor interpretation, can compromise the ability of mammography to detect cancers early and reduce deaths.

The only trial that has evaluated the quality of its mammography, and the accuracy of interpretation, has been the NBSS, and the results demonstrated that both were poor.[23] In that trial there was no special training for the radiologists or technologists, and the equipment only needed to provide low-dose exposures. Image quality was not stressed. The positioning of the breast was not optimized, and there was no standardization of the processing. In a blinded review by the reference radiologist, it was shown that 25% of the cancers that were diagnosed in the program could have been detected at least 1 year earlier. Without knowledge of these factors, the data from a trial cannot be accurately interpreted.

HOW SOON SHOULD RESULTS BE ANALYZED?

Many of the trials have analyzed and published their results too soon after the end of the screening period and produced false impressions. For there to be a benefit from screening, women with breast cancer in the control group must die. Thus, a long survival among controls may result in a delay in the appearance of benefit for the screened women. This is likely the case in the HIP study, where the benefit for women ages 40 to 49 did not begin to appear until 7 to 8 years after screening began. A similar result occurred when the Swedish screening programs were evaluated together.[24]

The evaluation of results too soon after the end of screening diminishes the true benefit of screening in two additional ways. Early analysis reduces the impact of benefit for cancers that were diagnosed during the incidence period of screening, while emphasizing the outcome from those prevalent at the first screen, which generally are more advanced. When the data are analyzed too soon, many cancers are not followed for at least the standard 5 years from diagnosis, and early reporting ignores the fact that mortality from breast cancer continues for 20 or more years after diagnosis.

Because only about 30% of cancers detected in a 5-year screening trial are detected in the first year, 5 years of follow-up after initial screening means that a large number of women with cancer will have been followed for less than 5 years. Many of the study cancers will have a much smaller "at risk" period. For example, if 3.2 cancers per 1000 women are detected at the first (prevalence) screen and then each subsequent year 1 cancer is detected per 1000 women (incidence), the measure of benefit in terms of reduced mortality at 5 years after the initial screening of 100,000 women would provide 720 cases of breast cancer in the study. However, by the fifth year of follow-up, 100 women with incident cancers would have been followed for only 1 year since diagnosis, 100 women followed for 2 years, 100 for 3 years, and 100 for 4 years. Only the women diagnosed with the 320 prevalent cancers would have been followed for the standard 5 years of survival. Early reporting biases the data with the results from prevalent cancers. These are less likely to show a benefit because the first year of screening reveals a higher percentage of cancers that are at a more advanced stage than subsequent "incident" cancers, and prevalent cancers are, thus, less likely to benefit from screening. The apparent benefit of screening is reduced by mixing prevalent cases with incident cases.

HOW SOON SHOULD A BENEFIT BE EXPECTED?

Some have expected that there should be an immediate mortality reduction from breast cancer screening. They have rejected the benefit that has appeared among younger women because it is delayed 5 to 8 years after the start of the screen. This is a perplexing position because the natural history of breast cancer makes it difficult to understand how an immediate benefit could occur.

Breast cancer is lethal only through its metastases. Screening can reduce mortality by eliminating the cancer before it can successfully metastasize. For the screened woman to demonstrate a benefit in an RCT, her counterpart in the control group must develop successful metastases after the time that the screened woman's cancer was discovered and treated, and the control woman must go on to die from those metastases. This is unlikely to occur within a few years of screening; rather, it is likely to appear 5 or more years later, and certainly not 1 to 2 years after the onset of screening. Researchers should be asking why mortality is reduced soon after screening begins in women 50 and over (perhaps the primary tumor somehow supports the metastatic growth) rather than disparaging the more biologically sound "delayed" benefit.

A delay in benefit may also occur in relation 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 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 group as the higher-growth-rate lesions. It is likely that the "delayed" benefit among younger women is due to the fact that the control group women live longer with their cancers.

WHAT ARE THE APPROPRIATE THRESHOLDS FOR INTERVENTION?

There is no test for breast cancer, short of a tissue diagnosis, that can accurately differentiate benign from malignant lesions. If early-stage breast cancer is to be found, false-positive tests are unavoidable.[25] Because some cancers do have fairly characteristic appearances on mammography, it is possible to have a majority of abnormalities that are sent on to biopsy prove to be malignant, but because many cancers have appearances that overlap with those of benign lesions, maintenance of a high positive predictive value virtually guarantees that early cancers will pass through the screen.[26] The best way to determine the success of a screen is to know the size and stage of the cancers at the time of diagnosis, and to determine what percent of cancers are diagnosed between screens (interval cancers). In some programs, to save money and to avoid biopsies with benign results, early cancers have been allowed to pass through the screen.[27]

HOW COMPLETE IS THE TRIAL— WHAT IS THE CONTROL OVER BIOPSIES AND THE PATHOLOGY INFORMATION?

Despite the fact that RCTs have been organized trials, the sequence of events after an abnormality is detected has not always been clear. In the NBSS, for example, diagnosis was often haphazard. The program specialist could only recommend a biopsy; it was the patient's own physician who made the final decision. According to the NBSS publication, 25% of biopsies recommended from the screening mammograms were not performed, likely delaying the diagnosis of breast cancers. Because pathologists who were not involved in the trial performed the pathologic analysis, the amount of data collected for cancers diagnosed in the NBSS is not extensive. This makes it very difficult to determine how successful the program was in detecting small cancers.

IS THE THERAPY THE SAME FOR BOTH GROUPS?

Not only are results from RCTs influenced by the procedures involved in detection and diagnosis, but it is not always clear that the women in the two groups who were diagnosed with breast cancer have been treated in the same fashion. Although it has never been documented, it is likely that there have been differences in the trials.

It is possible that women with early breast cancer detected by screening may have been undertreated, and the benefit of early detection may have been lost. Mammography has led to a marked increase in the detection of ductal carcinoma in situ (DCIS). Many have discounted the importance of DCIS. It is possible that these lesions have been undertreated, yet it has now been shown that there is at least a 20% recurrence rate of DCIS that has not been completely treated, and that 50% of these will recur as invasive cancers.[28]

WHY IS THERE DISAGREEMENT OVER SCREENING WOMEN AGES 40 TO 49?

Most now agree that there is a benefit from screening women ages 50 and over (there is no consensus as to the age at which screening is no longer valuable).[29] There has been a great deal of controversy over the benefits for women ages 40 to 49. As has already been proved, RCTs, if analyzed as they were designed, have shown a benefit for women who begin screening by age 40. Controversy has arisen from analyses that have broken the trials into subgroups of women ages 40 to 49 and those 50 and over. Because none of the trials were designed with sufficient numbers of women ages 40 to 49 to show a statistically significant benefit of 25% to 30%,[30] it is not surprising that the benefits that have appeared are not statistically significant. Given the fact that there have only been 30% as many women in the trials as would be needed to show a 25% or smaller benefit with statistical significance, it is mathematically impossible for this level of benefit to be statistically significant.[31] Thus, it is specious to suggest that the trials have been done and there is no benefit for women ages 40 to 49.[14] Although it is statistically inappropriate to divide the data into subgroups that lack statistical power and then

treat them as if they have such power, this has been done and has led to the confusion.

There is really no reason to expect that something "magical" happens at the age of 50. The data do not support this. Contrary to this widespread fallacy, the breast does not go from dense to fat at age 50 (or at menopause), making small, early breast cancers suddenly more visible by mammography. Changes in breast tissue density do not occur in all women, and among those in whom changes do occur, they may happen at any age, or slowly over time.[32] Some have argued that the positive predictive value of mammography suddenly changes at age 50.[33] This too is fallacious[34] and is an artifact of comparing women under age 50 (ages 30 to 49) with women ages 50 to 75 or older. Because the incidence of cancer increases steadily with age, analyses that group women in this way will be skewed by the extremes (the very low incidence at age 30 and the high incidence for older women). In fact, it is clear that the positive predictive value of mammography increases gradually with age and is a direct reflection of the prior probability of cancer in the population.

The confusion over the benefit of screening women ages 40 to 49 is due to improper and faulty analysis of the data.[35] If unplanned subgroup analysis is performed, then the analyst must be prepared to accept the fact that the data are unlikely to be statistically significant. Trends and other data then become important if appropriate medical advice is to be given.

DATA THAT SUPPORT SCREENING WOMEN AGES 40 TO 49

BREAST CANCER INCIDENCE

It is often suggested that breast cancer is not a big problem for women ages 40 to 49, and that these cancers constitute only 16% of the total cancers diagnosed each year. This is an artifact of comparing women ages 40 to 49 with all other women, and skewing the observation. Approximately 1.3 to 1.6 women in 1000 will develop breast cancer each year between the ages of 40 and 49. This rises to 2.2 to 2.6 for women ages 50 to 59, and 3.3 to 3.9 for women ages 60 to 69. On the basis of SEER estimates, and as a consequence of the large number of women ages 40 to 49, in 1993 alone, 28,900 women in this age group were diagnosed with breast cancer. This is only 8% fewer than the number of women who were diagnosed between the ages of 50 and 59 (31,500).[36] Breast cancers among women ages 40 to 49 account for 16% of the total number of cancers, but women ages 50 to 59 account for only 17% of the total, and no decade of life accounts for more than 24% of the total. Because of the life expectancy of women younger than 50, more than 40% of the years of life lost to breast cancer are from women diagnosed before age 50.[37]

Despite the fact that the trials were not designed to evaluate these women separately, point estimates reveal a mortality reduction (fewer deaths from breast cancer) in five out of the eight trials[3] that ranges from 22% in the Edinburgh trial[38] to 23% of the HIP trial, 25% at Kopparberg, and 49% in the Malmö trial.[3]

The most recent data from Gothenburg reveal a 40% mortality reduction at 10 years.[39] Because of the larger-than-expected reduction for women ages 40 to 49, a new meta-analysis of all of the trials that had unscreened control groups (the Canadian trial provided the control women with a preliminary clinical examination) demonstrates a "statistically significant" benefit for screening women ages 40 to 49 (personal communication, Edward Hendrick PhD) The strength of the benefit is reflected in the fact that 30% of the women who died in the Gothenburg trial, among the screened women, had actually refused to be screened (they are still counted as having been screened). Thus, there is now "statistically significant" benefit for screening women ages 40 to 49, and, as was found in Gothenburg, the benefit is for women whose cancers were diagnosed before the age of 50.

Other studies provide evidence from which a benefit can be inferred. Data from the Breast Cancer Detection Demonstration Project (BCDDP) show that there is no significant difference between survival rates for women under age 50 and those ages 50 to 59, and no differences by age based on tumor size, grade and stage.[40] The Swedish Kopparberg trial confirmed that there is no difference in survival for younger women compared with older women when their cancers are collated by size, stage, and histologic grade.[41] Older women have shown a statistically significant mortality reduction as a result of down-staging from screening. Mammography has been shown to down-stage tumors among younger women as well, and a mortality reduction can be expected when they are screened appropriately.

Modern mammography screening trials show that the ability of mammography to detect small, early-stage cancers in women ages 40 to 49 is similar to its ability in women ages 50 to 59.[42]

INFORMING WOMEN OF THE POSSIBLE "HARM" FROM SCREENING

To bolster their argument against screening younger women, opponents of screening have turned to the fact that many concerns raised by mammography prove to be false.[43] This is true. There are women who have abnormalities on their mammograms that prove to be benign. Most of these are resolved through a few extra projections (x-rays) or by ultrasound. Some women require a tissue sampling (needle biopsy or excisional biopsy) to determine the etiology of the finding by mammography. Because of the overlap in appearances between benign and malignant processes, false-positive results are unavoidable if cancers are to be found at an earlier and more highly curable stage. This has been the case with cervical cancer screening, and there is no reason to expect that breast cancer screening would be any different.

False-positive results, however, are not unique to mammography screening. There are as many, if not more, biopsies with benign results that are engendered by CBE as there are by mammography. In our own screening program, only 7% of the biopsies recommended as a result of an abnormal clinical

examination (and a normal mammogram) prove to be cancer, whereas 20% of biopsies recommended by mammography alone reveal malignancy. When cancers are detected by CBE, they are more often at a larger size and later stage than those detected by mammography.[44]

From a scientific perspective, if the goal is to reduce the false-positive results that lead to "unnecessary" biopsies, CBE should be eliminated before mammography. As stated previously, the true positive rate for both tests increases with the age of the women being evaluated because of the fact that the prior probability of the disease increases with age. There is no sudden change in the positive predictive value of mammography at the age of 50.

HOW GREAT IS THE POTENTIAL BENEFIT FROM BREAST CANCER SCREENING?

The benefit from screening is likely to be even greater than has been shown in the trials. All of the trials have had women who were offered screening, but refused and are still counted as having been screened (noncompliance), and all of the trials have had women counted as unscreened controls who have had mammograms outside the trial (contamination) and are still counted as having not been screened. Many of the cancer deaths in the screened group have occurred among women who refused the invitation to be screened. In the Gothenburg trial, for example, among women ages 40 to 49, 30% of the women who died and are counted as having been screened refused screening. A large number of women who died in the "screened" group in Ostergotland had refused screening. These deaths dilute the results because it is likely that some of these lives would have been saved had the women agreed to screening. There are no data on how many "unscreened" control women who had mammograms had early cancers detected by those mammograms and were saved as a result. Because the number of women who had mammograms, despite being "unscreened" controls, has been as high as 35% (at Malmö), it is likely that some early cancers were detected as a result, improving the mortality among the control group and diluting the overall benefit from screening. These dilutional factors make it likely that the potential benefit from screening is higher than the data demonstrate.

ARE SCREENING GUIDELINES A PUBLIC HEALTH ISSUE?

Those involved in public health have suggested that screening is a public health issue.[45] John M. King, MBBS, FRCPA, has stated, "Screening is a public health activity concerned not with individuals, but with populations. . ."[46]

This is a fallacy that has been perpetuated.[47] In fact, screening is a public health issue only with respect to access and public financing. Tuberculosis management and immunization programs are public health issues because they benefit the entire population. Breast cancer screening benefits only the individual woman who is interested in reducing her risk of dying from breast cancer. Public health analysts would argue, for example, that noncompliance and contamination are "facts of life" that cannot be avoided, and so their analyses do not take into account the dilutional effects caused by contamination and noncompliance. An individual woman, however, might be interested in knowing that there was a high death rate from breast cancer in the screening trials among women who refused screening. Regardless, screening recommendations should be based on individual health concerns, and women and physicians should be provided with the best medical and scientific information available. "Society" may decide that screening is too expensive for public support, but guidelines should not be influenced by public health mandates that are often governed by political concerns.

COST/BENEFIT

Despite the fact that screening is not a public health issue, access and cost are behind much of the discussion, and cost/benefit analyses are dictating much of the politics behind screening.[48] Cost/benefit analysis is beyond the scope of this paper, but based on an analysis by Rosenquist and Lindfors, the cost per year saved for screening women ages 40 to 49, assuming a 30% mortality reduction, is approximately $26,200. If these women are screened until the age of 85, the cost is $16,800. These costs are well within the range of accepted health interventions.[49]

CONCLUSION

There is clear evidence that screening for breast cancer with mammography can reduce the death rate. There is no reason to believe that this only begins to work at the age of 50. The data strongly support a benefit for screening beginning at least by the age of 40.

The data also strongly suggest that women ages 40 to 49 should be advised to be screened every year, rather than every 2 years, and an annual screen is likely to benefit women over 50 as well. The only reason to go to a longer interval for older women is economic, and this should be acknowledged. Screening by mammography and CBE, if performed properly, can be expected to reduce the death rate from breast cancer for women age 40 and older by 25% to 30%.

REFERENCES

1. Kessler LG: The relationship between age and incidence of breast cancer. Cancer 69:1896, 1992
2. Bassett LW, Liu TH, Giuliano AE, et al: The prevalence of carcinoma in palpable vs impalpable mammographically detected lesions. AJR Am J Roentgenol 157:21, 1991
3. Carter CL, Allen C, Henson DE: Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. Cancer 63:181, 1989
4. Tabar L, Duffy SW, Krusemo UB: Detection method, tumor size and node metastases in breast cancers diagnosed during a trial of breast cancer screening. Eur J Cancer Clin Oncol 23 (7):959, 1987
5. Shapiro S, Venet W, Strax P, et al: Periodic Screening for Breast Cancer: The Health Insurance Plan Project and its Sequelae, 1963–1986. Baltimore, The Johns Hopkins University Press, 1988
6. Swedish overview

7. Tabar L, Fagerberg CJG, Gad A, et al: Reduction in mortality from breast cancer after mass screening with mammography. Lancet 1:829, 1985
8. Nystrom L, Rutqvist LE, Wall S, et al: Breast cancer screening with mammography: Overview of Swedish randomised trials. Lancet 341:973, 1993
9. Anderson I, Aspergen K, Janzon L, et al: Mammographic screening and mortality from breast cancer: The Malmö Mammographic Screening Trial. BMJ 297:943, 1988
10. Frisell J, Glas U, Hellstrom L, et al: Randomized mammographic screening for breast cancer in Stockholm. In Breast Cancer Research and Treatment pp 45–54. Boston, Martinus–Nijhoff, 1986
11. Roberts MM, Alexander TJ, et al: Edinburgh trial of screening for breast cancer: Mortality at seven years. Lancet 335:241, 1990
12. Kopans DB, Feig SA: The Canadian National Breast Screening Study: A critical review. AJR Am J Roentgenol 161:755, 1993
13. Shapiro S: Screening: Assessment of current studies. Cancer 74:231, 1994
14. Fletcher SW, Black W, Harris R, et al: Report of the International Workshop on Screening for Breast Cancer. J Natl Cancer Inst 85:1644, 1993
15. Lachin JM: Introduction to sample size determination and power analysis for clinical trials. Control Clin Trials 2:93, 1981
16. Moher D, Dulberg C, Wells GA: Statistical power, sample size, and their reporting in randomized controlled trials. JAMA 272:122, 1994
17. Elwood JM, Cox B, Richardson AK: The effectiveness of breast cancer screening by mammography in younger women. Online J Curr Clin Trials 32, 1993
18. Moskowitz M: Breast cancer: Age-specific growth rates and screening strategies. Radiology 161:37, 1986
19. Tabar L, Faberberg G, Day NE et al: 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 55:547, 1987
20. Sickles EA, Kopans DB: Deficiencies in the analysis of breast cancer data. J Natl Cancer Inst 85:1621, 1993
21. Muir BB, Kirkpatrick A, Roberts MM, et al: Oblique-view mammography: Adequacy for screening. Radiology 151:39, 1984
22. Andersson I, Hildell J, Muhlow A, et al: Number of projections in mammography: Influence on detection of breast disease. Am J Roentgenol 130:349, 1978
23. Kopans DB: Screening for breast cancer and mortality reduction among women 40–49 years of age. Cancer 74:311, 1994
24. Nystrom L, Rutqvist LE, Wall S, et al: Breast cancer screening with mammography: Overview of Swedish randomised trials. Lancet 341: 973, 1993
25. D'Orsi CJ: To follow or not to follow, that is the question. Radiology 184:306, 1992
26. Kopans DB, Swann CA: Observations on mammographic screening and the false positive mammograms. AJR 150:785, 1988
27. Peer PG, Holland R, Jan HCL, et al: Age-specific effectiveness of Nijmegen Population-Based Breast Cancer-Screening Program: Assessment of early indicators of screening effectiveness. J Natl Cancer Inst 86:436, 1994
28. Fisher B, Costantini J, Redmond C, et al: Lumpectomy compared with lumpectomy and radiation therapy for the treatment of intraductal breast cancer. N Engl J Med 328:1582, 1993
29. Kopans DB: Mammography screening for women over 65. J Gerontol 47:59, 1992
30. Eckhardt S, Badellino F, Murphy GP: UICC meeting on breast-cancer screening in pre-menopausal women in developed countries. Int J Cancer 56:1, 1994
31. Kopans DB, Halperin E, Hulka CA: Statistical power in breast cancer screening trials and mortality reduction among women 40–49 with particular emphasis on The National Breast Screening Study of Canada. Cancer 74:1196, 1994
32. Kopans DB: Conventional wisdom: Observation, experience, anecdote, and science in breast imaging. AJR 162:299, 1994
33. Kerlikowske K, Grady D, Barclay J, et al: Positive predictive value of screening mammography by age and family history of breast cancer. JAMA 270:2444, 1993
34. Kopans DB: The use of mammography for screening. JAMA 271:982, 1994
35. Sickles EA, Kopans DB: Deficiencies in the analysis of breast cancer screening data. J Natl Cancer Inst 85:1621, 1993
36. Smith RA: Epidemiology of breast cancer in A Categorical Course in Physics: Technical Aspects of Breast Imaging, 2nd ed, pp 21–33. RSNA Publications, Radiological Society of North America, Oak Brook, IL 1993
37. Shapiro S, Venet W, Strax P, et al: Periodic Screening for Breast Cancer: The Health Insurance Plan Project and its Sequelae, 1963–1986. Baltimore, The Johns Hopkins University Press, 1988
38. Moss S, Alexander F: Presentation to the International Union Against Cancer (UICC) meeting on breast cancer screening in pre-menopausal women in developed countries. Geneva, Switzerland, September 29–October 1, 1993
39. Bjurstam N, Bjorneld L: Mammography screening in women aged 40–49 years at entry. Results of the randomized, controlled trial in Gothenburg, Sweden. Presented to the 26th National Conference on Breast Cancer, Palm Springs, CA May 8–13, 1994
40. Smart CR, Hartmann WH, Beahrs OH, et al: Insights into breast cancer screening of younger women. Cancer 72:1437, 1993
41. Tabar L: New Swedish breast cancer detection results for women aged 40–49. Cancer 72:1437, 1993
42. Sickles EA, Kopans DB: Deficiencies in the analysis of breast cancer data. J Natl Cancer Inst 85:1621, 1993
43. Harris R: Breast cancer among women in their forties: Toward a reasonable research agenda. J Natl Cancer Inst 86:410, 1994
44. Bassett LW, Liu TH, Giuliano AE, et al: The prevalence of carcinoma in palpable vs impalpable mammographically detected lesions. AJR Am J Roentgenol 157:21, 1991
45. Goldberg K: NCI: Facts don't support screening mammography for 40–49 year olds. Cancer Lett 19 (37):1, 1993
46. King J: Letter to the Editor. Mammography screening for breast cancer. Cancer 73:2003, 1994
47. Kaluzny AD, Rimer B, Harris R: The National Cancer Institute and guideline development: Lessons from the breast cancer screening controversy. J Natl Cancer Inst 86:901, 1994
48. Eddy DM, Hasselblad V, McGivney W, et al: The value of mammography screening in women under 50 years. JAMA 259 (10):1512, 1988
49. Rosenquist CJ, Lindfors KK: Screening mammography in women aged 40–49 years: Analysis of cost-effectiveness. Radiology 191: 647, 1994

Some compounds discussed in this publication may be in investigative stages and not yet approved by the FDA for uses mentioned. The reader is urged to check the package insert for each drug for any change in indications and dosage and for added warnings and precautions.

Mammography
Debbie

OCT 19 1994

Facsimile Transmission Cover Sheet

OFFICE OF DISEASE PREVENTION AND HEALTH PROMOTION
OFFICE OF HEALTH PLANNING AND EVALUATION
OFFICE OF THE ASSISTANT SECRETARY FOR HEALTH
330 C STREET, S.W., Room 2018
Washington, D.C. 20201

Fax: (202) 690-6603

DATE: 10/18/94

Total # of Pages (incl. cover sheet): 3 pages

ADDRESSEE: S. Broder, P. Greenwald, B. Kramer, B. Smith, S. Blumenthal, H. Varmus, P.Lee, D.Kesler/Merkatz, S.Love
H.S. Adams

BUILDING/ROOM #: (or address)

FAX #: 301 402 0338, 301 496 8667, 690-6960, 301 443-3100, ~~301 496 9931~~, 310 794 9461, 404 325-2548, 401-4005

PHONE #: 301 402 3700
011 46 18503431

FROM: Dr. Devra Davis

BUILDING/ROOM #:

PHONE #: 202 690 8033

Comments:

Note new fax number: 202 205 4840

FINAL DRAFT
REPLY TO KOPANS FOR JAMA

Unlike Dr. Kopans, we believe that public policies about screening should be based on verifiable, published information, and not on unpublished analyses--even those legitimated as letters to journals. The unpublished meta-analysis to which Kopans refers excludes the major Canadian study, which was the only study explicitly designed to assess the value of screening women under 50, and includes unpublished data from the Gothenburg study. Thus, significance is only achieved by the selective inclusion of data.

Kopans' unreferenced allegation that radiation after age 40 carries no increased risk of cancer is wrong, as two recent studies show[1]. While contemporary, radiation-induced cancer risks from mammography are much lower than those in past studies, the younger a person is when regular radiation exposure begins, the greater her lifetime risk [2]

Moreover, providing mammographic screening to women in their 20s and 30s, as some are doing today, could place them at increased radiogenic risk, both from the cumulative impact of exposures over a lifetime and from any inherited susceptibility. Some of the 5% of all breast cancer cases that stem from inheritance of the BRCA1 gene may be more susceptible to radiation, as are patients with the relatively rare Ataxia Telangiectasia gene.[3]

Kopans provides no substantiation for his charge that we have grossly overestimated the number of procedures engendered by screening; and several studies have confirmed that the rate of such procedures is far higher in women under age 50.[4]

In the meantime, what can be done? Efforts to identify women under 50 at highest risk for breast cancer and provide appropriate clinical screening for them may well prove to be valuable.[5] And any woman at any age with a suspicion or symptom needs to be evaluated by a health professional. But, routine mammographic screening of asymptomatic women under the age of 50 results in unnecessary surgery and anxiety, with no reduction in deaths.[6] So, we reiterate our call to ensure that women over 50 are provided life-saving screening, and to see that we reach the post-mammography era, when improved methods of breast cancer control, prevention, and early detection will supplant this technology for all women.

When health care policy is data driven, rather than opinion inspired, we will all be better off. We do no less when we build bridges and superhighways; we should do no less, when it comes to providing health services.

Devra Lee Davis

Susan M. Love

1. Mattsson, A. Ruden, B.-I. Hall, P. Wilking, N. Rutqvist, L.E. Radiation-Induced Breast Cancer: Long-Term Follow-up of Radiation Therapy for Benign Breast Disease. J Natl Cancer Inst 1986;7789-92.; Inskip, P.D., Stovall, M., and Flannery, J.T. (1994) Lung Cancer Risk and Radiation Dose Among Women Treated for Breast Cancer, J. Natl Cancer Inst., 86:13; 983-88.

2. National Research Council, Board on Radiation Effects Research, Report, 1994 VI Washington: National Academy Press. New England Journal of Medicine.

3. Swift,M.,Rettnauer,P. Morrell,D., Chase,C. Breast and Other Cancer in Families with Ataxia-Telangiectasia. New England Journal of Medicine 1987;316:1830-1835.

4.Kerlikowske, K., Grady, D., Barclay J., Sickles, E.A., Eaton, A., Ernster, V. (1993) Positive predictive value of screening mmammography by age and family history of breast cancer, JAMA, 270: 2444-50.; Lantz, P.M. Remington, P.L. Newcomb, P.A. Mammography screening and increased incidence of breast cancer in Wisconsin (1991) J. Natl Cancer Inst 83: 1540-46.

5. Vogel, VG Screening Younger Women at Risk for Breast Cancer (1994) J. Natl Cancer Institute Monographs 16:55-60.

6. Eddy, D. Assessing Health Practices and Designing Practice Policies: the Explicit Approach (1992) Philadelphia: American College of Physicians.

Statement

For Response to Inquiries

NATIONAL CANCER INSTITUTE

Office of Cancer Communications
Building 31, Room 10A24
Bethesda, Maryland 20892

National Institutes of Health

December 3, 1993

NCI Press Office
(301) 496-6641

Breast Cancer Screening

The National Cancer Institute (NCI) today released the following statement concerning breast cancer screening:

> There is a general consensus among experts that routine screening every 1 to 2 years with mammography and clinical breast examination can reduce breast cancer mortality by about one-third for women ages 50 and over.
>
> Experts do not agree on the role of routine screening mammography for women ages 40 to 49. To date, randomized clinical trials have not shown a statistically significant reduction in mortality for women under the age of 50.

The statement represents a summary of scientific fact about effectiveness, that is, the ability of mammography, coupled with appropriate treatment, to reduce the mortality from breast cancer. It summarizes scientific knowledge derived from two decades of clinical trials research. The statement is a successor to a "working guideline" formulation drafted in 1987 and will be revised as new information is developed.

Background Information

The National Cancer Institute is the lead federal agency for research on the causes, prevention, diagnosis and treatment of cancer. NCI conducts ongoing evaluations of the results of cancer research, and in late 1991 began the process of examining clinical trial evidence of the value of screening mammography.

(more)

NCI convened an International Workshop on Breast Cancer Screening in February 1993. The results from eight randomized clinical trials were reviewed. The workshop conclusions reinforced the advisability of screening for women ages 50 to 69, and stated that the effects of screening in women ages 40 to 49 do not demonstrate a statistically significant reduction in mortality to date.

Between May and December 1993, scientific data from clinical trials, including the workshop results, were reviewed by a number of scientific organizations, health groups, and advisory boards.

# # #

Press documents can be downloaded from Compuserve®. They are located in the SciNews-MedNews library" which is located in the Journalism forum (GO JFORUM).

COMMENTARY

The National Cancer Institute and Guideline Development: Lessons From the Breast Cancer Screening Controversy

*Arnold D. Kaluzny, Barbara Rimer, Russ Harris**

On December 6, 1993, the National Cancer Institute (NCI) issued the following statement:

> There is a general consensus among experts that routine screening every one to two years with mammography and clinical breast examination can reduce breast cancer mortality by about one third for women ages 50 and over.
>
> Experts do not agree on the role of routine screening mammography for women 40-49. To date, randomized clinical trials have not shown a statistically significant reduction in mortality for women under the age of 50.

The statement represents the culmination of a process that has its roots in a long-standing debate regarding the role of routine screening mammography. The debate began again in the fall of 1992 when the most recent data (*1*) from breast cancer screening trials found no evidence that mammography and clinical breast examination were effective in reducing mortality for women in their 40s. In the past year, the debate has been particularly acrimonious and visible and was confusing and disconcerting to both the scientific community and the general public.

Unlike previous discussions of this issue, the past year's debate took place in a climate in which there was a greater awareness of what "guidelines" are and what elements are required to create them. Thus, the question about what is the most appropriate guideline was joined by a second question: Is the NCI an appropriate agency to formulate guidelines at all?

When the new evidence from breast cancer screening trials became available, the NCI began a formal review process in February 1993 by organizing an International Workshop on Screening for Breast Cancer. The workshop considered results from eight randomized controlled trials of breast cancer screening and gave specific attention to trials with new information on breast cancer mortality. These trials included the Canadian National Breast Screening Studies and the Swedish trials. The workshop also reviewed findings from the Health Insurance Project (HIP) and the Edinburgh trials. The workshop was open to the general public, with attendance not only of the invited speakers, who represented all of the major international trials, but also of the media, volunteer and professional organizations, and advocacy groups, including the American Cancer Society, the American College of Radiology, and the Women in Health Network.

The workshop report was presented to the President's Cancer Panel in May 1993 and was subsequently published in the Journal (*2*). Based on this report, a statement was presented for review by the National Cancer Advisory Board (NCAB). The NCAB referred this matter to the Board of Scientific Counselors (BSC) of the NCI's Division of Cancer Prevention and Control. The BSC, which is composed of a diverse group of experts, including epidemiologists, clinicians, and biostatisticians, met in October 1993 and heard testimony from more than 30 different speakers. This meeting was an open process of inquiry with broad-based participation. Speakers included each of the members of the report-writing team from the International Workshop on Screening for Breast Cancer as well as representatives of professional organizations (e.g., American College of Physicians, American College of Radiology, Oncology Nursing Society, American College of Obstetrics and Gynecology, and the American Academy of Family Physicians) and representatives of breast cancer advocacy organizations, consumer health organizations, and the American Cancer Society. The meeting was well attended by the media and included representatives of the Centers for Disease Control and Prevention (CDC), the Food and Drug Administration, and other federal agencies.

On the basis of the testimony and subsequent deliberations, the BSC released the following statement concerning the effectiveness of breast cancer screening:

> There is general consensus among experts that the National Cancer Institute 1987 screening mammography recommendations for breast cancer for women 50 years of age and over (i.e., annual screening) be maintained.

**Affiliations of authors:* A. D. Kaluzny, R. Harris, University of North Carolina, Chapel Hill.

B. Rimer, Duke University, Durham, N.C.

Correspondence to: Arnold D. Kaluzny, Ph.D., 725 Airport Road Bldg., Campus Box 7590, University of North Carolina, Chapel Hill, NC 27599-7590.

See "Note" section following "References."

For those women under 50, the National Cancer Institute will provide a summary of existing evidence and data and suggests these be discussed with each woman's physician or health care provider.

This recommendation was presented to the NCAB on November 22, 1993; strong opinions were heard. Even prior to the meeting, a motion was circulated and eventually passed that recommended that the NCI defer action on recommending any changes to breast cancer screening guidelines at that time. Two weeks later, the NCI issued its December 6 statement, positioned not as a guideline but as a statement of the evidence.

It would be difficult for anyone to disagree that the NCI's final statement is indeed a statement of fact. There are those (*3*), however, who argue that other, indirect evidence, taken together with the costs of screening, justify a guideline for screening women in their 40s. On the issue of financial cost, the NCI statement is silent. Other, similarly controversial issues in cancer prevention and treatment abound in breast cancer, as with other diseases. With the pace of new research, more controversy can be expected. Should we screen men with prostate-specific antigen for prostate cancer? Should we increase screening for people with genetic markers? What role should the NCI play in developing guidelines for these situations?

Practice guidelines are at least as old as textbooks of medicine. Yet, only since the 1970s and the pioneering work of Frame and Carlson (*4*) and the Canadian Task Force on the Periodic Health Examination (*5*) has there been specific attention paid to the methodology of developing guidelines. This quest has been continued and extended by the U.S. Preventive Services Task Force (*6*) in the 1980s and, most recently, by Eddy (*7*). As noted above, this methodologic discussion has led to a new awareness of what steps are necessary to develop credible guidelines.

The development of any guideline must weigh potential benefits of the procedure against what is known about risks and costs (both economic and human). If research demonstrates that benefits are substantial and risks/costs are minimal, then we can assume that the public would want the procedure. The guideline should recommend the procedure. On the other hand, if research demonstrates minimal (or no) benefits and substantial risks/costs (as we believe is the case with screening for breast cancer in women in their 40s), then there is justification for not recommending the procedure.

Problems arise when there is insufficient evidence to determine benefits and risks/costs, when the benefits and risks/costs are close enough so that it is not clear which is greater and/or when the comparison involves known risks against theoretical benefits. In each case, the preferences of patients must be assessed rather than assumed.

The NCI is in an excellent position to contribute to this process. It has access to the latest research and has a highly competent staff and consultants to review and interpret the findings. It can comment with authority on both potential benefits and health risks of various cancer prevention and treatment strategies. Indeed, the NCI has a responsibility to speak out about the facts regarding medical care and environmental or other factors that affect the public's health. This is what the NCI has done with respect to breast cancer screening.

The NCI is, however, less well equipped to carry out other parts of the process of developing guidelines. It is, for example, not as expert in assessing the financial costs of health care strategies as other groups, and it is not well equipped to scientifically assess the preferences of groups of patients (a process very different from the approach of hearing testimony from advocacy groups). Finally, weighing potential benefits against costs is not an area of particular NCI expertise.

In December 1989, President Bush signed into law an Omnibus Reconciliation Act, creating the Agency for Health Care Policy and Research (AHCPR) on the same level as the CDC and the National Institutes of Health. This agency was charged with "arranging for" the development and periodic review and updating of "clinically relevant guidelines that may be used by physicians, educators, and health care practitioners to assist in determining how diseases, disorders, and health conditions can most effectively and appropriately be prevented, diagnosed, treated, and managed clinically." As part of this mandate, the AHCPR has adopted and continues to refine a rigorous methodologic approach to ensure the reliable and valid development of guidelines. Moreover, the Institute of Medicine (*8*) has convened a panel on clinical practice guidelines specifically to encourage a careful, rigorous approach to their development.

Given the organizational mechanism and the available methods for developing practice guidelines, the NCI's December statement on breast cancer screening makes eminent good sense. It has made a statement about the scientific facts and left for others such as the AHCPR the challenge of developing a guideline. Others without the expertise to carry out all the steps of guideline development would do well to follow the NCI's lead.

The issue is not to recreate the process of developing guidelines within the NCI or to limit the NCI's role in promoting the public's health. Instead, the challenge is to use the mechanisms and processes that have proved to be effective in other settings to best inform physicians and the general public. We believe the best role for the NCI in the development of clinical guidelines is the one Dr. Broder has so well articulated—to be the science agency. In this model, the NCI would accumulate knowledge, define the state-of-the-art, and serve as part of the guidelines development team, headed by the AHCPR or other mandated agencies. Ultimately, however, just as the science of cancer is best left to the experts, so is the science of guideline development.

References

(*1*) Miller AB, Baines CJ, To T, et al: Canadian National Breast Screening Study: breast cancer detection and death rates among women aged 40 to 49 years. Can Med Assoc J 147:1459-1476, 1992

(*2*) Fletcher SW, Black W, Harris R, et al: Report of the international workshop on screening for breast cancer. J Natl Cancer Inst 85:1644-1656, 1993

(*3*) Kopans DB: Letter to the *New York Times*, January 6, 1994

(*4*) Frame PS, Carlson SJ: A critical review of periodic health screening using specific screening criteria. J Fam Pract 2:283-289, 1975

(5) The periodic health examination: Canadian Task Force on the Periodic Health Examination. Can Med Assoc J 121:1193-1254, 1979
(6) US Preventive Services Task Force: Guide to Clinical Preventive Services: an Assessment of the Effectiveness of 169 Interventions. Baltimore: Williams & Wilkins, 1989
(7) Eddy D: Assessing Health Practices and Designing Practice Policies: the Explicit Approach. Philadelphia: American College of Physicians, 1992
(8) Institute of Medicine: Guidelines for Clinical Practice. Washington, DC: Academic Press, 1992

Note

Manuscript received January 28, 1994; accepted March 31, 1994.

May We Agree to Disagree, or How Do We Develop Guidelines for Breast Cancer Screening in Women?

*G. Marie Swanson**

In the spring of 1994, we face a number of controversies concerned with breast cancer in women. One of these controversies has raised the interest, the voices, and even the blood pressure of researchers, breast cancer advocates, public health professionals, and government officials: What shall be the guidelines for screening for breast cancer? We no longer have a consensus on this policy. Researchers, advocates, public health professionals, and public policy makers disagree among themselves and with each other. The approach of this commentary is to consider principal issues that play a role in the development of policies about cancer screening in general and then to apply those issues to breast cancer screening.

Before we evaluate diverse public policy positions about screening for breast cancer, it will be useful to delineate the three major components of public policy that are considered when establishing such guidelines. The general elements of establishing a guideline for cancer screening include the following: 1) research results assessing the efficacy of a screening modality, 2) risks and benefits to the public health of a screening modality, and 3) economic impact of a screening guideline.

Today, the accepted research evidence of cancer-screening efficacy is a reduction in mortality in the population screened when compared with a population not screened. The accepted research design for producing this evidence is a clinical trial. Public health benefits of cancer screening include reduction in physical morbidity for the screened patient who is later diagnosed with cancer, reduction in costs of care for the screened cancer patient, reduction in psychosocial morbidity for the screened patient and her family, and access to methods of early detection for the general population. Public health risks of cancer screening include costs and morbidity associated with screening modalities that have poor sensitivity and specificity, lack of access to methods of screening that are widely utilized in the health care system, and public perceptions that certain populations are excluded by policy from access to specific methods of cancer screening. The economic impact of a screening guideline includes the costs of implementing a guideline, the costs of limiting access to specific populations, and the costs of excluding defined populations from access to screening. These costs are both monetary and social.

In my view, the fundamental question is how do we balance these three elements of a public policy on breast cancer screening? Should we limit public policy concerned with cancer screening to statements that are derived solely from evidence provided by clinical trials that demonstrate a reduction in mortality? Should research data serve as the initiator of public policy when such evidence is available, with public health and economic concerns having equal weight in the translation of research results into public health practice? Even if research evidence is our primary source of public policy concerned with screening for cancer, should we demand evidence for all demographic subsets of the population before the screening procedure is made available to those populations? Is it reasonable to restrict access only to those age groups, ethnic groups, racial groups, or socioeconomic groups that were specifically evaluated through clinical trials?

As an exercise in the development of public policy for breast cancer screening, I propose the following three approaches to establishing a guideline: 1) a research-based approach, 2) a public health-based approach, and 3) an economics-based approach. With the exception of the research-based approach, multiple options are available. Rather than attempting to delineate each of the possible alternatives, I will suggest a policy that could reasonably be derived within that approach.

Research-based breast cancer-screening guideline. Screening for breast cancer, utilizing the combined modalities

**Affiliation of author:* Cancer Center and Department of Medicine, Michigan State University, East Lansing.
Correspondence to: G. Marie Swanson, Ph.D., M.P.H., Cancer Center and Professor of Medicine, Michigan State University, A-128 East Fee Hall, East Lansing, MI 48824.
See "Note" section following "References."

IN THIS ISSUE

Breast Screening: How Effective?

A study from the Netherlands finds no apparent advantage to mammographic screening for women under age 50, although screening did benefit older women. Among those under 50, no substantial difference in the proportion of positive-lymph-node cancers was found between tumors detected by screening and those detected clinically. To reduce cancer mortality, screening must detect earlier stage disease than that found clinically.

Peer et al. (p. 436) analyzed data, after eight rounds of biennial screening, on more than 40 000 women invited to participate. Screening detected 37% of cancers among women under age 50, 48% among women age 50-69, and 35% among women age 70 or older.

In women over 50, about one third of screening-detected invasive tumors were advanced stage; however, more than two thirds of screening-detected tumors were advanced stage in women under 50. Investigators conclude that detecting invasive breast cancer in its premetastatic stage is more difficult in younger women than in older women.

In an editorial, Harris (p. 410) says that data from the Peer study and from other studies provide base-line detection rates for measuring new screening technology. But the data do not yet warrant the mounting of another randomized, controlled trial to definitively answer the question of whether to screen women in their forties, he says. Instead, Harris suggests a different research agenda—including surveillance of screening populations, case–control studies, and improvement of mammography technology and practice—that is potentially more productive.

Cost-Effectiveness of Therapy

In 1989, the National Cancer Institute issued an announcement advising physicians of the benefits of the combined use of levamisole and fluorouracil as adjuvant treatment for patients with stage III colon cancer. A study by Brown et al. (p. 424) suggests that the cost-effectiveness of this therapy is favorable and that the research investment promises a high economic return due to the deaths averted.

The investigators used the computer simulation model CAN*TROL to estimate costs and benefits for a population cross section receiving adjuvant therapy between the years 1990 and 2000, with follow-up accounting through 2020. Results showed a treatment cost of $2094 to $5000 per year of life saved. These values, investigators say, reflect a favorable cost-effectiveness for the treatment.

Further analysis indicated that during the years of life saved by the treatment, patient earnings would add $2.01 billion to the economy. Subtracting a research investment of $10.84 million and treatment/care costs of $338 million leaves an economic return of $1.66 billion.

Can Contraceptive Regimen Cut Breast Cancer Risk?

Indirect evidence suggests that a new hormonal contraceptive regimen, using small amounts of estrogen and progestogen, may reduce breast cancer risk while maintaining the beneficial effects of standard oral contraceptives on ovarian cancer risk, according to a study by Spicer et al. (p. 431). The regimen uses gonadotropin-releasing hormone agonist to suppress ovarian function; this approach permits using much less estrogen and progestogen than do standard contraceptives.

Spicer et al. compared changes in mammographic densities between base line and 1 year for 13 women given the regimen (7.5 mg intramuscular leuprolide acetate every 28 days, 0.625 mg oral conjugated estrogen for 6 days weekly, 10 mg oral medroxyprogesterone acetate for 13 days every fourth 28-day cycle) and seven women not treated. Results showed a significant density reduction for women on the regimen.

Indirect evidence links reduction of densities to reduced breast cancer risk. Increased densities are associated with increased risk, and early menopause is associated with reduced densities and risk.

In an editorial, Feig (p. 408) writes that the hormonal contraceptive regimen of the Spicer study represents a novel approach to reducing breast cancer risk. Feig outlines several questions about the regimen that should be addressed by future research, including, if the regimen is stopped, whether mammographic densities would return and breast cancer risk would remain low. He also points out additional benefits the regimen might have, such as improving the sensitivity and diagnostic specificity of mammographic screening.

Taxol as Radiosensitizer

Paclitaxel (Taxol) can sensitize some, but not all, human tumor cells to radiation to about the same degree as do other chemotherapeutics, a study shows. Paclitaxel blocks cells in mitosis, the most radiosensitive stage of the cell cycle. Mitotic blockade appears necessary but not sufficient for radiosensitization.

Liebmann et al. (p. 441) studied the effects of 0-10 000 n*M* paclitaxel on radiosensitivity of human breast (MCF-7), lung (A549), ovary (OVG-1), and pancreas (PC-Sh) adenocarcinoma cells exposed to 0-10 Gy radiation. Investigators also analyzed cell-cycle characteristics of cells exposed to similar drug concentrations for 6-72 hours.

All cell lines developed G_2/M blocks after 24 hours' exposure to 100-10 000 n*M* paclitaxel. But lines exhibited varying levels of radiosensitivity, ranging from a 1.8 sensitizer enhancement ratio (SER; radiation dose that kills 90% of control cells:radiation dose that kills 90% of paclitaxel-exposed cells) for MCF-7 cells to no radiosensitization for A549 cells.

Breast Cancer Among Women in Their Forties: Toward a Reasonable Research Agenda

*Russell Harris**

Many women in their forties feel confused and betrayed that they have been left with nothing to do about their high risk of developing breast cancer. They are aware of their increased risk ("one in 10 women" is now "one in eight"), and there is a growing realization that mammography, breast examination, and even breast self-examination are not effective in reducing the risk of dying of breast cancer for women aged 40-49 years (*1*).

Cancer control researchers need to hear the complaints of these women and to develop a reasonable research agenda to address their concerns. What is a reasonable agenda? To address this question, I will first review the breast cancer-screening debate and the problems of evaluating new technology in an effort to learn what direction research about screening should take. I will then make several other observations about research priorities for breast cancer control.

The Debate and Evaluating Modern Mammography

As most women in their forties know, there has been a vigorous debate during the past year about the effectiveness of breast cancer screening for women aged 40-49 years. The debate has centered on the seven randomized, controlled trials that have provided direct evidence about this issue. As a group, these studies provide no evidence that screening is effective among women in their forties, yet they do give clear and consistent evidence that mammography and clinical breast examination reduce breast cancer mortality by about 30% for women aged 50-69 years (*1*).

Some critics (2) claim that these studies are irrelevant to the question of whether modern mammography could reduce breast cancer mortality for women under 50, primarily because they used older mammography equipment, sometimes with only one view per breast and sometimes at intervals of longer than 1 year. It is a fair criticism, although it is one that could be made about many technologies that have not been shown to be effective. Research to evaluate a technology takes time, sometimes long enough that the technology has moved ahead before the study is completed.

That does not mean, of course, that we should not require technology to be evaluated before encouraging widespread use. Furthermore, it does not mean that the burden of proof should be shifted: Those who propose widespread use of a technology must offer strong evidence that it is effective and that, under usual situations, the benefits of the technology outweigh its risks and costs.

How then are we to evaluate new technology and, specifically, new, improved mammography? Should we begin a new randomized, controlled trial, which would need to involve hundreds of thousands of women and could take 15-20 years to organize and complete, every time there is a claim of a breakthrough in mammography technology? It is a vexing question, one for which there are only partial answers.

One clear statement can be made. Before mounting more randomized, controlled trials of breast cancer screening, we should require clear evidence that the new technology is better at detecting early breast cancer than the old technology. Early detection of breast cancer is not synonymous with reduction in mortality, but it is a necessary precondition.

The usual way we evaluate detection is by following so-called "intermediate measures" for screened populations of truly asymptomatic women similar to those who are candidates for screening (as opposed to referred or volunteer populations). These measures (*3,4*) include such indicators as the prevalence of cancer found by screening, the sensitivity of screening (i.e., the proportion of all cancers that emerge within the screening interval that are detected by screening), and the stage (size and lymph node status) of cancer at diagnosis. It is best to demonstrate improvement in these measures at second and later screens, because previously unscreened groups harbor more and larger cancers, making initial screens appear more sensitive.

The study by Peer et al. (*5*) reported in this issue of the Journal uses these measures to examine cancer detection in the Nijmegen population-based breast cancer-screening program. The study finds little evidence of improved detection of breast cancer for women under 50 years of age. Screen-detected cancers did not show a shift to earlier stage at diagnosis compared to clinically detected cancers among women under 50 as they did among women aged 50-69 years. This observation is similar to findings from the Health Insurance Plan Project (*6*) and the National Breast Screening Study from Canada (*7*). A generous calculation of sensitivity from the Nijmegen data (i.e., including cancers detected at first screen and including ductal carcinoma in situ cases as detected cancers) is 49% for women under 50 years of age, 65% for women aged 50-69 years, and 74% for women aged 70 years or more [calculated from Table 2 (*5*)]. These figures are on the lower end of the range of sensitivity from similar calculations on data from several of the randomized, controlled trials (*1*). In those studies, the sensitivities ranged from 53% to 81% for women aged 40-49 years and from 63% to 88% in women aged 50-59 years. These findings are

**Correspondence to:* Russell Harris, M.D., M.P.H., UNC-Lineberger Comprehensive Cancer Center, University of North Carolina/Chapel Hill, CB #7300, Mason Farm Rd., Chapel Hill, NC 27599-7300.

See "Note" section following "References."

not surprising, given previous studies using older mammography technology (the Nijmegen program began in 1975), single view, and a screening interval of every 2 years.

These data and others from randomized, controlled trials provide "base-line" detection rates against which to measure the new technology. There are claims (*2*) that more recent and intensive mammography is far better at detection. These new data need to be examined critically to determine whether the population screened was a selected one, whether the women were truly asymptomatic, and whether these figures are for initial or later screens.

Is Evidence of Improved Detection Enough to Mount Another Randomized, Controlled Trial?

As noted above, evidence of improved detection is necessary but not sufficient to recommend widespread screening. There must also be strong evidence of a decrease in breast cancer mortality or an improvement in quality of life. If indeed we are presented with valid and reproducible evidence of improved detection of cancer for women in their forties, should we then proceed with another randomized, controlled trial to answer this question? Although a case can be made to move ahead with such a study (*8*), there are at least three reasons to consider approaching the problem first with other research, using surveillance and case–control methodologies.

The first reason is that the previous randomized, controlled trials have not demonstrated a clear relationship between detection and mortality reduction for women aged 50-69 years. In this age group, all screening programs have found an approximately 30% reduction in mortality, regardless of whether mammography was vintage 1960s, 1970s, or 1980s; regardless of whether one or two views were done; regardless of whether the interval was annual or every 2 years (or longer); and regardless of whether clinical breast examination was also performed. This outcome may be because it is possible for mammography to find a cancer before it would have been detected clinically and still not before some "critical point" where response to treatment is improved. Thus, improved detection, although necessary, does not equal reduced mortality.

The second reason for not immediately beginning a large randomized, controlled trial of screening for women in their forties is that the potential benefit is relatively small. Even if, for example, the actual benefit is a 30% reduction in breast cancer mortality (similar to the reduction found for older women), that translates into prolonging the life of one woman from screening 1000 women annually for 10 years (i.e., 10000 mammograms)!

The third reason is that there is the potential for doing more harm than good. An important effect of mammographic screening in all age groups has been the large increase in the number of cases detected of ductal carcinoma in situ. That is certainly shown in the report by Peer et al. (*5*). The natural history of this lesion is as yet unclear. If the natural history of most of these cases is to progress to invasive cancer and if they can be cured only by finding them in this stage (rather than waiting to detect them after age 50), then this may turn out to be a much desired effect. If, however, a number of these cases never progress or grow so slowly that lives can still be prolonged by screening at a later age, then screening women in their forties may only produce unnecessary and premature treatment and anxiety.

Directions for Screening Research

There is another reasonable agenda for screening research other than a randomized, controlled trial. The first is surveillance of populations with different screening behaviors. Especially useful would be measures of mammography quality and intensity, measures of detection (e.g., proportion of total cancers found by screening and stage distribution of cancers detected), measures of compliance to screening (e.g., proportion of women in various age groups who are screened with mammography and/or clinical breast examination), and breast cancer mortality for the same population. Changes in screening behavior could then be correlated with changes in detection and, eventually, mortality for various age groups. Although this approach would take time, we should certainly learn a considerable amount before the 15 years that a randomized, controlled trial would require.

A second part of this agenda that could be completed relatively quickly is a case–control study. As there is now variation in screening behavior among women in their forties, case–control studies of the association between screening and late-stage breast cancer in this age group could be conducted. These studies could be done in areas performing high-quality mammography.

Further development of mammography technology and practice should move forward. Improving sensitivity is a reasonable goal, but equal effort should be exerted to improving specificity (i.e., decreasing false-positive results). If specificity could be substantially improved, the risk and cost of screening would be reduced, thus making the value of mammography clear for even small benefits. One interesting way of improving the specificity of screening would be to define low-risk groups (perhaps through genetic markers or lack thereof) who do not need screening at all.

There is a need to develop further the technique of the clinical breast examination. This long-recommended but underdeveloped examination can clearly be improved, and the early results of the Canadian National Breast Screening Study (*7*) imply that it has the potential to be more accurate than we know.

Directions for Other Research

It is clear that women of all ages often misperceive their risk of developing and dying of breast cancer (*9*). We need more effective ways of helping them better understand their true risk and how that change in risk perception affects behavior. We also need research to learn how to help physicians educate their patients about accurate breast cancer risk.

Continued efforts to understand the natural history of ductal carcinoma in situ may also be considered cancer control research. The frequency with which modern mammography detects this pathologic lesion means that its contribution to mortality needs to be examined critically.

Finally, the enormous efforts in primary prevention, chemoprevention, identifying genetic markers, and even treatment of breast cancer may have a profound effect on screening and breast cancer control. At present, screening has been the most effective tool against breast cancer, and its failure for women under 50 has seemed a particularly cruel blow. But it is unlikely that screening mammography will be the long-term vehicle for controlling this disease.

Although it is tempting for us as researchers to propose the definitive study—a randomized, controlled trial—to answer the question of cancer control for women in their forties, there is another long and potentially more productive research agenda that could give useful information in a shorter time frame than that required for a randomized, controlled trial. It may also lead us away from mammography altogether, as we develop more accurate tools to deal with this disease. Let us hope that the "post-mammography" era arrives soon and that the newer methods of breast cancer control are as effective for women in their forties as for older women.

References

(1) Fletcher SW, Black W, Harris R, et al: Report of the International Workshop on Screening for Breast Cancer. J Natl Cancer Inst 85:1644–1656, 1993

(2) Sickles EA, Kopans DB: Deficiencies in the analysis of breast cancer screening data. J Natl Cancer Inst 85:1621–1624, 1993

(3) Day NE, Willams DRR, Khaw KT: Breast cancer screening programmes: the development of a monitoring and evaluation system. Br J Cancer 59:954–958, 1989

(4) Morrison AS: Screening in Chronic Disease, 2nd ed. New York: Oxford Univ Press, 1992

(5) Peer PGM, Holland R, Hendriks JHCL, et al: Age-specific effectiveness of the Nijmegen population-based breast cancer-screening program: assessment of early indicators of screening effectiveness. J Natl Cancer Inst 86:436–441, 1994

(6) Shapiro S, Venet W, Strax P, et al: Periodic Screening for Breast Cancer: the Health Insurance Plan Project and Its Sequelae, 1963-1986. Baltimore: Johns Hopkins Univ Press, 1988

(7) Miller AB, Baines CJ, To T, et al: Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years [published erratum appears in Can Med Assoc J 148:718, 1993]. Can Med Assoc J 147:1459–1476, 1992

(8) Chalmers TC: Screening for breast cancer: what should national health policy be [editorial]? J Natl Cancer Inst 85:1619–1621, 1993

(9) Harris RP, Fletcher SW, Gonzalez JJ, et al: Mammography and age: Are we targeting the wrong women? A community survey of women and physicians. Cancer 67:2010–2014, 1991

Note

Manuscript received February 7, 1994; accepted February 9, 1994.

REFERENCE REAGENTS FOR MURINE AND HUMAN CYTOKINES

The Biological Response Modifiers Program (NCI), the Division of Microbiology and Infectious Diseases (NIAID), and the National Institute for Biological Standards and Control (United Kingdom) have made available reference reagents for murine and human cytokines. The reagents are available in small amounts (approx. 1 μg/sample) for use in the calibration of *in vitro* bioassays and in-house standards only and are not to be used for experimental purposes.

HUMAN REFERENCE REAGENTS CURRENTLY AVAILABLE:

IFN-α	IL-1-α	IL-3	IL-6	G-CSF	TGF-β1
IFN-β	IL-1-β	IL-4	IL-7	GM-CSF	TNF-α
IFN-γ	IL-2	IL-5	IL-8	M-CSF	TNF-β
		LIF			

MURINE REFERENCE REAGENTS:

IFN-α	IL-3
IFN-β	IL-4
IFN-γ	GM-CSF
TNF-α	

TO OBTAIN THESE REAGENTS, CONTACT:

Dr. Craig W. Reynolds
Biological Response Modifiers Program
NCI-FCRDC
Building 1052, Room 253
Frederick, MD 21702-1201
FAX: 301-846-5429

Shipments will be made collect express. Please allow 3 to 4 weeks for delivery.

Factors associated with breast structure in breast cancer patients. Cancer 43:1895–1899, 1979
(22) Arthur JE, Ellis IO, Flowers C, et al: The relationship of "high risk" mammographic patterns to histological risk factors for development of cancer in the human breast. Br J Radiol 63:845–849, 1990
(23) Siegel S: Nonparametric Statistics. New York: McGraw-Hill, 1956
(24) Petersen OW, Höyer PE, van Deurs B: Frequency and distribution of estrogen receptor-positive cells in normal, nonlactating human breast tissue. Cancer Res 47:5748–5751, 1987
(25) Kaufman Z, Garstin WI, Hayes R, et al: The mammographic parenchymal patterns of women on hormonal replacement therapy. Clin Radiol 43:389–392, 1991
(26) Berkowitz JE, Gatewood OM, Goldblum LE, et al: Hormonal replacement therapy: mammographic manifestations. Radiology 174:199–201, 1990
(27) Spicer DV, Pike MC, Pike A, et al: Pilot trial of a gonadotropin hormone agonist with replacement hormones as a prototype contraceptive to prevent breast cancer. Contraception 47:427–444, 1993
(28) Lemay A, Maheux R, Faure N, et al: Reversible hypogonadism induced by a luteinizing hormone-releasing hormone (LH-RH) agonist (Buserelin) as a new therapeutic approach for endometriosis. Fertil Steril 41:863–871, 1984
(29) Henzl MR, Kwei L: Efficacy and safety of nafarelin in the treatment of endometriosis. Am J Obstet Gynecol 162:570–574, 1990
(30) Friedman AJ: Treatment of leiomyomata uteri with short-term leuprolide followed by leuprolide plus estrogen-progestin hormone replacement therapy for 2 years: a pilot study. Fertil Steril 51:526–528, 1989

Notes

Affiliations of authors: D. V. Spicer, A. Pike (Department of Medicine), G. Ursin, M. C. Pike (Department of Preventive Medicine), Y. R. Parisky, J. G. Pearce (Department of Radiology), D. Shoupe (Department of Obstetrics and Gynecology), Kenneth J. Norris Jr. Comprehensive Cancer Center, University of Southern California School of Medicine, Los Angeles.

Supported by generous donations from the following contributors to the Kenneth J. Norris Jr. Comprehensive Cancer Center's research funds: the Ida Miles Foundation, the Candle Foundation, the Firestein/Gertz Gift, and the Permanent Charities.

Drs. Pike and Spicer are associated with Balance Pharmaceuticals, Inc., a company set up to develop the contraceptive regimen discussed here.

We wish especially to thank the women who so public-spiritedly volunteered to be in this study and who have been quite remarkable in their commitment to this research.

Manuscript received August 25, 1993; revised December 10, 1993; accepted December 23, 1993.

Age-Specific Effectiveness of the Nijmegen Population-Based Breast Cancer-Screening Program: Assessment of Early Indicators of Screening Effectiveness

*Petronella G. M. Peer, Roland Holland, Jan H. C. L. Hendriks, Marcel Mravunac, André L. M. Verbeek**

Background: **The benefits of mammographic screening for breast cancer are not clear for women less than 50 years old.** ***Purpose:*** **Our aim was to evaluate the effectiveness of breast cancer screening in different age groups.** ***Methods:*** **A mammographic breast cancer-screening program with a 2-year screening interval has been under way in Nijmegen, The Netherlands, since 1975. After eight rounds, more than 40000 women have been invited to participate. All breast cancer cases diagnosed in the invited population, whether detected by screening or clinically, have been considered in this study. The age groups are younger than 50 years, 50-69 years, and 70 years or older at last invitation before diagnosis. Our assessment of the effectiveness of screening is based on the proportion of screen-detected cancers among all cancers and on the disease stage at diagnosis.** ***Results:*** **For women younger than 50, compliance was 75%, and 37% (85/230) of the cancers were detected by screening. The age group 50-69 years had a compliance of 65% with 48% (288/595) of cancers detected by screening, and the group 70 years or older had a 25% compliance with 35% (108/305) of cancers detected by screening. The sizes of the cancers detected by screening were smaller than those detected clinically for all age groups. The age group under 50 showed no substantial difference in the proportion of positive axillary lymph node status between screening-detected and clinically diagnosed cancers, while among older women, the proportion of lymph node involvement was substantially higher for clinically detected cases. A significantly lower frequency of advanced stages was observed in screen-detected compared with clinically diagnosed cancers for women 50 years of age or older ($P<.001$) but not in women under 50 ($P = .35$).** ***Conclusions:*** **No positive effect of the biennial screening program is apparent for women under age 50. For women aged 50 and above, the screen-detected cancers have a more favorable stage distribution than clinically diagnosed cancers, a prerequisite for a reduction in breast cancer mortality. [J Natl Cancer Inst 86:436–441, 1994]**

Since the 1960s, several screening programs for breast cancer detection have been carried out in North America and Western Europe. For women aged 50-69 years, the results in general consistently show benefits of mammographic screening for breast cancer, such as a better stage distribution at diagnosis and a subsequent reduction in breast cancer mortality (*1-5*). In women younger than 50 years, however, the benefits of screening are not as clear. So far, none of the trials (*5-7*) have been able to show convincing evidence of a benefit for women in this age group.

These discouraging results might be explained by a follow-up that is too short. They may also be explained by other factors, such as frequency of screening, sensitivity of the mammographic screening test, less aggressive treatment offered to women with mammographically detected breast cancers, and the biological nature of the tumor (*8*).

**Correspondence to:* Petronella G. M. Peer, M.S., Department of Epidemiology, Medical Informatics, and Statistics, University of Nijmegen, P. O. Box 9101, 6500 HB Nijmegen, The Netherlands.

See "Notes" section following "References."

Although breast cancer mortality is the ultimate evaluation measure, early measures may be applied to determine the potential effectiveness of a screening program, inasmuch as the effect on mortality can only be assessed long after the start of the screening program.

An important initial measure of the effectiveness of the screening program is the proportion of cancers detected by screening among the total number of cancers, which is partly influenced by the compliance rate. Another initial measure of the screening program's effectiveness is the disease stage at diagnosis. Patients with cancers detected by screening are expected to benefit from the screening program in that they will present with an earlier disease stage, a prerequisite for any reduction in breast cancer mortality (*9*).

In this study, these two early indicators of effectiveness of screening (i.e., the proportion of cancers detected at screening among the total number of cancers and the stage distribution of the cancers) were evaluated in an ongoing program in women of three age groups. These groups consisted of women younger than 50 years, 50-69 years old, or 70 years old or older at last invitation to participate in the screening program before breast cancer diagnosis.

The present analysis is based on the results of eight rounds of biennial screening in Nijmegen, The Netherlands. It comprises all cancers detected or diagnosed in the period from 1975 to 1992.

Study Population and Methods

In the city of Nijmegen (150000 inhabitants), a population-based biennial screening program for breast cancer was set up in 1975 (*10*). Women were invited to participate in the program by birth cohort. In the first screening round, all women aged 35-64 years at January 1, 1975 (n = 23000), were invited to participate. In the subsequent screening rounds, older women (n = 7700) were also invited. Since round 5, additional women who had reached the age of 40 were invited to participate.

Single-view mammography was carried out as the only screening modality every 2 years. In rounds 1-3, a single lateral view of each breast was taken. From round 4 onward, a mediolateral–oblique view was taken instead of a single lateral view. By the end of 1990, eight screening rounds had been carried out; round 9 started in 1991.

The present study includes all breast cancer cases that were diagnosed in the invited population, before they received the invitation to participate in round 9, among both participants and nonparticipants. Patients with lobular carcinomas in situ were not included, because they are not treated as breast cancer patients. The effect of screening was assessed in the whole target population, not just in those who participated in the screening. Since 1975, data on all Nijmegen patients with a primary breast cancer have been carefully recorded by the local cancer registry of the Department of Pathology and the Department of Diagnostic Radiology of University Hospital, University of Nijmegen. This data file has been merged with the regional cancer registry of the Comprehensive Cancer Center East (Integraal Kankercentrum Oost, Nijmegen) and the National Computerized Archive of Pathology Reports (Landelijk Geautomatiseerd Pathologisch Anatomisch Archief, Amsterdam, The Netherlands).

Among the program participants, cancers either could be detected at screening (n = 481) or could be diagnosed in the interval between a screening examination that gave negative results and the subsequent scheduled screening, the so-called "interval cancers" (n = 280). A cancer was classified as a "nonparticipant cancer" if it was diagnosed in a woman after she had received an invitation to participate in the screening and had chosen not to and before she had received an invitation to the next screening (n = 369).

We considered two early indicators of screening effectiveness on future breast cancer mortality. The first indicator was the proportion of cancers detected at screening. We distinguished between detection at screening after the first invitation and after a later invitation. The relatively high detection rate at first invitation would not be representative of the ongoing screening situation.

The second indicator was the stage distribution of the cancers. Tumor size and axillary lymph node involvement of screen-detected and clinically diagnosed cancers were compared among women who were in one of the following age groups at last invitation before diagnosis: younger than 50 years, 50-69 years, and 70 years or older.

Only the sizes of invasive lesions were included. The largest diameter of the tumor mass was assessed mammographically in 90% of the cases. If this diameter could not be assessed mammographically because the margins of the tumor were unclear or because the tumor was radiographically occult, the diameter was established on the basis of findings from the histologic slides and the specimen radiograph.

In the early years of the screening program, the axillary lymph nodes were not routinely examined histologically, but the results regarding lymph node status were presented from round 4 onward. In 17 patients with clinically diagnosed cancers, all but two among women aged 50 years or older, no lymph node dissection was performed because the tumor was at an advanced stage and was inoperable. These cases have been recorded as node positive.

The benefits of screening will come mainly from the improvements in prognosis seen in screen-detected cancers. Because of length-time–biased sampling, the aggressiveness of cancers detected at a first screening is expected to be different from that seen at later screenings. Breast cancers detected at a screening examination subsequent to a previous screening examination that showed negative results give a better indication of any screening effect. That is why a distinction was made between breast cancers detected at a first screening and those detected at successive screenings. A cancer detected at screening in a woman who did not participate in the previous round was classified as "detected at first attendance."

To determine the statistical significance of differences in proportions and of interactions, statistical analyses of contingency tables were performed (*11*). A two-sided significance level of .05 was used. The statistical tests used were based on likelihood ratio statistics. In the case of two-dimensional contingency tables, these tests were similar to the familiar Pearson chi-square tests. Differences in the relationships among the different age groups were also tested with likelihood ratio statistics as described in (*11*).

Results

After eight rounds, more than 40000 women have been invited to participate. A total of 18000 women received their first invitation to participate in the screening while they were younger than 50 years of age, 15000 women were between 50 and 69 years of age at first invitation, and 7000 women were 70 years of age or older. In round 8 (1989-1990), 5500 of the invited women were still younger than 50 years of age, 15000 were between 50 and 69 years of age, and 8800 were 70 years old or older. The attendance rate stabilized at approximately 70% for women younger than 50 years of age and at about 60% for women aged 50-69 years, while in the group of women aged 70 years or older, less than 25% of the invited women participated in the latest rounds (rounds 4-8).

The proportion of cancers detected at screening, including in situ and invasive cancers, differed significantly between the age groups ($P<.001$). Table 1 shows that only 37% of the breast cancers diagnosed in women under 50 years of age were detected at screening, 38% of the cancers were classified as interval cancers, and 25% were classified as nonparticipant cancers. In the 50- to 69-year-old age group, 48% of the cancers were detected at screening, 26% were classified as interval cancers, and 26% were classified as nonpartici-

Table 1. Number of women screened at first or successive invitations and number of breast cancers detected at screening or clinically diagnosed among participants and nonparticipants in the subsequent periods between screening invitations

Screening rounds 1-8	Age at screening invitation, y		
	<50	50-69	≥70
First invitation			
No. of women screened	14852	11842	2460
Attendance rate	81%	78%	34%
Cancer at screening	29 (50%)	67 (66%)	28 (54%)
Interval cancer*	25 (43%)	18 (18%)	6 (12%)
Nonparticipant cancer†	4 (7%)	17 (17%)	18 (35%)
Total	58	102	52
Successive invitations			
No. of women screened	32617	62845	10883
Attendance rate	72%	63%	24%
Cancer at screening	56 (33%)	221 (45%)	80 (32%)
Interval cancer*	63 (37%)	136 (28%)	32 (13%)
Nonparticipant cancer†	53 (31%)	136 (28%)	141 (56%)
Total	172	493	253
Total invitations			
No. of women screened	47469	74687	13343
Attendance rate	75%	65%	25%
Cancer at screening	85 (37%)	288 (48%)	108 (35%)
Interval cancer*	88 (38%)	154 (26%)	38 (12%)
Nonparticipant cancer†	57 (25%)	153 (26%)	159 (52%)
Total	230	595	305

*Cancer diagnosed in the interval between a screening examination that gave negative results and the subsequent scheduled screening.

†Cancer diagnosed in a woman after she had received an invitation to participate in the screening and had chosen not to and before she had received an invitation to the next screening.

pant cancers (Table 1). Despite the low attendance rate in the oldest age group (≥70 years), 35% of the cancers were detected at screening, and the majority (52%) of the cancers in this age group were classified as nonparticipant cancers (Table 1). For all age groups, no improvement in the proportion of screen-detected cancers was seen when we compared the results obtained in the first few rounds with those obtained in the more recent rounds.

Table 2 shows the size of the invasive cancers by method of detection (i.e., screening or clinical diagnosis between two successive scheduled screening visits or among women who declined to participate) in three groups formed according to age at the last invitation to participate before diagnosis. The tumors observed in screen-detected cancers were significantly smaller than those observed in clinically diagnosed cancers. This difference was more pronounced in women 50 years of age or older (P = .06). In the group younger than 50 years, nearly 20% of the cancers detected at a successive screening examination had a diameter of 10 mm or less, while the percentage of such cancers was twice as high in the age group 50 or older. The proportion of clinically diagnosed cancers of 10 mm or less was about 10%, irrespective of age group. A small, but not statistically significant, decrease in the size of screen-detected cancers was seen in the group under age 50 during the period that mediolateral–oblique views were taken compared with the earlier time period when a single lateral view was taken.

The number of ductal carcinomas in situ is also shown in Table 2. The number detected at screening in younger women was significantly greater than that detected at screening in older women (P<.01).

Table 3 shows the axillary lymph node status of screen-detected and clinically diagnosed cancers from round 4 onward. There was a marked difference between the age groups. In women 50 years of age or older, about one fifth of the cases detected at a successive screening showed axillary lymph node involvement, which was significantly lower than in clinically diagnosed cancers (P<.001). In the group younger than 50, nearly 40% of the cases detected at a successive screening showed axillary lymph node involvement, which was not significantly different from that found in clinically diagnosed cancers (P = .38).

The size of the tumor and lymph node status were combined to define stage of disease (*9*). Stage I cancer was defined as a tumor with a diameter of 20 mm or less and showing no involvement of the lymph nodes. When the lymph nodes were involved or when the tumor diameter was larger than 20 mm, the cancer was classified as stage II+. Table 4 shows the stage distribution of invasive cancers. A significantly lower frequency of advanced stages was observed in screen-detected compared with clinically diagnosed cancers in women 50 years of age or older (P<.001) but not in women in the younger age group (P = .35).

For all age categories, the biopsy rates were highest in the initial screening round and decreased during the program (toward round 8). For the group under age 50, rates declined from 7.3 per 1000 women screened to 4.1; for those aged 50-69 years, from 11.1 to 3.3; and for those 70 years old or older, from 13.2 to 7.3.

Discussion

Women with screen-detected cancers are expected to benefit from a screening program. Accordingly, a first indicator of the potential benefit associated with screening is the proportion of screen-detected cancers. In our study, in women younger than 50 years of age, 37% of the cancers were screen detected; in women 50-69 years old, 48% were screen detected; and in women 70 years of age or older, 35% were screen detected. In other studies (*1*,*2*,*4*,*12*), the percentages of cancers detected at screening were considerably higher: between 60% and 70%. These higher percentages can be partly explained by a higher attendance rate of 80% or more (*4*,*12*) or a more frequent screening schedule (*1*,*2*). The percentage of screen-detected cancers in the oldest age group in our study was relatively high (35%), considering the low participation rate (overall,

Table 2. Invasive breast cancer cases by tumor size from screening rounds 1 to 8

Breast cancer*	Age at last screening invitation before diagnosis, y† <50	50-69	≥70
Screen-detected			
First attendance‡			
Invasive tumor size, mm			
≤10	5 (19)	27 (29)	8 (17)
11-20	19 (70)	49 (52)	24 (51)
>20	3 (11)	18 (19)	15 (32)
Total	27	94	47
DCIS	10 (27)	6 (6)	7 (13)
Successive attendances			
Invasive tumor size, mm			
≤10	6 (19)	64 (42)	20 (42)
11-20	19 (61)	71 (46)	23 (48)
>20	6 (19)	19 (12)	5 (10)
Total§	31	154[1]	48
DCIS	17 (35)	33 (18)	6 (11)
Clinically diagnosed			
Interval cancer‖			
Invasive tumor size, mm			
≤10	11 (14)	16 (11)	7 (20)
11-20	36 (44)	68 (47)	19 (54)
>20	34 (42)	61 (42)	9 (26)
Total§	81[1]	145[1]	35[1]
DCIS	6 (7)	8 (5)	2 (5)
Nonparticipant cancer¶			
Invasive tumor size, mm			
≤10	4 (8)	15 (10)	7 (5)
11-20	24 (46)	38 (26)	39 (27)
>20	24 (46)	93 (64)	98 (68)
Total§	52[2]	146[2]	144[10]
DCIS	3 (5)	5 (3)	5 (3)

*DCIS = ductal carcinoma in situ.
†Values in columns = number of cases (%).
‡Includes screen-detected patients who did not participate in the previous round of screening.
§Superscript numbers indicate the number of invasive cancers of unknown sizes.
‖Cancer diagnosed in the interval between a screening examination that gave negative results and the subsequent scheduled screening.
¶Cancer diagnosed in a woman after she had received an invitation to participate in the screening and had chosen not to and before she had received an invitation to the next screening.

25%) in this age group. This result might be explained by self-selection of women with an elevated risk for breast cancer, such as familial history. Another explanation may be that a portion of these cancers would never have been detected during the patient's life if no screening had been carried out (overdiagnosis).

A prerequisite for a reduction in breast cancer mortality is a more favorable stage distribution in screen-detected cancers compared with clinically diagnosed cancers. This condition was fulfilled in all age groups with regard to tumor size at presentation. Although mammographically detected cancers were smaller than those detected by palpation, this difference was less apparent for younger women than for older women. The proportion of small tumors among the screen-detected cancers in the age group 50-69 years old in our study was comparable to that of other studies (*4,13,14*).

Fewer cases with axillary lymph node metastases were observed among the cancers detected at successive screenings compared with clinically diagnosed cases in women aged 50 years or more but not in those under age 50. About 20% of the screen-detected cancers in the group of women 50-69 years old showed axillary lymph node metastases in our study; this percentage is comparable to that reported in other studies (*4,13,14*).

Approximately one third of the screen-detected invasive cancers in the women 50 years of age or older were advanced (stage II+), which is still acceptable according to minimum criteria for screening performance (*9*). However, in the group of women younger than 50, as many as two thirds had cancers in an advanced stage at diagnosis. Detailed information on the disease stage at diagnosis for the group younger than age 50 has been published for only two screening projects (*7,14*). Our results are consistent with the results of those studies. The results of the only randomized trial published so far (*7*), in part specifically designed for women aged 40-49 years, showed no reduction in the cumulative occurrence of large or lymph node-positive tumors in the study group compared with the control group.

The benefit of mammographic screening seems to differ in young and older women. The reduction in size observed in screen-detected cancers compared with clinically diagnosed cancers was accompanied by fewer cancers with axillary lymph node metastases only in the women 50 years old or older. Cross-classification of tumor size and lymph node status (not tabulated here) showed that 40% of younger women with very small invasive lesions (≤10 mm) had positive nodes—a considerable percentage. In women 50 years of age or older, the percentage was 20%. Advancing the time of diagnosis does not seem to alter the propensity for axillary lymph node invasion in women younger than 50 years of age. A different tumor biology in older and younger women might explain this finding. If in younger women breast cancers behave as an early systemic disease, i.e., they disseminate early in their evolution before they are detectable by current screening methods, advancing the time of diagnosis will have little effect on the development of metastasis (*15*). In older women, breast cancer may behave as a progressive disease, which renders the time of diagnosis more relevant to the question of whether or not metastasis has already occurred (*16*).

The as yet unanswered question regarding ductal carcinoma in situ is the likelihood of its progression to invasive cancer. Ductal carcinomas in situ are detected more frequently at screening of younger women. If progressive, this process may take many

Table 3. Breast cancer cases by axillary lymph node status from screening rounds 4 to 8

Breast cancer	Age at last screening invitation before diagnosis, y		
	<50	50-69	≥70
Screen-detected			
First attendance*			
Negative node†	8	18	18
Positive node	5 (38%)	9 (33%)	6 (25%)
Total‡	13	27[1]	24[2]
Successive attendances			
Negative node†	19	101	34
Positive node	12 (39%)	28 (22%)	9 (21%)
Total‡	31	129	43[4]
Clinically diagnosed			
Interval cancer§			
Negative node†	26	63	18
Positive node	28 (52%)	30 (32%)	8 (31%)
Total‡	54	93[3]	26[5]
Nonparticipant cancer‖			
Negative node†	22	51	40
Positive node	16 (42%)	57 (53%)	50 (56%)
Total‡	38[2]	108[1]	90[32]

*Includes screen-detected patients who did not participate in the previous round.
†Ductal carcinoma in situ cases have been included as negative.
‡Superscript indicates the number of cancers with unknown axillary node status.
§Cancer diagnosed in the interval between a screening examination that gave negative results and the subsequent scheduled screening.
‖Cancer diagnosed in a woman after she had received an invitation to participate in the screening and had chosen not to and before she had received an invitation to the next screening.

Table 4. Invasive breast cancer cases by stage from screening rounds 4 to 8

Breast cancer	Age at last screening invitation before diagnosis, y		
	<50	50-69	≥70
Screen-detected			
First attendance*			
Stage I	3	14	11
Stage II+	5 (63%)	12 (46%)	9 (45%)
Total†	8	26[1]	20[1]
Successive attendances			
Stage I	4	74	26
Stage II+	13 (76%)	32 (30%)	13 (33%)
Total†	17	106	39[3]
Clinically diagnosed			
Interval cancer‡			
Stage I	17	39	12
Stage II+	34 (67%)	49 (56%)	12 (50%)
Total†	51	88[3]	24[5]
Nonparticipant cancer§			
Stage I	13	24	14
Stage II+	22 (63%)	81 (77%)	82 (85%)
Total†	35[2]	105	96[22]

*Includes screen-detected patients who did not participate in the previous round.
†Superscript numbers indicate the number of invasive cancers with unknown stage.
‡Cancer diagnosed in the interval between a screening examination that gave negative results and the subsequent scheduled screening.
§Cancer diagnosed in a woman after she had received an invitation to participate in the screening and had chosen not to and before she had received an invitation to the next screening.

years, and a long delay is to be expected before the benefit of mammography, in terms of a reduction in breast cancer mortality, becomes apparent.

Our data substantiate the hypothesis that it is more difficult to detect invasive breast cancer in its premetastatic stage in younger women than in older women.

References

(1) Andersson I, Aspegren K, Janzon L, et al: Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. BMJ 297:943–948, 1988

(2) Roberts MM, Alexander FE, Anderson TJ, et al: Edinburgh trial of screening for breast cancer: mortality at seven years. Lancet 335:241–246, 1990

(3) Frisell J, Eklund G, Hellström L, et al: Randomized study of mammography screening—preliminary report on mortality in the Stockholm trial. Breast Cancer Res Treat 18:49–56, 1991

(4) Tabár L, Fagerberg G, Duffy SW, et al: Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol Clin North Am 30:187–210, 1992

(5) Nyström L, Rutqvist LE, Wall S, et al: Breast cancer screening with mammography: overview of Swedish randomised trials. Lancet 341:973–978, 1993

(6) Day NE: Screening for breast cancer [published erratum appears in Br Med Bull 47:786, 1991]. Br Med Bull 47:400–415, 1991

(7) Miller AB, Baines CJ, To T, et al: Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years [published erratum appears in Can Med Assoc J 148:718, 1993]. Can Med Assoc J 147:1459–1476, 1992

(8) Breast cancer screening in women under 50 [editorial]. Lancet 337:1575–1576, 1991

(9) Day NE, Williams DR, Khaw KT: Breast cancer screening programmes: the development of a monitoring and evaluation system. Br J Cancer 59:954–958, 1989

(10) Peeters PH, Verbeek AL, Hendriks JH, et al: Screening for breast cancer in Nijmegen. Report of 6 screening rounds, 1975–1986. Int J Cancer 43:226–230, 1989

(11) Bishop YMM, Feinberg SE, Holland PW: Discrete Multivariate Analysis: Theory and Practice. Cambridge, Mass: MIT Press, 1975

(12) Frisell J, Eklund G, Hellström L, et al: The Stockholm breast cancer screening trial—5-year results and stage at discovery. Breast Cancer Res Treat 13:79–87, 1989

(13) Anderson TJ, Lamb J, Donnan P, et al: Comparative pathology of breast cancer in a randomised trial of screening. Br J Cancer 64:108–113, 1991

(14) Paci E, Ciatto S, Buiatti E, et al: Early indicators of efficacy of breast cancer screening programmes. Results of the Florence District Programme. Int J Cancer 46:198–202, 1990

(15) Fisher B: Laboratory and clinical research in breast cancer—a personal adventure: the David A. Karnofsky Memorial Lecture. Cancer Res 40:3863–3874, 1980

(16) Tabár L, Fagerberg G, Day NE, et al: Breast cancer treatment and natural history: new insights from results of screening. Lancet 339:412–414, 1992

Notes

Affiliations of authors: P. G. M. Peer, A. L. M. Verbeek, Department of Epidemiology, Medical Informatics, and Statistics, University of

SPECIAL ARTICLE

Report of the International Workshop on Screening for Breast Cancer

*Suzanne W. Fletcher, William Black, Russell Harris, Barbara K. Rimer, Sam Shapiro**

Background: **Over the past 30 years, eight major randomized controlled trials of breast cancer screening—with mammography and/or clinical breast examination—have been conducted. Results from several trials have been updated during the past year, and initial results of three other trials have been reported.** *Purpose:* **The National Cancer Institute held an International Workshop on Screening for Breast Cancer in February 1993 to conduct a thorough and objective critical review of the world's most recent clinical trial data on breast cancer screening, consider the new evidence, assess the current state of knowledge, and identify issues needing further research.** *Methods:* **Investigators representing the eight randomized controlled trials of breast cancer screening in women aged 40-74 presented published and unpublished data. Evidence relating to the effectiveness of breast cancer screening in different age groups, especially women aged 40-49, was presented.** *Results:* **For women aged 40-49, randomized controlled trials consistently demonstrated no benefit from screening in the first 5-7 years after study entry. A meta-analysis of six trials found a relative risk of 1.08 (95% confidence interval = 0.85-1.39) after 7 years' follow-up. After 10-12 years of follow-up, none of four trials have found a statistically significant benefit in mortality; a combined analysis of Swedish studies showed a statistically insignificant 13% decrease in mortality at 12 years. Only one trial (Health Insurance Plan) has data beyond 12 years of follow-up, and results show a 25% decrease in mortality at 10-18 years. Statistical significance of this result is disputed, however. In women aged 50-69, all studies show mortality reductions; three of four studies show reductions of about 30% at 10-12 years after study entry. Results from two of these trials were statistically significant. Too few women over age 70 have been included in studies for adequate analysis.** *Conclusions:* **For women aged 40-49, randomized controlled trials of breast cancer screening show no benefit 5-7 years after entry. At 10-12 years, benefit is uncertain and, if present, marginal; thereafter, it is unknown. For women aged 50-69, screening reduces breast cancer mortality by about a third. Currently available data for women age 70 or older are inadequate to judge the effectiveness of screening.** *Implications:* **Randomized trials have provided stronger scientific evidence regarding the effectiveness of screening for breast cancer than for any other cancer. However, much still needs to be learned. Periodic gatherings of scientists in the field should speed the process. [J Natl Cancer Inst 85:1644–1656, 1993]**

Breast cancer is the most common nonskin cancer in women in the United States and several other Western countries. Except for lung cancer, breast cancer causes more deaths than any other cancer in women (*1*). During 1993, approximately 182000 women in the United States will be diagnosed with breast cancer, and an estimated 46000 will die of the disease (*2*). These facts alone make breast cancer an important health issue for all women, health care workers, and medical scientists.

The risk of developing breast cancer increases markedly with age; more than 75% of breast cancers occur in women who are older than 50 (*3*). The annual risk of developing breast cancer is approximately one in 3700 for a woman 30-34 years old, but it is one in 235 for a woman 70-74 (*4*). Therefore, although the lifetime risk of breast cancer is high, one in eight (*1*), most women are not at high risk of developing the disease while they are young.

The incidence of breast cancer varies considerably from country to country (*5*). In the United States, the reported incidence has increased steadily since the 1940s, particularly during the past decade (*2,6*). Many physicians and researchers believe that most of the recent reported increase reflects a more widespread use of mammography rather than

*See "Notes" section following "References."

a true increase in incidence (*7*). Meanwhile, mortality rates have been stable (*1,2*).

Methods to prevent breast cancer from occurring have not yet been established, and scientific efforts have concentrated on developing new and improved ways to treat the cancer once it occurs. In addition, efforts have focused on early detection of breast cancer in women through screening so that ensuing treatment can be more effective and save more lives.

Studies of breast cancer screening—including mammography, clinical breast examination, and breast self-examination—have been conducted for at least 30 years. During this period, in addition to several case–control and cohort studies, eight major randomized controlled trials of breast cancer screening have been conducted (Tables 1 and 2) (*8-41*), an unparalleled accomplishment in research on cancer detection. The randomized trials alone have included nearly 500000 women. After early results of several of the trials were reported, organizations and expert groups around the world began making recommendations for breast cancer screening at the population level (*42-50*). In general, these groups have agreed that scientific evidence supports breast cancer screening with mammography alone or mammography and clinical breast examination in women aged 50 years or older. On the other hand, recommendations regarding breast cancer screening for women 40-49 years old have varied, with some groups recommending screening and others not. Most groups have not made specific recommendations about women older than 70. Other areas of uncertainty or disagreement among recommendations include the frequency of screening and the screening tests to be used (mammography, clinical breast examination by health professionals, and/or breast self-examination). With the appearance of breast cancer screening recommendations, screening practices have become increasingly common in several countries.

Over the past year, several randomized clinical trials of breast cancer screening have been updated; initial results of three other trials (two in Canada and one in Göteborg, Sweden) have also been reported. Finally, a meta-analysis (*41*) was recently completed to analyze the impact of screening on women ages 40-49 in the randomized trials. As a result of these developments, the National Cancer Institute (NCI) convened an International Workshop on Screening for Breast Cancer on February 24 and 25, 1993, in Bethesda, Maryland. The goals were to consider the new evidence, to conduct a thorough and objective critical review of the world's clinical trial data on breast cancer screening, to assess the current state of knowledge, and to identify issues needing further research.

The workshop was not intended to produce recommendations for breast cancer screening, and none have been made. Because the workshop was open to the public, with lay and press people in attendance, the NCI hoped that its presentations and discussions would help medical scientists, practitioners, and the public at large to gain a better understanding of the current evidence on the effectiveness of breast cancer screening.

The workshop focused on the scientific evidence regarding effectiveness of breast cancer screening, especially in women 40-49 years of age, but also in other age groups of women. In addition, participants evaluated the scientific evidence regarding the performance characteristics of the three common screening tests: mammography, clinical breast examination, and breast self-examination. The NCI is aware of many other important scientific issues relating to breast cancer screening, including cost-effectiveness of various screening strategies and development and evaluation of more effective implementation of breast cancer screening, but these issues were not emphasized in the workshop. This article summarizes the evidence presented by investigators and discussants at the workshop. Evidence previously re-

Table 1. Selected characteristics of the design and conduct of eight randomized controlled trials of breast cancer screening

Study (year begun)	Age at entry, y	Screening modality*	Periodicity, mo	Randomization	Sample size		% screened at 1st examination
					Study	Control	
HIP (1963)	40-64	2-view MM + CBE	12	Individual	30239	30756	67
Sweden							
Two-county (1977)	40-74	1-view MM	24 (age <50 y) 33 (age ≥50 y)	Cluster: geographic	78085	56782	89
Malmö (1976)	45-69	2-view MM	18-24	Cluster: birth cohort	21088	21195	74
Stockholm (1981)	40-64	1-view MM	28	Cluster: birth cohort	39164	19943	81
Göteborg (1982)	40-59	2-view MM	18	Individual (age 40-49 y) Cluster (age 50-59 y)	20724	28809	84
Edinburgh (1976)	45-64	2-view MM + CBE initially (later, usually 1-view MM)	12 24	Cluster: physician	23226	21904	61
Canada NBSS1 (1980)	40-49	2-view MM + CBE	12	Individual: volunteers	25214	25216	~100†
Canada NBSS2 (1980)	50-59	2-view MM + CBE versus CBE only	12	Individual: volunteers	19711	19694	~100†

*CBE = clinical breast examination; MM = mammography.

†Study design included randomization of volunteers after clinical breast examination; accordingly, virtually 100% had their first screening examination.

Table 2. Age at entry, duration of follow-up, and relative risk for deaths due to breast cancer, Study group compared with control group: eight randomized controlled trials

Trial	Age at entry, y	Duration of follow-up, y	Relative risk (95% confidence interval)	
			All ages	Younger than 50 y
HIP	40-64	10	0.71 (0.55-0.93)*	0.77 (0.50-1.16)
Sweden				
Kopparberg†	40-74	12	0.68 (0.52-0.89)	0.75 (0.41-1.36)
Ostergotland†	40-74	12	0.82 (0.64-1.05)	1.28 (0.76-2.33)
Malmö	45-69	12	0.81 (0.62-1.07)	0.51 (0.22-1.17)
Stockholm	40-64	8	0.80 (0.53-1.22)	1.04 (0.53-2.05)
Göteborg	40-59	7	0.86 (0.54-1.37)	0.73 (0.27-1.97)
All centers	40-74	7-12	0.76 (0.66-0.87)	0.87 (0.63-1.20)
Edinburgh‡	45-64	10	0.84 (0.63-1.12)	0.86 (0.41-1.80)
Canada				
NBSS1§	40-49	7		1.36 (0.84-2.21)
NBSS2‖	50-59	7	0.97 (0.62-1.52)	

* 18 years from entry, relative risk is 0.77 (95% confidence interval = 0.61-0.97).
† Two-county Swedish trial.
‡ Preliminary data presented at the International Workshop on Screening for Breast Cancer, February 24-25, 1993.
§ Comparison is between annual mammography plus clinical breast examination versus clinical breast examination at initial screening, subsequently no examination.
‖ Comparison is between annual mammography plus clinical breast examination versus annual clinical breast examination alone.

ported in the literature (*see* reference list) and unpublished information provided by the workshop participants have been incorporated into this presentation.

In assessment of the effect of screening, results from randomized controlled trials are preferred, because, if performed properly, randomized controlled trials avoid biases that are difficult to control in other types of studies and help assure the comparability of groups. The NCI workshop, therefore, primarily considered results from eight randomized trials of breast cancer screening. Special attention was given to the quality of these trials in terms of 1) whether randomization resulted in comparable groups of women in the study and control groups, 2) the degree of compliance in the group of women offered screening, 3) the amount of "contamination" in the control group of women not offered screening (i.e., the percentage of the control group of women who received breast cancer screening), 4) the quality and frequency of the screening test, 5) the adequacy of follow-up of women in both the screening and control groups, 6) the validity of outcome ascertainment, 7) the magnitude of difference in mortality outcomes between women offered screening and those not offered screening and whether any differences seen could have occurred by chance alone, and 8) the generalizability of the results to other groups of women.

Evidence relating to the performance characteristics of the three different screening tests was examined. Mammography was emphasized because of the concentration in more recent studies on this detection modality. The most important performance characteristic of a screening test is its accuracy. Measures related to accuracy include sensitivity, specificity, and positive and negative predictive value; the evidence regarding these properties was examined. Positive predictive value for breast cancer screening tests assumes special significance because of the emotional reaction of women to breast cancer and the potential negative effects of false-positive tests. With this in mind, information about the rate of screening examinations with false-positive results, which lead to further investigations including biopsy, was sought. It is difficult to measure sensitivity directly, so we examined different ways to measure this test performance characteristic.

Clinical Trials

This section includes a discussion of the data from randomized trials for women in three age groups: 40-49, 50-69, and 70 or older. The data, which are summarized in Tables 1-4, are derived from published reports, updated information presented at the workshop (much of which is in press), and discussion at the workshop. The section concludes with a summary of the evidence for women in the three age groups.

Randomized Trials in Women Aged 40-49 Years

Overview of trials. The best evidence for determining the value of breast cancer screening is derived from randomized trials in which a group of women offered screening is compared with a group not offered screening. The following major randomized trials included women aged 40-49: the Health Insurance Plan (HIP) (*8-10*), Swedish two-county (*11-16*), Malmö (*17,18*), Stockholm (*19-23*), the first Canadian National Breast Screening Study (NBSS1) (*32*), Edinburgh (*34-36*), and Göteborg (*40*) (Table 1). Each varied in terms of frequency of screening, duration of follow-up, compliance, and the screening test (e.g., mammography alone or mammography and clinical breast examination with or without training in breast self-examination).

Of these trials, only the HIP study showed a possible benefit for women aged 40-49 years. This benefit occurred 10-18 years after entry in the study, with a 25% decrease in mortality from breast cancer; statistical significance of the result is disputed (*8,10*). Moreover, the benefit was not

Table 3. Relative risk for deaths due to breast cancer. Study group compared with control group in women aged 50 y and older at entry—HIP, Sweden, and Edinburgh

Trial Age at entry, y	Duration of follow-up, y	Relative risk (95% confidence interval)
HIP	10	
50-64		0.68 (0.49-0.96)
Swedish centers combined*	12	
50-59		0.71 (0.57-0.90)
60-69		0.71 (0.56-0.91)
70-74		0.94 (0.60-1.46)
Edinburgh†		
50-64	10	0.85 (0.63-1.14)

*Data presented at the International Workshop on Screening for Breast Cancer, February 24-25, 1993, and represent the findings in the Swedish two-county study and Malmö combined.

†Preliminary data presented at the International Workshop on Screening for Breast Cancer, February 24-25, 1993.

evident until 7-8 years of follow-up had been completed. The combined Swedish trials showed a 13% decrease in mortality for study group women aged 40-49 after 12 years of follow-up, but the difference from mortality in the control group was not statistically significant (Tables 2 and 4) (*40*). In a meta-analysis of the results from five randomized trials after 7 years of follow-up, Elwood et al. (*41*) concluded that the "overall {breast cancer mortality} ratio in the intervention to the control groups was 0.99, indicating virtually identical breast cancer death rates, and no observable benefit from screening in younger women. The data are consistent with a range of odds ratios from 0.74 to 1.32, that is from a 26% reduction to a 32% increase in mortality in association with screening." When the 7-year results of the Canadian trial were included, the relative risk was 1.08 (95% confidence interval = 0.85-1.39). The meta-analysis did not include the Göteborg trial and did not take into account the recent re-analysis of the combined Swedish trials. Thus, the randomized trials are consistent in showing no benefit at 5-7 years after entry, an uncertain and marginal benefit at 10-12 years, and an unknown benefit thereafter.

The negative consequences of screening (e.g., false negatives and false positives) are more frequent in women aged 40-49 than in women aged 50-69 years. For example, the Swedish two-county trial showed the sensitivity of mammography to be 60% in younger women, compared with 86% in older women (*14*). Several of the trials found an early, small increase in mortality for the study group women, although this increase was not statistically significant (*41*).

The major elements of the randomized trials are summarized in Tables 1, 2, and 4. One valid criticism of some of the trials is that they may have had insufficient power to allow researchers to conduct adequate subgroup analyses of women aged 40-49 years and, especially, analyses of women aged 40-44 and 45-49. However, the hypothesis that best fits the available data for women younger than 50 is that screening in this age group has no effect. The meta-analysis by Elwood et al. (*41*) showed a best estimate of no effect and a wide range of uncertainty in both directions—wide enough to include all of the derived results of the trials, both positive and negative.

Recruitment and randomization. Methods for recruitment and randomization have varied from trial to trial. In the HIP trial, 62000 age-eligible women were randomly allocated to the control or study (screening) group (*8*). Letters of invitation were sent to those allocated to the study group, and efforts were made to encourage participation. In the Swedish two-county trial, sampling was by clusters—all women within a defined geographic area were invited to participate and were compared with women in a noninvited geographic community (*11*). Women in the Malmö trial were randomly assigned by birth cohort into control and study groups and then invited to participate through letters (*17*). In the Stockholm trial, cluster sampling was used, and then women were randomly assigned individually by birth date. Written invitations were sent to every woman in the study group, and follow-up letters were sent 3 months later if no reply was received (*21,22*). Göteborg researchers used a similar strategy, except that women aged 40-49 years were randomly assigned by individuals and those aged 50-59 by clusters (*40*). The Edinburgh trial used a form of cluster randomization in which general practices were stratified by size; then, all of the women registered with each practice were randomly assigned to the study or control group. Women in the study group were invited by letter (*34*). This study is the one trial in which significant differences were found in socioeconomic status between the study and the control groups, and subsequent analyses attempted to control for this difference by an adjustment for socioeconomic status (*36*).

Table 4. Swedish trial results: woman-years of exposure, numbers of deaths due to breast cancer, and relative risk, Study group compared with control group in women aged 40-49 y at entry*

Screening center	Woman-years		Deaths due to breast cancer		Relative risk (95% confidence interval)
	Study	Control	Study	Control	
Kopparberg†	107000	56000	26	18	0.75 (0.41-1.35)
Ostergotland†	104000	106000	24	19	1.28 (0.70-2.33)
Malmö‡	46000	47000	8	16	0.51 (0.22-1.17)
Stockholm	107000	64000	20	12	1.04 (0.53-2.05)
Göteborg	64000	77000	6	10	0.73 (0.27-1.92)
All centers	428000	350000	84	75	0.87 (0.63-1.20)

*Data presented at the International Workshop on Screening for Breast Cancer, February 24-25, 1993.

†Swedish two-county study.

‡Ages 45-49.

The randomization scheme for the Canadian trial was different from those for the other randomized trials. In each of the other trials, population-based lists of women were assembled, and women were randomly assigned to the study or control groups. Women in the study group were then invited by personal letter to participate. In the Canadian study, on the other hand, volunteers were recruited through a variety of methods, including personal letters, mass media campaigns, and presentations to groups in the community. Thus, volunteer women were first recruited in each of 15 study centers. All eligible volunteers were given a clinical breast examination by a nurse or physician and then randomly assigned according to a plan that stratified eligible women by 5-year age groups within centers. A local coordinator used lists with preprinted identification numbers and group designations. Because more women who were later found to have advanced cancers were initially assigned to the study group, questions have been raised as to possible problems with the randomization process. However, no clustering was observed in numbers of advanced cancer cases by screening center (Miller A: personal communication).

Overall, the randomization schemes in the trials were effective in producing similar groups of women. Only the Edinburgh trial had a clear imbalance of the arms with respect to socioeconomic status (*34*).

Compliance. As Table 1 demonstrates, compliance in the randomized trials varied considerably. At the first screening, the proportion of women who were screened ranged from 61% in the Edinburgh trial (*35*) and 67% in the HIP trial (*8*) to 74% in the Malmö trial (*17*), 81% in the Stockholm trial (*21*), and 89% in the Swedish two-county trial (*11*). Compliance in the Canadian trials for the first screen was, by definition, almost 100%, because volunteers were used (*32,33*). These rates of participation are impressive. In all cases, the measure of effect of screening on mortality reduction included the experiences of the participants and of the nonparticipants (i.e., those who refused screening). Low compliance would diminish differences between the study and control groups, if any, but such an effect is unlikely to have been large enough to alter these studies' findings. Furthermore, compliance rates for women aged 40-49 were at least as good as those for women aged 50-69.

Contamination. With the exception of the HIP study, contamination is a problem to a greater or lesser extent in all of the trials. Screening among women in the control group would attenuate the impact of the intervention and decrease any differences in outcome between the study and control groups. Each of the randomized trials varied in the extent to which contamination occurred. In the Stockholm and Edinburgh trials, the amount of screening outside the trial was not known. In others, considerable screening occurred outside the trial. For example, in the Malmö trial, 24% of the women in the control group overall and 35% of those aged 45-49 had mammograms outside of the study (some for screening and some for diagnostic reasons) (*17*). Contamination was inherent in the Swedish two-county trial design because the control group was invited to participate in screening 3-6 years after the start of the trial (*14*). In Canada, about 7% of the women in the study group and 26% of those in the control group were screened outside of the trial (*32*). One and a half percent of the control group received diagnostic mammograms at the time of entry into the study. In the HIP trial, few women were screened outside the trial, because the secular trend to obtain a mammogram that is so clearly apparent in the United States today had not yet begun.

Quality and frequency of breast screening. Great variation in the quality and frequency of breast screening exists among the screening modalities included (Table 1) in the trials. All trials had mammography, but they differed in the number of views of each breast. The HIP trial included two views at each examination, Malmö and Edinburgh had two views in at least the initial examination, and the Swedish two-county and Stockholm studies had one-view mammograms at all examinations. Tabar et al. (*16*) have concluded, based on an analysis of the interval cases in the Swedish two-county trial, that a contributing factor to no benefit being found in that trial was the decision to use only one-view mammography among women 40-49 years old at entry. Table 1 summarizes the screening modality and screening interval, which ranged from 12 to 33 months, for all trials.

The HIP, Canadian, and Edinburgh trials included routine clinical breast examinations. Except for the Canadian study, little information is available regarding the quality of the performance of the tests. In the HIP trial, most of the examinations were conducted by surgeons, the rest by internists. All examiners were instructed by a well-trained surgeon, and the procedure was monitored (*8*). In the Edinburgh trial, physicians and nurses were used (*36*). A study of the nurses' performance is under way. Observations in the Canadian study suggest high-quality performance by nurse examiners who performed clinical breast examinations in 12 of the 15 centers, following careful training (*29,31*).

In assessing the quality of mammography, only the Canadian study's investigators reported results of external reviewers (*26-28*). The reviews of mammogram films have been cited as evidence of poor quality mammography in the Canadian trial, at least in the early years of the study (*51*).

Advancement in the quality of mammograms differentiates the study environment during the 1960s and early 1970s (when the HIP study was performed) from that of later years when all of the other trials were under way. Major improvements occurred in mammography around 1985, and further improvements occurred over the next several years. For example, before 1985, grids were not widely used. In 1985, 59% of radiologists reported using grids, with increases to 86% in 1988 and 95% in 1992 (*52*). In judging the likely benefit from mammography screening, an important question should be raised: Are results from a trial involving many radiologists who meet acceptable standards for community practice but have varying degrees of proficiency in obtaining quality films and interpreting them more informative than the results from a trial in which a uniform degree of excellence is achieved? From this standpoint, criticism of the Canadian trial regarding quality of mammography in the early stages of the study may have less of an impact on the interpretation of the results than has

been advanced by some critics (*51*).

Adequacy of follow-up. Data from the trials are available for varying periods of follow-up from entry, thereby limiting the ability to combine study results (Table 2). Several of the studies provide findings for up to 12 years of follow-up. The HIP study provides follow-up information up to 18 years from entry. Analysis of these data suggests that longer periods of follow-up may be needed than are available in reports from the other trials (*53*). Furthermore, given the year-to-year changes in measures of screening effectiveness, average periods of exposure may not provide a satisfactory basis for drawing conclusions about the effectiveness of screening women who are 40-49 years old at time of entry into the program. It has been suggested that, due to lower than expected mortality from breast cancer, at least 10 years' follow-up is required for 80% power to detect a mortality decrease as large as 37% (Ederer F: personal communication).

Determination of outcome. Agreement is uniform that mortality from breast cancer, and not survival among cases detected, is the primary measure of outcome to use in evaluating the effectiveness of screening at all ages. Two other issues are also important. The first is age at diagnosis. The basic analysis in a randomized controlled trial is age at entry into the trial. For long-term follow-up of screening, a secondary analysis might address age at diagnosis. The second issue relates to determining the underlying cause of death. The HIP and Canadian studies conducted blinded reviews of the cause of death from death records and from physician, hospital, and autopsy records early in the trials; researchers in Sweden have recently conducted such a review for the Stockholm, Malmö, Göteborg, and Swedish two-county trials (*40*). During this review, several changes in classification were apparently made. It is important that, for each study, the results and changes in classification (initial versus final classification) be reported numerically. A criticism of the Canadian NBSS trials has been that deaths in the control group of women may have been ascertained less completely than those for the study group.

Other measures, such as case survival rates and detection rates of advanced cancers, can only be used to suggest the impact that screening might have on mortality. Survival rates are limited by the healthy volunteer effect, lead time, and length bias and cannot be used directly to measure changes in mortality. Stage of disease—late stage (measured by rate per unit population) in particular—is useful, but it is necessary to consider the likelihood that in the first round of screening, and possibly in the second or third, screening will lead to detection of more breast cancers at all stages compared with detection in the absence of screening.

Magnitude of difference. When the magnitude of difference is weighed, two questions must be considered: 1) How large an effect do the trials aim to find? 2) What is the effect found? All but the Canadian NBSS planned to test for an effect over a wide age range, for example, ages 40-64 (HIP) and 40-74 years (Swedish two-county). The power of the samples was designed to detect reductions in breast cancer mortality of 20%-35%, usually at a reasonable level of certainty. In most cases, the time interval over which the reduction in mortality would occur was not clearly specified.

All trials have produced data by age subgroups, but in all cases, the subgroup results had wide confidence intervals. It was speculated that even if no single trial was large enough to produce a statistically significant differential between the study and control groups in a particular age subgroup, a pattern across studies might emerge that would increase confidence in drawing conclusions about the effectiveness of screening.

The Canadian NBSS1 was the only trial specifically designed to test effectiveness of breast cancer screening in women aged 40-49 years. Because of the large improvements in mammography since the HIP study, the study was designed to detect a 40% differential with a high level of confidence (i.e., a power of 90%) (*32*). This percentage has proven to be high because the mortality rate for the control group was derived from national data rather than from a cohort of women with no history of breast cancer. Further, the calculation did not account for the fact that all women, in both the study and control groups, received an intervention at the initial screening examination.

Generalizability. The European studies and the HIP trial were strengthened by the availability of registries of all age-eligible women who could be invited to participate. Similarly, the Swedish two-county trial and the Malmö and Stockholm trials were population-based studies with participation rates ranging from 74% to 89% in women aged 40-49. Thus, the ability to generalize from these trials is extremely high. Generalizability of the Edinburgh trial is affected by socioeconomic status differences that were found between the study and control groups. A reservation cited for the HIP trial is the fact that large improvements have occurred in the capabilities of mammography to detect breast cancer since the HIP trial was conducted.

The most important difference between the Canadian study and the other trials is the use by the Canadian study of volunteers who selected themselves to participate, whereas the other trials used a natural aggregation of women who were assigned randomly. However, the radiology facilities used in the Canadian trial were those already existing in the community, and the results are more likely to be generalizable than are results from a study performed under difficult-to-achieve ideal conditions.

All of the trials differ from usual care in the United States in that they were conducted in systems in which cost was not a barrier to participation and in which the costs of follow-up were covered by the national health care system. In the United States, not all residents are covered by insurance, and the extent of coverage for diagnostic work-ups varies from plan to plan. In fact, if compliance was higher in these trials than would be expected in the general U.S. population, the benefit seen in the trials would be an overestimate.

Trial results for women aged 40-49 through the first 7 years of follow-up are remarkably consistent even though the studies were conducted at different times, in different countries, with varying screening intervals, and with one- or two-view mammograms. This consistency adds weight to the evidence of the effect of screening through the first 7 years of follow-up. For longer periods of follow-up, some trials

show a benefit, while others demonstrate a negative effect. What is apparent is that none of the studies, with the possible exception of the HIP trial, showed a statistically significant benefit for women younger than 50.

Randomized Trials in Women Aged 50-69 Years

The most convincing evidence for the benefit of mammographic screening has been shown for women aged 50-69 (Table 3). In the Swedish trials, at 12 years from entry, the relative risk for breast cancer death in the study group compared with that in the control group was 0.71 for women aged 50-69. In the HIP trial, the odds ratios were 0.69 for women aged 50-59 and 0.67 for women aged 60-64; 18 years after entry, these figures changed to 0.77 and 0.83, respectively. Results from other trials for shorter periods, as brief as 7 years, are favorable for women aged 50 years or older. Because women in the control group received yearly clinical breast examinations, the Canadian study cannot measure the effect of screening on mortality for women aged 50 or older. Thus, the evidence from the randomized trials consistently shows a substantial benefit of screening: Mortality is decreased about 30% for women aged 50-69 at 10-12 years after study entry. This finding holds despite the fact that only the Swedish two-county results and the HIP results for women aged 50-59 met a test of statistical significance.

Effectiveness of Screening in Women Aged 70 Years or Older

Most trials have had an upper age limit of 64 years. Two have gone beyond this age: The Swedish two-county trial included women up to age 74, and the Malmö study included women up to age 69. In the former, at ages 70-74, there were only two screenings, and the evidence related to mortality was inconclusive (relative risk = 0.94; 95% confidence interval = 0.60-1.64) (*40*). Few women 70 or older participated; therefore, the numbers were insufficient for adequate analysis. The authors concluded that subgroup differences could not be detected when women aged 40-74 were considered but that power was not adequate to answer questions about the benefit of screening for women older than 70.

In the United States, screening decreases after age 65, although incidence of breast cancer continues to increase (*54*,*55*). Unfortunately, data from the international trials are inadequate to draw a conclusion regarding the benefit of screening women older than age 70. The National Institutes of Health Breast Screening Forum urged that decisions regarding screening for older women be based on the woman's general health, be made jointly by the woman and her physician, and reflect a consideration of comorbid disease or other medical conditions (*56*).

Summary

The randomized trials of women aged 40-49 are consistent in showing no statistically significant benefit in mortality after 10-12 years of follow-up. This is true even with a recent meta-analysis, a combined analysis of four Swedish studies based on more current data, and updated data from the Edinburgh trial.

In contrast, results for women aged 50-69 in all randomized trials confirmed the benefit that was first seen in the HIP trial. Elwood et al. (*41*) concluded from an analysis of five of the six published studies with 7 years of follow-up that a 34% reduction in mortality had been achieved among women in the study groups, with reductions ranging from 20% in Edinburgh to 43% in Stockholm. It is not clear at exactly what age the benefit is achieved. Further specification of the age at which benefit occurs would be useful.

Too few data are available regarding the effectiveness of screening for women older than 70 years. Further research is needed to determine whether an age threshold exists beyond which screening ceases to be a benefit. In any case, the woman's physiologic status should be considered in making screening recommendations.

Intermediate Measures: Performance of Screening Tests

Reduction of morbidity and mortality from breast cancer is the ultimate goal of breast cancer screening. To this end, it is useful to examine screening test performance in a standardized, systematic manner. Measures of the accuracy of screening tests and of procedures performed and cancers detected (e.g., biopsy rate, positive predictive value, rate of prevalent cancers detected at first screening, prevalence ratio, and percentage of early- or late-stage cancers) can help compare screening programs across trials, thus providing insight into which screening program is best. These measures also are useful in comparing screening programs in randomized trials with those in community practice.

Intermediate measures may also help determine when advances in a screening test make a new randomized controlled trial desirable; unless the advance can be shown to result in an important increase in sensitivity or cancer detection, it is unlikely that a new trial will show a reduction in mortality. Measures such as sensitivity are complex to calculate and are affected by factors other than the test itself. Some of these factors include whether the test is performed during the first (prevalence) or later (incidence) screening round, whether the patient population is volunteer or geographically defined, the degree to which the study population is aware of breast cancer, whether single or multiple screening modalities are used, whether carcinoma in situ is counted as a detected cancer, and the age distribution of the screened population. These factors, combined with the fact that some studies have not reported these data, limit our ability to answer important questions from the randomized controlled trials of breast cancer screening.

Screening test performance and other intermediate measures reported in randomized trials of breast cancer screening are listed in Tables 5-8. In several trials, more than one screening test was used; it is difficult to disentangle the separate characteristics and effects of tests used in tandem. Thus, we have calculated performance measures for the

Table 5. Definitions of screening test performance and other intermediate measures

Measure	Definition	Comments
Sensitivity (detection method)	Ratio of screen-detected cancers to screen-detected plus interval cancers.	May overestimate sensitivity. Includes cancers with lead times longer than 1 year.
Sensitivity (incidence method)	1 minus ratio of interval rate (in screening group) to incidence rate in controls.	May underestimate sensitivity. Excludes cancers with lead times longer than 1 year.
Prevalence	Number of prevalent cancers at first screening per 1000 screening exams.	Dependent on prevalence of cancer and sensitivity of screening.
Prevalence ratio	Ratio of prevalence at first screening to incidence in controls.	Indirect measure of sensitivity (detection method).
Specificity (work-up)	Ratio of true negative to true negative plus false-positive screening exams.	Positive with respect to requiring further work-up.
Specificity (biopsy)	Ratio of true negative to true negative plus false-positive screening exams.	Positive with respect to requiring surgical biopsy.
Positive predictive value (work-up)	Ratio of cancer cases to all cases requiring further evaluation in screened group.	Most dependent on specificity and prevalence.
Positive predictive value (biopsy)	Ratio of cancer cases to all cases requiring surgical biopsy in screened group.	Most dependent on specificity and prevalence.
Benign to malignant ratio	Ratio of benign to malignant biopsies.	Most dependent on specificity and prevalence.
Biopsy rate	Number of screening exams requiring biopsy per 1000 exams.	Dependent on sensitivity and specificity of screening and prevalence of cancer.

Table 6. Estimates of screening test performance and other intermediate measures in seven randomized controlled trials

Screening test or intermediate measure	Trial (reference No.)*							
	HIP (8)	Malmö (17,63)	Swedish two-county (11,14)	Edinburgh (35,37)†	Stockholm (19,21)	Canada NBSS1 (ages 40-49 y) (32)†	Canada NBSS2 (ages 50-59 y) MM + CBE (33)†	Canada NBSS2 (ages 50-59 y) CBE only (33)†
Sensitivity‡ (detection method)	74	79	76	88	68§	81	88	63
Sensitivity‖ (incidence method)	77	68	60-70	79	52-75¶	58	72	46
Prevalence** (per 1000 at first screen)	2.72	7.5	6.0	6.2	4.0	3.89	7.20	3.45
Prevalence ratio**	1.39	2.8	2.9	3.2	2.82	2.74	3.51	1.68
Specificity†† (work-up)	98.5	96-97	95-98	96-97	95-97	82-93	83-96	89-96
Specificity†† (biopsy)	99.8	NA	99.4-99.9	98.7-99.9	99.8-99.9	96.6-98.8	96.5-99.2	99.1-99.7
Positive predictive value (work-up)	12	10-22	12	4-15	8-10	2	4-6	2-4
Positive predictive value (biopsy)	20	33-61	50-75	36-64	62-78	12	20	28
Biopsy rate (per 1000 exams)	10	NA	4-12	3-12.5	4-6	15-38	12-42	4-12
Benign to malignant ratio	3.9	NA	0.33-1	0.56-1.8	0.28-0.62	7.3	2.7-4.8	1.6-3.3
% in situ (entire study group)	12.8	15.8	8.4	10.1	10.5	18.9	16.7	5.8
% stage I (entire study group)	56	50	53	32	65	NA	NA	NA
% node-negative (entire study group)	57	NA	71	63	NA	65	69	67

*CBE = clinical breast examination; MM = mammography; NA = not applicable. Note that ranges pertain to different rounds within each trial.
†Estimates are based on the screening process as a whole. The individual contributions of MM and CBE are not included.
‡Based on screen-detected and interval cancers from all rounds.
§Based on first round only.
‖Based on rate of interval cancers in study group and incidence rate in control group from all rounds.
¶Reported separately for different intervals after previous screening.
**Dependent on prevalence, spectrum of cancer in population, and sensitivity of screening.
††Based on each round.

Table 7. Estimates of screening test performance among women aged 40-49 y at entry in three randomized trials

Screening test performance	Swedish two-county	Stockholm	Canada NBSS1
Sensitivity (detection method)	62	53*	81
Sensitivity (incidence method)	45-62†	39	58
Prevalence (per 1000 at first screening)	2.09	2.06	3.89
Prevalence ratio	1.99	1.74	2.74

*Based on first round only.
†Range for 1st- and 2nd-year interval after previous screening.

Table 8. Estimates of screening test performance among women aged 50-59 y at entry in three randomized trials*

Screening test performance	Swedish two-county	Stockholm†	Canada NBSS2 (MM + CBE)	Canada NBSS2 (CBE only)
Sensitivity (detection method)	76	73‡	88	63
Sensitivity (incidence method)	37-83§	61	72	46
Prevalence (per 1000 at first screening)	4.67	4.84	7.20	3.45
Prevalence ratio	2.50	2.93	3.51	1.68

*CBE = clinical breast examination; MM = mammography.
†Age ranged from 50 to 64 y at entry.
‡Based on first round only.
§Reported separately for round and different intervals after previous screen (*14*).

entire screening program in each trial, not for each test alone (Tables 6-8).

Sensitivity of breast cancer screening differs depending on how it is calculated. When screen-detected cases are in the numerator and all cases are in the denominator (detection method, Table 5), sensitivity varies from 63% to 88% (Table 6). When the rate of detected cases is inferred from the difference between the expected incidence and the interval cancer rate (incidence method, Table 5), sensitivity varies from 46% to 79% (Table 6), depending on the age of the population, the modality of screening, and the round of screening.

When measured under comparable conditions, the sensitivity of breast cancer screening was lower for women aged 40-49 years than for women older than 50 (Tables 6 and 7), probably because of the relative lack of contrast between cancer and normal glandular tissue with both mammography and clinical breast examination. More rapid growth of cancers in younger women is another possible reason.

The data in Tables 6-8 may underestimate sensitivity for more recent studies relative to earlier ones, because breast cancer awareness in the population has increased, thus increasing the likelihood that cancers would be detected by women themselves as "interval" cancers. Breast cancer awareness may also vary among nations, thus affecting comparisons between countries. The degree to which this factor affects reported sensitivity estimates is unknown. More recent trials show an approximately twofold increase in prevalence at first screening and in prevalence ratio, compared with results in the HIP study. This finding probably indicates an improvement in cancer detection in the more recent trials. However, the increase also could be due to an increase in the prevalence of breast cancer in the general population, a change in pathologic criteria for cancer, the detection of cancers that would never have caused clinical disease, or a combination of these factors.

The specificity of screening depends heavily on how a positive test result is defined. When "positive" is defined as referral for any further evaluation, specificity ranges from about 82% to 98%. When "positive" is defined as having a surgical biopsy, specificity is consistently higher, ranging from 96.6% to 99.9%. Positive predictive values for any further evaluation vary little across studies, but the positive predictive value for biopsy varies greatly, from 20% in the HIP trial and the Canadian NBSS1 trial of younger women to more than 70% in the Swedish two-county and Stockholm trials (Table 6). This result may reflect a greater tendency to perform biopsies in North America, perhaps because of the pressures of malpractice litigation. Biopsy rates were highest in the Canadian NBSS studies and lowest in the Swedish studies. At the biopsy rates of the Canadian NBSS2 study, assuming that false-positive results are randomly distributed, an individual woman would have a 10%-20% chance of having a biopsy at some time during 5 years of screening with annual mammography and clinical breast examination.

The pathologic spectrum of cancers detected has varied among the studies, although not all studies have reported either stage or nodal status. The percentage of tumors that were carcinomas in situ was highest in the Canadian NBSS studies. The natural history of this pathologic finding is not clear. However, it may be that many carcinomas in situ do not progress to cause clinical disease (*57*). Intermediate measures should be calculated both with and without carcinomas in situ, a practice seldom performed.

Several important questions can be addressed with these data:

1) Has breast cancer screening improved over time?

Some have argued that breast cancer screening techniques have markedly improved over time, and, thus, previous studies are irrelevant with respect to the extent of mortality reduction with current methods. To examine this question, it is important first to determine whether the sensitivity and size and stage of the detected cancers have improved over the course of the trials (i.e., higher prevalences at first screening, higher prevalence ratios, and lower interval cancer rates). Although cancer detection appears higher in the studies that followed the HIP trial, sensitivity is not consistently better (Tables 6-8). It is uncertain to what extent recent improvements in mammography or clinical breast examination have resulted in increased sensitivity.

Another issue is whether gains in screening test sensitivity are translated directly into reductions in breast cancer mortality. If the sensitivities in Table 6 are compared with the relative risks in Table 2, no clear relationship between the two is found. Perhaps the best example of this is the Canadian NBSS2 trial. Although the sensitivities of the two programs varied from 46% to 72%, breast cancer mortalities at 7 years' follow-up were identical (*33*). This result suggests that gains in screening test sensitivity do not necessarily result in reduced mortality. Improvement in the sensitivity of screening invariably improves survival from

ime of diagnosis but does not necessarily reduce mortality, because of lead time and length biases. Therefore, a new randomized trial is necessary to determine whether an improved and more sensitive screening program would lead to fewer breast cancer deaths.

2) Is one modality of screening better than another?

Although all eight breast cancer screening trials included mammography in their screening programs, only four included a clinical breast examination and two, breast self-examination. In the HIP and Edinburgh studies, the quality and technique of clinical breast examinations were incompletely described. In three other studies, clinical breast examination was combined with mammography, and direct measures of sensitivity, specificity, and cancer detection rates are not possible. In the Canadian NBSS2 trial, on the other hand, sensitivity (with the detection method) can be calculated for women aged 50-59. Sensitivity of clinical breast examination at first screen was 81%; for subsequent screening rounds, sensitivity was 46%-68%. In an earlier report from this same study (*58*), specificity was low (88%-96%), with positive predictive values of 2.9%-4.2%.

At 7 years' follow-up, the Canadian NBSS2 study suggests that, for women aged 50-59, mammography does not decrease breast cancer mortality beyond the reduction achieved through well-performed clinical breast examinations. The effectiveness of clinical breast examination alone in reducing mortality from breast cancer has not been tested directly.

No randomized study has directly tested the effectiveness of breast self-examination in reducing mortality from breast cancer (*59-62*); thus, this screening test was not emphasized in the workshop. A nonrandomized trial (*39*) has shown confusing but generally negative results, and an uncontrolled study (Gastrin G, et al.: Incidence and mortality from breast cancer in the MAMA program for breast screening in Finland, 1973-1986. Manuscript submitted for publication) has shown some evidence of benefit. However, a case–control study (*61*) has shown that breast self-examination as it is usually conducted is of no benefit. Randomized clinical trials under way in Russia and Shanghai will likely provide more information on the effectiveness of breast self-examination, although the results may not be applicable to North America because these trials do not address the question of whether breast self-examination contributes any mortality reduction beyond that achieved by mammography and clinical breast examination.

The sensitivity of breast self-examination has not been directly measured, but an upper limit can be calculated by assuming that all interval cases in the trials were detected by breast self-examination, an approach that probably overestimates true sensitivity (*62*). The sensitivity of breast self-examination then becomes one minus the sensitivity of the screening program, a figure that would range between 12% and 25% for the detection method and, ignoring the extremes, between 20% and 50% for the incidence method.

3) Is single-view mammography as good as double-view?

The HIP, Malmö, and Canadian NBSS trials used two-view mammography. The Edinburgh trial used two-view mammography in its initial screen, but one-view thereafter. In the Swedish two-county and Stockholm trials, only single-view mammography was used. When these trials are compared, it is not clear that the number of mammographic views had much impact on sensitivity or on effectiveness. In any case, there are trade-offs between sensitivity and specificity. Further research is necessary to determine the effect of adding the craniocaudal to the mediolateral oblique projection.

4) Is a longer interval between screening rounds as good as a shorter interval?

The interval between screening rounds has varied in the trials from 12 months (HIP and Canadian NBSS studies) to 33 months (Swedish two-county study). Again, it is not clear that studies with shorter intervals had greater sensitivity (Tables 1 and 6). Because the effectiveness of screening women aged 50 or older varied little across the studies, there is little evidence to show that shorter screening intervals in these women were more effective in reducing breast cancer mortality.

Summary and Future Research Directions

Breast Cancer Screening in Women Aged 40-49 Years

The most important issue addressed during the workshop is the effectiveness of screening women aged 40-49 years. For this age group, it is clear that in the first 5-7 years after study entry, there is no reduction in mortality from breast cancer that can be attributed to screening. There is an uncertain and, if present, marginal reduction in mortality at about 10-12 years. Only one study (HIP) provides information on long-term effects beyond 12 years, and more information is needed.

More research results should be available soon. Continued analysis from the combined Swedish experience is needed. The control groups of women in several of the Swedish studies are now being screened, which may limit the conclusions that can be drawn about long-term effectiveness. The degree to which outcomes were reclassified when individual trials were combined needs to be clarified.

Continued follow-up of the Canadian trial should also help determine whether and to what degree long-term effects from breast cancer screening occur among women aged 40-49. It is worrisome that more patients in the screening group had advanced tumors, and this fact may be responsible for the results reported to date. Detailed review of the randomization procedures at each center would be helpful to determine with as much certainty as possible whether any breach in protocol occurred. The technical quality of mammography early in this trial is of concern, but it is not clear to what extent mammography quality affected the study outcome. Combined sensitivity of mammography and clinical breast examination in the Canadian trial was at least as good as that of mammography in the Swedish trial.

Follow-up in the Edinburgh trial should be useful as well, although to date, the number of breast cancer deaths in women younger than 50 is small. The large randomized controlled trial of breast cancer screening starting at age 40, which is now in progress in the United Kingdom

(Chamberlain J: personal communication), may provide the most definitive evidence about the effectiveness of breast cancer screening in women younger than 50 years. It is unlikely that large randomized clinical trials for women 40-49 years of age will be mounted elsewhere in the Western world.

A second meta-analysis of the data from all available trials of screening in women aged 40-49 may be useful, especially when longer follow-up is available and when the effect of reclassification is clarified in the combined Swedish studies. Such a meta-analysis should use the raw data from each of the trials.

Breast Cancer Screening in Women Aged 50-69 Years

For women aged 50-69, the evidence presented at the Workshop strengthens the scientific observation that screening leads to reduced breast cancer mortality. Every study presented found a protective effect for women in this age group. The combined analyses of the four Swedish trials provide increased precision of the estimated mortality reduction seen in the individual trials. That analysis and the meta-analysis by Elwood et al. (*41*) raise our confidence that screening women aged 50-69 reduces mortality 30%-35%. The results of the Canadian study of women aged 50-59 do not contradict those from the Swedish studies, because the Canadian study addressed a different question: the incremental effect of adding mammography to a careful physical examination when screening for breast cancer.

Three important questions relating to screening in women aged 50-69 remain unanswered. First, what is the optimal interval for mammography? The Swedish studies suggest that a screening mammogram as infrequent as every 33 months reduces breast cancer mortality, at least in a population with a high compliance rate and in a setting with high-quality mammography. Estimates of lead time in this age group range from 21 months in the HIP study to 42 months in the Swedish two-county study. These data raise the possibility that a screening interval of 12 months may not be necessary in this population. Further studies to determine the optimal screening interval would be useful and are being planned in the United Kingdom.

Second, what is the most effective screening modality? Previous studies have demonstrated that mammography and physical examination detect breast cancer in a complementary manner. Initial results from the Canadian study suggest that a careful clinical breast examination may be as effective in decreasing breast cancer mortality as mammography, although more follow-up is needed. Research should be undertaken to determine (*a*) which components of the clinical breast examination correlate with early cancer detection and (*b*) how to acquire the necessary clinical skills.

Finally, is a single-view mammogram as effective as a two-view mammogram in women of this age group? Trials varied in the number of views they used. Further research should be carried out to determine the degree to which adding the craniocaudal projection to the mediolateral oblique projection changes intermediate measures such as sensitivity and specificity.

Because the trials we reviewed provide clear evidence of a difference in the effectiveness of breast cancer screening between younger and older women, further studies to determine the mechanisms for this difference should assume high priority. Possible biological reasons, such as menopause, should be studied.

Breast Cancer Screening in Women Aged 70 Years and Older

Women in their 70s are at high risk for breast cancer. The currently available clinical trial data for these women are inadequate to judge the effectiveness of screening, because the numbers of women were small, the compliance was poor, and the screening episodes were too few. Because of the prevalence of screening in North America, it is probably not feasible to conduct a trial in which the control group is not screened. It may be possible to study the effectiveness of different screening intervals of mammography using a randomized controlled trial. Furthermore, clinical breast examination may be particularly effective in older women, and trials should evaluate this possibility.

Other Issues

To date, breast cancer screening has been evaluated in a number of specific age groups, usually by 10-year intervals. These groupings are arbitrary and without biological justification and are used probably because of the relatively small numbers of participants in individual trials. If data from all trials could be combined, it might be possible to examine the effectiveness of breast cancer screening in age groupings as small as 1 year. If so, beginning screening at a less arbitrary age than the beginning of the 5th or 6th decade might be beneficial.

Discussions during the workshop dealt with the vexing problem of age at diagnosis versus age at entry in all the trials, and one analysis was proposed (Alexander F: personal communication). Methodologic research should be undertaken to resolve this issue. The new trial in the United Kingdom plans to include only women aged 40-41 years at entry so that during the course of the study (7 years) no woman will be diagnosed with breast cancer after age 50.

Known risk factors for breast cancer are important but do not provide the basis for selective screening within age groups. New developments in molecular genetics and in the field of biomarkers may lead to much more powerful predictors of breast cancer risk. In the future, it may be possible to identify women who are at such low risk that they will not need to be screened.

An important aspect of breast cancer screening is communication with women regarding breast cancer risks by age and the benefits of breast cancer screening. Presently, perceptions of breast cancer risk are highest in women of lowest risk and lowest in women of highest risk by age group (*64-66*). More effective ways of describing risks to women must be developed. Likewise, better communication techniques are needed when discussing screening abnormalities with women to avoid adverse psychological

sequelae from false-positive examinations.

Throughout the workshop, many issues were raised regarding quality assurance and quality control. In screening, quality is a concern throughout the entire process, from the technical aspects of screening tests, to communication with women, to follow-up of abnormalities, and to compliance with recommendations and screening schedules. Quality assurance and quality control procedures need to be developed and evaluated in all these areas if we are to realize the potential of breast cancer screening.

Standardized data on performance of screening tests should be developed and routinely reported by all trials of breast cancer screening. One model for such calculations is provided by Day et al. (*67*). All trials should report results with and without carcinomas in situ.

Although cost-effectiveness was not the focus of the workshop, updated analyses should be performed as the precision of benefit and costs is clarified in the ongoing trials. Ultimately, the practice of breast cancer screening must take into account both benefits and costs. It is particularly important that such analyses include the cost of working up false-positive results.

Randomized controlled trials have given the world a great deal of information about the effectiveness of breast cancer screening. Because of these trials and the work of the dedicated investigators who directed and participated in them, we have stronger scientific evidence about screening for breast cancer than any other cancer. The 2-day NCI workshop made clear, however, that much work remains. Periodic gatherings of scientists in this field should speed the process and aid in this important work.

References

(*1*) Miller BA, Ries LA, Hankey BF, et al: Cancer Statistics Review: 1973–1989. DHEW Publ No. (NIH)92-2789. Bethesda, Md: National Cancer Institute, 1992

(*2*) Boring CC, Squires TS, Tong T: Cancer statistics, 1993. CA Cancer J Clin 43:7–26, 1993

(*3*) Surveillance, Epidemiology, and End Results: Incidence and mortality data, 1973–77. DHEW Publ No. (NIH)81-2330. Bethesda, Md: Public Health Service, 1981

(*4*) Ries LAG, Hankey BF, Miller BA, et al: Cancer statistics review 1973–88. DHEW Publ No. (NIH)91-2789. Bethesda, Md: National Cancer Institute, 1991

(*5*) Muir C, Waterhouse J, Mack T, et al, eds: Cancer Incidence in Five Continents, vol 5. Lyon: IARC, 1987

(*6*) Harris JR, Lippman ME, Veronesi U, et al: Breast cancer. N Engl J Med 327:319–328, 1992

(*7*) Miller BA, Feuer EJ, Hankey BF: The increasing incidence of breast cancer incidence since 1982: relevance of early detection. Cancer Causes Control 2:67–74, 1991

(*8*) Shapiro S: Periodic Screening for Breast Cancer: the Health Insurance Plan Project and Its Sequelae, 1963-1986. Baltimore: Johns Hopkins Univ Press, 1988

(*9*) Shapiro S, Venet W, Strax P, et al: Ten- to fourteen-year effect of screening on breast cancer mortality. J Natl Cancer Inst 69:349–355, 1982

(*10*) 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 80:1125–1132, 1988

(*11*) Tabar L, Fagerberg CJ, Gad A, et al: Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. Lancet 1:829–832, 1985

(*12*) Holmberg LH, Tabar L, Adami HO, et al: Survival in breast cancer diagnosed between mammographic screening examinations. Lancet 2:27–30, 1986

(*13*) Tabar L, Fagerberg G, Duffy SW, et al: The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. J Epidemiol Community Health 43:107–114, 1989

(*14*) Tabar L, Fagerberg G, Duffy SW, et al: Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol Clin North Am 30:187–210, 1992

(*15*) Tabar L, Fagerberg CG, South MC, et al: The Swedish two-county trial of mammographic screening for breast cancer: recent results on mortality and tumor characteristics. *In* Cancer Screening (Miller AB, Jr, Chamberlain J, Day NE, et al, eds). Cambridge: Cambridge Univ Press, 1991, pp 23–36

(*16*) Tabar L, Fagerberg G, Day NE, et al: What is the optimum interval between mammographic screening examinations? An analysis based on the latest results of the Swedish two-county breast cancer screening trial. Br J Cancer 55:547–551, 1987

(*17*) Andersson I, Aspegren K, Janzon L, et al: Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. BMJ 297:943–948, 1988

(*18*) Janzon L, Andersson I: Malmö mammographic screening trial. *In* Cancer Screening (Miller AB Jr, Chamberlain J, Day NE, et al, eds). Cambridge: Cambridge Univ Press, 1991, pp 637–644

(*19*) Frisell J, Eklund G, Hellstrom L, et al: Analysis of interval breast carcinomas in a randomized screening trial in Stockholm. *In* Breast Cancer Research and Treatment, vol 9. Boston: Martinus-Nijhoff, 1987, pp 219–225

(*20*) Frisell J, Eklund G, Hellstrom L, et al: The Stockholm breast cancer screening trial—5-year results and stage at discovery. Breast Cancer Res Treat 13:79–87, 1989

(*21*) Frisell J, Eklund G, Hellstrom L, et al: Randomized study of mammography screening—preliminary report on mortality in the Stockholm trial. Breast Cancer Res Treat 18:49–56, 1991

(*22*) Frisell J, Glas U, Hellstrom L, et al: Randomized mammographic screening for breast cancer in Stockholm. *In* Breast Cancer Research and Treatment, vol 8. Boston: Martinus-Nijhoff, 1986, pp 45–54

(*23*) von Rosen A, Frisell J, Glas U, et al: Non-palpable invasive carcinomas from the Stockholm screening project. Acta Oncol 28:23–27, 1989

(*24*) Baines CJ: Impediments to recruitment in the Canadian National Breast Screening Study: response and resolution. Controlled Clin Trials 5:129–140, 1984

(*25*) Baines CJ, McFarlane DV, Miller AB: Sensitivity and specificity of first screen mammography in 15 NBSS centres. Can Assoc Radiol J 39:273–276, 1988

(*26*) Baines CJ, McFarlane DV, Miller AB: The role of the reference radiologist. Estimates of inter-observer agreement and potential delay in cancer detection in the National Breast Screening Study. Invest Radiol 25:971–976, 1990

(*27*) Baines CJ, McFarlane DV, Wall C: Audit procedures in the National Breast Screening Study: mammography interpretation. Can Assoc Radiol J 37:256–260, 1986

(*28*) Baines CJ, Miller AB, Kopans DB, et al: Canadian National Breast Screening Study: assessment of technical quality by external review. AJR Am J Roentgenol 155:743–747, 1990

(*29*) Baines CJ, Miller AB, Bassett AA: Physical examination. Its role as a single screening modality in the Canadian National Breast Screening Study. Cancer 63:1816–1822, 1989

(*30*) Baines CJ, To T, Wall C: Women's attitudes to screening after participation in the National Breast Screening Study. A questionnaire survey. Cancer 65:1663–1669, 1990

(*31*) Miller AB, Baines CJ, Turnbull C: The role of the nurse-examiner in the National Breast Screening Study. Can J Public Health 82:162–167, 1991

(*32*) Miller AB, Baines CJ, To T, et al: Canadian National Breast Screening Study. 1. Breast cancer detection and death rates among women ages 40 to 49 years. Can Med Assoc J 147:1459–1476, 1992

(*33*) Miller AB, Baines CJ, To T, et al: Canadian National Breast Screening Study. 2. Breast cancer detection and death rates among women 50 to 59 years. Can Med Assoc J 147:1477–1488, 1992

(*34*) Alexander F, Roberts MM, Lutz W, et al: Randomisation by cluster and the problem of social class bias. J Epidemiol Community Health 43:29–36, 1989

(*35*) Roberts MM, Alexander FE, Anderson TJ, et al: Edinburgh trial of screening for breast cancer: mortality at seven years. Lancet 335:241–246, 1990

(*36*) Roberts MM, Alexander FE, Anderson TJ, et al: The Edinburgh randomised trial of screening for breast cancer: description of method. Br J Cancer 50:1–6, 1984

(*37*) Chamberlain J, Coleman D, Ellman R, et al: Sensitivity and specificity

of screening in the UK trial of early detection of breast cancer. *In* Cancer Screening (Miller AB Jr, Chamberlain J, Day NE, et al, eds). Cambridge: Cambridge Univ Press: 1991, pp 3–17
(*38*) Trial of early detection of breast cancer: description of method. Br J Cancer 44:618–627, 1981
(*39*) UK Trial of Early Detection of Breast Cancer Group: First results on mortality reduction in the UK Trial of Early Detection of Breast Cancer Group. Lancet 2:411–416, 1988
(*40*) Nyström L, Rutqvist LE, Wall S, et al: Breast cancer screening with mammography: an overview of the Swedish randomised trials. Lancet 341:973–978, 1993
(*41*) Elwood JM, Cox B, Richardson AK: The effectiveness of breast cancer screening by mammography in younger women. Online J Curr Clin Trials, 1993.
(*42*) American Cancer Society: Summary of current guidelines for the cancer-related checkup: recommendations. New York: American Cancer Society, 1988
(*43*) National Cancer Institute: Working guidelines for early detection: rationale and supporting evidence to decrease mortality. DHEW: Bethesda, Md: National Cancer Institute, 1987
(*44*) American Medical Association: Mammography screening in asymptomatic women 40 years and older (Resolution 93, 1-87). Report of the Council on Scientific Affairs, Robert F (A-88). Chicago: American Medical Association, 1988
(*45*) American College of Obstetricians and Gynecologists: Committee on Professional Standards. Standards for Obstetric-Gynecologic Services, 6th ed. Washington, DC: American College of Obstetricians and Gynecologists, 1985
(*46*) American College of Radiology: Policy statement: guidelines for mammography. Reston, Va: American College of Radiology, 1982
(*47*) The periodic health examination. 2. 1985 update. Canadian Task Force on the Periodic Examination. Can Med Assoc J 134:724–727, 1986
(*48*) Hayward RS, Steinberg EP, Ford DE, et al: Preventive care guidelines: 1991. Ann Intern Med 114:758–783, 1991
(*49*) Frame PS: A critical review of adult health maintenance. Part 3: Prevention of cancer. J Fam Pract 22:511–520, 1986
(*50*) US Preventive Services Task Force: Guide to Clinical Preventive Services: an Assessment of the Effectiveness of 169 Interventions: Report of the U.S. Preventive Services Task Force. Baltimore: Williams & Wilkins, 1989
(*51*) Kopans DB: The Canadian screening program: a different perspective. AJR Am J Roentgenol 155:748–749, 1990
(*52*) Conway BJ, Suleiman OH, Rueter FG, et al: Does Credentialing Make a Difference in Mammography? Center for Devices and Radiological Health. Presented at the 79th Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, Ill, 1992
(*53*) Smart CR: The role of mammography in the prevention of mortality from breast cancer. Cancer Prevention, pp 1–16, 1990
(*54*) Constanza ME: Breast cancer screening in older women: overview. J Gerontol 47:1–3, 1992
(*55*) Constanza ME, Annas GJ, Brown ML, et al: Supporting statements and rationale. J Gerontol 47:7–16, 1992
(*56*) Screening recommendations of the forum panel. J Gerontol 47:5, 1992
(*57*) Schnitt SJ, Silen W, Sadowsky NL, et al: Ductal carcinoma in situ (intraductal carcinoma) of the breast. N Engl J Med 318:898–903, 1988
(*58*) Baines CJ, Miller AD, Bassett AA: Physical examination. Its role as a single screening modality in the Canadian National Breast Screening Study. Cancer 63:1816–1822, 1989
(*59*) Baines CJ, To T: Changes in breast self-examination behavior achieved by 89,835 participants in the Canadian National Breast Screening Study. Cancer 66:570–576, 1990
(*60*) Morrison AS: Is self-examination effective in screening for breast cancer? J Natl Cancer Inst 83:226–227, 1991
(*61*) Newcomb PA, Weiss NS, Storer BE, et al: Breast self-examination in relation to the occurrence of advanced breast cancer. J Natl Cancer Inst 83:260–265, 1991
(*62*) O'Malley MS, Fletcher SW: US Preventive Services Task Force. Screening for breast cancer with breast self-examination: a critical review. JAMA 257:2196–2203, 1987
(*63*) Hurley SF, Kaldor JM: The benefits and risks of mammographic screening for breast cancer. Epidemiol Rev 14:101–130, 1992
(*64*) Harris RP, Fletcher SW, Gonzalez JJ, et al: Mammography and age: are we targeting the wrong women? A community survey of women and physicians. Cancer 67:2010–2014, 1991
(*65*) Mah Z, Bryant H: Age as a factor in breast cancer knowledge, attitudes and screening behaviour. Can Med Assoc J 146:2167–2174, 1992
(*66*) Baines CJ: Women and breast cancer: is it really possible for the public to be well informed? Can Med Assoc J 146:2147–2148, 1992
(*67*) Day NE, Williams DR, Khaw KT: Breast cancer screening programmes; the development of a monitoring and evaluation system. Br J Cancer 59:954–958, 1989

Notes

Affiliations of authors: S. W. Fletcher, *Annals of Internal Medicine*, American College of Physicians, Philadelphia, Pa.
W. Black, Department of Radiology, Dartmouth-Hitchcock Medical Center, Lebanon, N.H.
R. Harris, University of North Carolina School of Medicine, Chapel Hill.
B. K. Rimer, Duke Comprehensive Cancer Center, Durham, N.C.
S. Shapiro, Department of Health Policy and Management, Johns Hopkins School of Hygiene and Public Health, Baltimore, Md.
Correspondence to: Suzanne W. Fletcher, M.D., American College of Physicians, Independence Mall West, Sixth St. at Race, Philadelphia, PA 19106-1572.

We thank the staff at the National Cancer Institute's Early Detection Branch for technical support; Pamela Fried, Managing Editor, *Annals of Internal Medicine*, for editing; and Torii Ransome and Suzanne Brownholtz for secretarial assistance.
Manuscript received April 29, 1993; revised July 15, 1993; accepted July 28, 1993.

If you're looking for some good reading, you've just found it. The free Consumer Information Catalog.

The Catalog lists about 200 federal publications, many of them free. They can help you eat right, manage your money, stay healthy, plan your child's education, learn about federal benefits and more.

So sharpen your pencil. Write for the free Consumer Information Catalog. And get reading worth writing for.

References

(1) National Cancer Institute: Consensus Development Meeting on Breast Cancer Screening. US Department of Health, Education and Welfare, Public Health Service, National Cancer Institute. DHEW Publ No. (NIH)78-1257, September, 1977

(2) Bailar JC 3d, Smith EM: Progress against cancer? N Engl J Med 314:1226–1232, 1986

(3) Elwood JM, Cox B, Richardson AK: The effectiveness of breast cancer screening by mammography in younger women. Online J Curr Clin Trials (serial online), Feb 25 (Doc No 32), 1993

(4) Fletcher SW, Black W, Harris R, et al: Report of the International Workshop on Screening for Breast Cancer. J Natl Cancer Inst 85:1644–1656, 1993

(5) Chalmers TC: Editorial: mammography after 30 years. Online J Curr Clin Trials May 28, (Doc No 66), 1993

(6) Miller AB, Baines CJ, To T, et al: Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years [published erratum appears in Can Med Assoc J 148:718, 1993]. Can Med Assoc J 147:1459–1476, 1992

(7) Miller AB, Baines CJ, To T, et al: Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years [published erratum appears in Can Med Assoc J 148:718, 1993]. Can Med Assoc J 147:1477–1488, 1992

Notes

Correspondence to: Thomas C. Chalmers, M.D., Medical Research Consulting, 32 Pinewood Village, West Lebanon, NH 03784.

Manuscript received September 17, 1993; accepted September 17, 1993.

Deficiencies in the Analysis of Breast Cancer Screening Data

*Edward A. Sickles, Daniel B. Kopans**

Breast cancer is a major, and increasing, public health problem in the United States and most other industrialized countries. Since there is not yet an established means of preventing breast cancer, efforts to combat the disease have focused on early detection through screening to increase the success of existing treatment. There is general consensus that screening is both safe and effective in women aged 50 and older, but considerable controversy remains over the utility of screening women aged 40-49.

The report by Fletcher et al. published in this issue of the Journal (*1*) is a summary of a breast cancer screening workshop convened by the Early Detection Branch of the National Cancer Institute in February 1993. The vast majority of data presented at the workshop involved only the eight randomized controlled trials of breast cancer screening, despite the availability of an abundance of additional evidence. The article by Fletcher et al. (*1*) reflects this bias by discounting the importance of nonrandomized controlled trial data. Screening data from the eight randomized controlled trials are discussed for women aged 40 and older, with special focus on women in the specific age group 40-49, the period during which screening is most controversial.

Fletcher et al. briefly cite many of the problems that have been raised concerning the validity and applicability of the screening data presented at the workshop. However, the reader is left with the implication that these problems are relatively unimportant, when in fact they severely limit both statistical validity and clinical applicability. In this editorial, we present a much more critical analysis of available screening data and also provide what we believe to be a more balanced overview of current knowledge on the strengths and limitations of screening. Paralleling the article by Fletcher et al. (*1*), we concentrate on the evidence from randomized controlled trials of breast cancer screening in women aged 40-49. The major problems concerning the randomized controlled trial data for this age group include lack of statistical power, other flaws in experimental design, and technical deficiencies in implementation of study protocols. Some specific problems are common to all randomized controlled trials, others to individual randomized controlled trials.

Statistical Power

A randomized controlled trial of breast cancer screening must involve sufficient numbers of women who develop sufficient numbers of breast cancers resulting in sufficient numbers of breast cancer deaths, such that a reduction in the number of deaths by screening can be demonstrated with statistical validity. In designing a trial, planners make a calculation of statistical power that estimates how large the study and control populations should be to demonstrate the anticipated benefit. The smaller the expected benefit, the larger the number of women needed in the trial; conversely, if the number of women in the trial is small, only a very large benefit can be demonstrated. If the number of women is too small to begin with, or if dilution effects from noncompliance (refusal of screening) or control group bias (screening in control group) are not accounted for by increasing the number of women studied, then the ability of the trial to demonstrate a benefit is compromised (*2*).

With the exception of the National Breast Screening Study of Canada (NBSS) (*3,4*), for which only preliminary data are available, the randomized controlled trials were designed to evaluate screening in women aged 40 and older as a single

*See "Notes" section following "References."

group, not as separate trials for women aged 40-49 and women aged 50 and older. Therefore, the size of study and control groups were determined for women of all ages starting at age 40. These randomized controlled trials consistently show an overall benefit for screening beginning at age 40. However, confusion has arisen from the retrospective stratification by age of the women. This unplanned subgroup analysis drastically reduces the ability of any of the trials to provide statistically significant "proof" of benefit. In fact, none of the randomized controlled trials has sufficient power to demonstrate an expected 25%-30% reduction in mortality for women aged 40-49 because the trials were not designed to have the power for small subgroup analysis.

Fletcher et al. briefly discuss this point but then proceed to cloud the issue by citing a recent meta-analysis (*5*), which showed that even the combined data fail to show a statistically significant benefit. This meta-analysis has two important weaknesses. First, it was performed with the use of old data and did not include most of the updated results presented by Fletcher et al. Second, even pooling the data from all eight randomized controlled trials produces insufficient statistical power to indicate presence or absence of benefit from screening. In the eight trials, there were only 167000 women (30% of participants) aged 40-49, a number too small to provide a statistically significant result.

The effects of dilution from noncompliance and from control group bias (contamination) on the calculated power of the trials also is briefly mentioned but then ignored in the article by Fletcher et al. (*1*). To compensate for dilution, the number of women in a trial must be increased. The initially calculated sample size must be multiplied by the factor $1/(1 - n - c)^2$, where n is the expected rate of noncompliance and c is the expected rate of contamination (*2*). If this adjustment is not made, the trial cannot achieve its predicted power to demonstrate the anticipated benefit. A major omission in the article by Fletcher et al. (*1*) is the failure to recalculate the statistical power of the randomized controlled trials on the basis of actual noncompliance and contamination rates rather than on the basis of the expected rates used in the power calculations for the trials. For example, in the NBSS, which did not compensate at all for noncompliance or contamination in its power calculation (*6*), Fletcher et al. note that 26% of control group women aged 40-49 had at least one mammography examination outside of the trial and that 1.5% of control group women had diagnostic mammography examination as they entered the study. The medical indication for these examinations is irrelevant, since diagnostic as well as screening mammography effectively screens both breasts. To compensate for the overall 27.5% rate of contamination, the number of women aged 40-49 in the NBSS should have been increased from 49000 to more than 93000 women.[1] Since the sample size was not increased, the NBSS does not have the power to demonstrate its own calculated 40% or higher benefit, let alone a more realistic 25%-30% benefit.

Despite the fact that all the randomized controlled trials involved too few participants to be expected to prove a benefit for women aged 40-49, Fletcher et al. present the data as scientifically valid and draw firm conclusions. The proper conclusion is that the randomized controlled trials, as designed and executed, both individually and collectively have insufficient statistical power to prove either benefit or lack of benefit for screening women aged 40-49.

We also dispute the conclusion that by 10-12 years after study entry, there is "marginal" reduction in breast cancer mortality among women aged 40-49. Five of the eight randomized controlled trials that involve women aged 40-49 are sufficiently mature to report data with 10-12 years of follow-up. Of these five, four indicate a mortality reduction ranging from 13% to 49%; this reduction certainly indicates a trend toward benefit. The average mortality reduction among all five trials is 17%; to us, this trend is more than "marginal," especially since other randomized controlled trial deficiencies likely have lessened the demonstrated benefit.

Other Flaws in Experimental Design

Most expert mammographers in the United States believe that the screening interval for women aged 40-49 should be 1 year. The lead time for mammographic screening is approximately 2 years for women aged 40-49 (*7,8*). Thus, a screening interval of 2 or more years should have a diminished impact on mortality reduction, since the number of poorer-prognosis cancers detected during the interval would increase while the number of less advanced, screening-detected cancers would decrease. However, in four of the eight randomized controlled trials, mammographic screening was performed every 2 years (or even less frequently) for women aged 40-49. Fletcher et al. mention this issue but then proceed to ignore it. The results in Table 2 of the article by Fletcher et al. (*1*) indicate that, among the five randomized controlled trials for which there is sufficient follow-up (10-12 years) to demonstrate mortality reduction, the three trials utilizing screening intervals of 2 or more years show an average mortality reduction of only 4% (relative risks = 0.75, 0.86, and 1.28). The two trials utilizing screening intervals of less than 2 years, on the other hand, show an average mortality reduction of 26% (relative risks = 0.51 and 0.77).

Mammography should be performed using two views per breast. Single-view mammography has been shown to reduce the ability to detect cancer by up to 11% (*9,10*). However, three of the eight randomized controlled trials that involved women aged 40-49 utilized single-view mammography alone, and another randomized controlled trial used single-view mammography in most women after a baseline two-view screening. Fletcher et al. also briefly mention this issue and then ignore it. Yet the results in Table 2 of the article by Fletcher et al. (*1*) again indicate that among the randomized controlled trials for which there were 10-12 years of follow-up, those utilizing single-view mammography either entirely or partially show an average mortality reduction of only 4% (relative risks = 0.75, 0.86, and 1.28). In contrast, trials utilizing two-view mammography show an average mortality reduction of 26% (relative risks = 0.51 and 0.77).

A final flaw in experimental design involves the NBSS. In this study, a clinical breast examination was performed prior to randomization that was not completely blinded, but then women found to have palpable breast masses were retained in the trial. Specifically, this approach involves two errors: First, encouraging inclusion of prevalent, advanced breast cancers, which cannot benefit from screening but are likely to contribute to early mortality, will dilute any benefit of screening, especially during the early years of follow-up. Second, this approach potentially compromises randomization if disproportionate numbers of patients with advanced cancers are allocated to control and screening groups. Unfortunately, such confounding actually occurred in the NBSS: Of the 24 patients with cancers detected at the initial screening who had four or more positive axillary lymph nodes, 19 (17 of whom had palpable tumors) were allocated to the screening group, whereas only five were allocated to the control group. Although there is no evidence of purposeful misallocation, the disproportionate placement of 14 more patients with advanced cancers in the screened group is sufficient to explain the 10 "excess" breast cancer deaths reported in this group. Certainly, this bias further compromises the ability of the NBSS to demonstrate an early benefit for screening women aged 40-49.

Deficiencies in Implementation of Study Protocols

The ability of mammography to detect nonpalpable breast cancer is very much dependent on image quality. Cancers that are not included on available images due to poor breast positioning cannot be identified, just as cancers may be missed if mammographic images are limited by poor contrast or other technical deficiencies. Mammographic image quality was not documented in most of the randomized controlled trials, but poor image quality has clearly been shown in the NBSS (*11,12*) and acknowledged in the Edinburgh trial as well (*13*). Utilization of readers unskilled in screening interpretation is also likely to compromise the ability of mammography to detect nonpalpable cancer. Again, poor quality interpretation has been documented in the NBSS (*14*) but has not been reported for the other randomized controlled trials.

Fletcher et al. discuss this important issue briefly but attempt to minimize the relevance of poor quality NBSS mammography. They imply that since image quality and interpretation in the NBSS apparently met then-acceptable standards for community practice in Canada, results were especially pertinent to real-world mammography.

This argument is highly misleading for two reasons. First, in assessing the effectiveness of mammography, a randomized controlled trial should evaluate the technique at its best. If found to be efficacious, *then* community standards can be raised to whatever performance level was used in the randomized controlled trial. Evaluation of mammography at a poor level of performance is likely to mask a benefit. Second, standards of community practice for mammography have dramatically improved since the completion of screening in the eight randomized controlled trials studied by Fletcher et al. As a result, mammographic images from the eight randomized controlled trials are not at all comparable with today's real-world mammography. Indeed, none of the randomized controlled trials would be found to meet the current standards of the Mammography Accreditation Program of the American College of Radiology (*15*), which now are being followed by 85%-90% of radiologists in the United States (*16*). There should be no doubt that relatively poor quality mammography in *all* the randomized controlled trials has reduced the ability of mammography to demonstrate the benefits achievable at current levels of performance.

Modern Mammography Versus Randomized Controlled Trial Mammography

The article by Fletcher et al. (*1*) correctly states that major improvements in mammography occurred around 1985 but fails to indicate that all the randomized controlled trials were well under way or had actually been completed prior to this time. The authors go on to suggest that these advances in mammography are relatively unimportant, but this unjustified conclusion was not based on direct comparison of mammographic performance before and after 1985. Actually, when the prognostic factors of breast cancers detected by modern mammography (after 1985) are compared with those detected in the randomized controlled trials, it is clear that a substantial improvement has indeed occurred. As yet unpublished results from the two largest modern breast cancer screening demonstration projects in North America (*17,18*), involving more than 150000 women, show that

1) whereas 22%-30% of randomized controlled trial-detected cancers were T_2 or larger lesions, those found by modern mammography in women aged 40-49 are T_2 or larger in only 12%-14% of cases;

2) whereas 29%-43% of randomized controlled trial-detected cancers were axillary lymph node positive, those found by modern mammography in women aged 40-49 are node positive in only 12%-13% of cases; and

3) whereas 24%-58% of randomized controlled trial-detected tumors were advanced cancers (stage 2 or higher), the corresponding results for modern mammography in women aged 40-49 are only 15%-19%.

Furthermore, the performance of modern mammography is better in women aged 40-49 than in those aged 50-69, unlike the results from the randomized controlled trials considered by Fletcher et al., which demonstrated superior performance for mammography in older women. These observations provide strong inferential evidence that the mammography used in the randomized controlled trials is not comparable with and that the results of the randomized controlled trials are not relevant to the mammography now available for women aged 40-49, which appears to be much more effective.

Conclusions

The article by Fletcher et al. (*1*) involves a one-sided analysis of the facts, ignoring many deficiencies in the

randomized controlled trials that need to be considered for proper interpretation of the data. Randomized controlled trials for women aged 40-49 do not have the statistical power to demonstrate either benefit or lack of benefit for screening. In addition, there is a clear trend indicating benefit for screening women aged 40-49 in randomized controlled trials having 10-12 years of follow-up, with an average mortality reduction of 17%, despite the fact that mammography in the trials was not optimized for these women. Inferential data from nonrandomized controlled trial studies suggests even greater benefit for modern mammography in women aged 40-49.

It is scientifically unjustified to claim that screening women aged 40-49 is ineffective. The only reason not to screen these women is economic. Articles such as the one by Fletcher et al. (*1*) may be used as distractions to avoid the important medical and political debate over whether the cost of screening younger women is justified. We would argue forcefully that many other medical and nonmedical endeavors are less cost-effective than breast cancer screening. However, if a consensus against screening women aged 40-49 is reached, this decision should be made because we are unwilling or unable to pay for it, not because of the misinformed belief that it produces no benefit.

References

(*1*) Fletcher SW, Black W, Harris R, et al: Report of the International Workshop on Screening for Breast Cancer. J Natl Cancer Inst 85:1644–1656, 1993

(*2*) Lachin JM: Introduction to sample size determination and power analysis for clinical trials. Controlled Clin Trials 2:93–113, 1981

(*3*) Miller AB, Baines CJ, To T, et al: Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years [published erratum appears in Can Med Assoc J 148:718, 1993]. Can Med Assoc J 147:1459–1476, 1992

(*4*) Miller AB, Baines CJ, To T, et al: Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years [published erratum appears in Can Med Assoc J 148:718, 1993]. Can Med Assoc J 147:1477–1488, 1992

(*5*) Elwood JM, Cox B, Richardson AK: The effectiveness of breast cancer screening by mammography in younger women. Online J Curr Clin Trials (serial online), Feb 25 (Doc No. 32), 1993

(*6*) Miller AB, Howe GR, Wall C: The National Study of Breast Cancer Screening Protocol for a Canadian randomized controlled trial of screening for breast cancer in women. Clin Invest Med 4:227–258, 1981

(*7*) Moskowitz M: Breast cancer: age-specific growth rates and screening strategies. Radiology 161:37–41, 1986

(*8*) Tabar L, Faberberg G, Day NE, et al: 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 55:547–551, 1987

(*9*) Andersson I, Hildell J, Muhlow A, et al: Number of projections in mammography: influence on detection of breast disease. AJR Am J Roentgenol 130:349–351, 1978

(*10*) Muir BB, Kirkpatrick AE, Roberts MM, et al: Oblique-view mammography: adequacy for screening. Work in progress. Radiology 151:39–41, 1984

(*11*) Baines CJ, Miller AB, Kopans DB, et al: Canadian National Breast Screening Study: assessment of technical quality by external review. AJR Am J Roentgenol 155:743–747, 1990

(*12*) Kopans DB: Response to Re: "Author of Canadian breast cancer study retracts warnings [letter]." J Natl Cancer Inst 84:1367–1368, 1992

(*13*) Roberts MM, Alexander FE, Anderson TJ, et al: Edinburgh trial of screening for breast cancer: mortality at seven years. Lancet 335:241–246, 1990

(*14*) Baines CJ, McFarlane DV, Miller AB: The role of the reference radiologist. Estimates of inter-observer agreement and potential delay in cancer detection in the National Breast Screening Study. Invest Radiol 25:971–976, 1990

(*15*) Hendrick RE: Quality assurance in mammography: accreditation, legislation, and compliance with quality assurance standards. Radiol Clin North Am 30:243–255, 1992

(*16*) Hendrick RE, Smith RA, Wilcox PA: ACR Accreditation and legislative issues in mammography. *In* Syllabus: a Categorical Course in Physics and Technical Aspects of Breast Imaging (Haus AG, Yaffe MJ, eds). Oak Brook, Ill: Radiological Society of North America, 1993. In press

(*17*) Burhenne LJ, Hislop TG, Burhenne HJ: The British Columbia mammography screening program: evaluation of the first 15 months. AJR Am J Roentgenol 158:45–49, 1992

(*18*) Sickles EA: Quality assurance: how to audit your own mammography practice. Radiol Clin North Am 30:265–275, 1992

Notes

[1] $49000 \times [1/(1-.275)^2] = 93222$. This calculation does not account for noncompliance, which would require an even further increase in sample size.

Affiliations of authors: E. A. Sickles, Department of Radiology, University of California School of Medicine, San Francisco.

D. B. Kopans, Department of Radiology, Massachusetts General Hospital, Boston, and Harvard Medical School, Boston, Mass.

Correspondence to: Edward A. Sickles, M.D., Department of Radiology, Box 0628, University of California School of Medicine, San Francisco, CA 94143-0628.

Manuscript received September 10, 1993; accepted September 14, 1993.

Need the latest information on clinical trials for HIV and AIDS?

Call the AIDS Clinical Trials Information Service:

1–800–TRIALS–A

(1–800–874–2572) Monday through Friday, 9 a.m. to 7 p.m. eastern time.

- Free up-to-date information about more than 200 federally and privately sponsored HIV and AIDS clinical trials and the drugs used in the trials.
- Personalized assistance from English- and Spanish-speaking health specialists.
- TTY/TDD access: 1–800–243–7012.

A service of the U.S. Department of Health and Human Services, Public Health Service. The Centers for Disease Control, Food and Drug Administration, National Institute of Allergy and Infectious Diseases, and National Library of Medicine have collaborated to provide this service.

must decide based on their training and experience what is truly known about an agent, which conclusions are too early to make, and, therefore, what data should be made available to a patient. Emerging data are seductive but often subject to various interpretations, as anyone who has attended a National Cancer Institute phase I meeting can readily attest.

The current phase I and phase II trial designs are products of evolution and are widely employed not for their statistical rigor, but because they provide data upon which decisions for future studies can be based. The designs also strike a balance between knowing exactly and unnecessarily treating patients with subtherapeutic doses or inactive drugs. The currently used trial designs are not perfect, and proposals to modify them usually attempt to either improve patient safety or increase the likelihood of response. However, these two elements frequently conflict with one another when a cancer patient is treated with cytotoxic chemotherapy. Many of the proposed modifications to the current phase I and phase II trial designs will shift the balance of these two factors. While a discussion of such proposals may be useful, the feasibility of any new trial design must ultimately be demonstrated by the successful conduct of the clinical study. We can look forward to the investigators at the University of Chicago and other institutions providing us with these reports in the years to come.

References

(1) Ratain MJ, Mick R, Schilsky RL, et al: Statistical and ethical issues in the design and conduct of phase I and II clinical trials of new anticancer agents. J Natl Cancer Inst 85:1637–1643, 1993

(2) Collins JM, Grieshaber CK, Chabner BA: Pharmacologically guided phase I clinical trials based upon preclinical drug development. J Natl Cancer Inst 82:1321–1326, 1990

(3) Penta JS, Rosner GL, Trump DL: Choice of starting dose and escalation for phase I studies of antitumor agents. Cancer Chemother Pharmacol 31:247–250, 1992

(4) Ratain MJ, Berezin F, Allen SL, et al: Population pharmacodynamic (PD) study of amonafide ,(BIDA): CALGB 8862. Proc ASCO 11:A290, 1992

(5) Ratain MJ, Mick R, Berezin F, et al: Phase I study of amonafide dosing based on acetylator phenotype. Cancer Res 53(10 Suppl):2304–2308, 1993

Notes

Correspondence to: Michael J. Hawkins, M.D., Departments of Medicine and Pharmacology, Georgetown University Medical Center, Lombardi Cancer Research Center, 3800 Reservoir Rd., Washington, DC 20007.

Manuscript received September 13, 1993; accepted September 14, 1993.

Screening for Breast Cancer: What Should National Health Policy Be?

*Thomas C. Chalmers**

For decades, it has been assumed that the lives of women with breast cancer could be prolonged if their tumors could be diagnosed early, before metastasis to other parts of the body. The addition of effective technology, in the form of radiological mammography, to supplement routine physical examinations and breast self-examination has been viewed as an inherently attractive program, and mammography has become widely used in the last 20 years.

By 1977, there was general agreement that women older than age 50 should have an annual breast physical examination and mammography. This consensus was codified by the National Cancer Institute (NCI) and the American Cancer Society (ACS) (*1*). However, there were serious doubts about the wisdom of encouraging annual mammograms for women younger than 50. Was the apparent lengthening of life merely a shortening of the prediagnosis lead time? Could the exposures to x rays at a younger age be harmful (2)? In 1980, undaunted by these worries, the ACS encouraged mammography for women under 50.

In 1988, led by the American College of Radiology, most organizations involved with cancer treatment agreed to a more stringent recommendation, which stated that beginning at age 40, annual or biannual mammography should be performed in addition to a yearly clinical examination. In 1991, the ACS reiterated its support of radiological screening in women under 50.

The landmark Health Insurance Plan (HIP) of New York study that began 30 years ago was the first of eight long-term clinical trials in which women were selected at random or randomly assigned either to a group receiving mammography or to a control group with no mammography. Results of that trial and those from the other six studies, which have been published during the last 8 years, have been combined in a meta-analysis by Elwood et al. (*3*) presented at an international workshop on screening for breast cancer summarized in this issue of the Journal (*4*). A representative of each trial was present and participated in the deliberations.

The charge to the workshop was not to make recommendations regarding mammography, but to seek a consensus on the data from the eight studies for the benefit of an advisory committee of the NCI, which would, in turn, make recommendations. The report from the workshop summarizes data on women aged 40-49 in the trials as follows: "The randomized trials of women ages 40 to 49 are consistent in

*See "Notes" section following "References."

showing no statistically significant benefit in mortality after 10 to 12 years of follow-up . . . . In contrast, results for women ages 50 to 69 in all randomized trials confirmed the disease postponement benefit that was first seen in the HIP trial." All representatives agreed that more randomized control trial data are needed for women who might have their first mammography after the age of 69.

Thus, mammography recommendations with regard to women under the age of 50 seem to have come full circle. However, it is important to point out that there is as yet no basis for denying potential efficacy in younger women. Except for the hard-to-explain shortening of time between random assignment and death from breast cancer found in the Canadian trial, harm has neither been suggested nor proven by the randomized control trial data presented at the workshop. Although mammography's combined effect on mortality from breast cancer in women 40-49 years of age is very close to zero, the 95% confidence interval includes a potentially beneficial lengthening as well as harmful shortening of the time until death from breast cancer of approximately 30%.

This analysis by workshop participants is a good example of how hard it is to prove a negative hypothesis in clinical medicine. The statistical power needed to accomplish that task depends both on the total numbers of study subjects and on the quality of the studies. Although almost half a million young women have been randomly assigned in the eight studies, the number that have had follow-up for 10 years or more is still relatively small, and several studies have found a delay of 8-12 years in the possible effect on mortality.

The quality and applicability of the studies have been vigorously challenged by radiologists, but they are not an unbiased group (*5*). Even so, I believe that there are numerous ways in which flaws in the studies could have resulted in a spurious answer. Randomization was sometimes by clusters of patients rather than by individuals. The technique of mammography was not always what is now considered optimal. The effects of crossovers were not adequately measured (they rarely are). The best of the studies in women aged 40-49 (*6*) is one of two recently reported trials comprising the Canadian National Breast Screening Study (NBSS). This trial was different from most of the others in that the effect of annual mammography and physical examination was compared with that of a single physical examination plus yearly follow-up with a self-administered questionnaire. Over 50000 women were randomly assigned. The risk of death from breast cancer was higher in the mammography group (odds ratio = 1.36; 95% confidence interval = 0.84-2.21)—approximately 35% higher. Even if some beneficial effect begins to appear in the next 5-10 years for women aged 40-49 at the start of the study, it is extremely unlikely that the observed trend will be completely reversed. For the second NBSS trial, involving women aged 50-59 (*7*), annual mammography combined with breast physical examination had produced no effect after 7 years compared with physical examination alone. Considering the NBSS trials, the apparent trend toward a good effect of mammography in the past studies in women of all ages could be attributed to awareness stemming from an annual examination rather than a specific effect of mammography.

So the pooled results of the definitive trials on mammography in women aged 40-49 are unimpressive for mortality efficacy, as pointed out in the accompanying article by Fletcher et al. (*4*). It is important to emphasize that we are discussing routine preventive measures for breast cancer in the general population, not for patients at high risk, such as those with a strong family history.

What should those individuals responsible for public health policy now recommend? I believe that after wallowing in the data for almost a year, I can answer that question with respectable credibility.

I propose two measures: First, because of the annoying uncertainty regarding the benefit of mammography for women aged 40-49, another study that corrects a major defect in all the studies reported to date should be conducted immediately. This defect is the relative crudeness and insensitivity of determining effects by decade rather than individual years of age at the time of randomization. We can remedy this problem by obtaining from all of the investigators all of the individual patient data on age and menopausal status at the time of randomization and timing of all events and compliance after randomization. Menopausal status is especially important because many bits of data suggest that cancer of the breast acts differently before and after menopause. Perhaps menopausal status is the critical cutoff point rather than the age of 50.

The NCI should see that such a "mega-analysis" is undertaken. A precedent has been set by the National Heart, Lung, and Blood Institute. Individual data on all patients randomly assigned to coronary artery bypass surgery in seven trials started over 10 years ago are about to be published.

If such a study does not yield a satisfactory conclusion in 6-12 months and if the data from other trials now under way do not suggest a clear answer to the question, a large-scale international trial designed to avoid all past mistakes should be undertaken. This trial would, I hope, be sized to supplement the available information rather than test the null hypothesis as if no prior information existed.

My second proposal concerns third party reimbursement for mammography in women under the age of 50—or in premenopausal women, if menopause proves to be the critical cutoff point. I recommend that third parties pay for mammography in these women only if the women are randomly assigned in a peer-reviewed study. The point here is that it is foolish for some third parties to continue to pay for mammography in younger women as if it has been proven to be effective and for other third parties not to pay for mammography as if it has been proven ineffective. A board of experts could authorize exceptions such as geographic inaccessibility or other contraindications.

A policy of supporting needed clinical trials and not paying for unproven diagnostic and therapeutic interventions should be a cornerstone of any health care reform, because such a policy will improve quality while at the same time saving money. Mammography for younger women would be a great place to start.

References

(1) National Cancer Institute: Consensus Development Meeting on Breast Cancer Screening. US Department of Health, Education and Welfare, Public Health Service, National Cancer Institute. DHEW Publ No. (NIH)78-1257, September, 1977

(2) Bailar JC 3d, Smith EM: Progress against cancer? N Engl J Med 314:1226–1232, 1986

(3) Elwood JM, Cox B, Richardson AK: The effectiveness of breast cancer screening by mammography in younger women. Online J Curr Clin Trials (serial online), Feb 25 (Doc No 32), 1993

(4) Fletcher SW, Black W, Harris R, et al: Report of the International Workshop on Screening for Breast Cancer. J Natl Cancer Inst 85:1644–1656, 1993

(5) Chalmers TC: Editorial: mammography after 30 years. Online J Curr Clin Trials May 28, (Doc No 66), 1993

(6) Miller AB, Baines CJ, To T, et al: Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years [published erratum appears in Can Med Assoc J 148:718, 1993]. Can Med Assoc J 147:1459–1476, 1992

(7) Miller AB, Baines CJ, To T, et al: Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years [published erratum appears in Can Med Assoc J 148:718, 1993]. Can Med Assoc J 147:1477–1488, 1992

Notes

Correspondence to: Thomas C. Chalmers, M.D., Medical Research Consulting, 32 Pinewood Village, West Lebanon, NH 03784.

Manuscript received September 17, 1993; accepted September 17, 1993.

Deficiencies in the Analysis of Breast Cancer Screening Data

*Edward A. Sickles, Daniel B. Kopans**

Breast cancer is a major, and increasing, public health problem in the United States and most other industrialized countries. Since there is not yet an established means of preventing breast cancer, efforts to combat the disease have focused on early detection through screening to increase the success of existing treatment. There is general consensus that screening is both safe and effective in women aged 50 and older, but considerable controversy remains over the utility of screening women aged 40-49.

The report by Fletcher et al. published in this issue of the Journal (*1*) is a summary of a breast cancer screening workshop convened by the Early Detection Branch of the National Cancer Institute in February 1993. The vast majority of data presented at the workshop involved only the eight randomized controlled trials of breast cancer screening, despite the availability of an abundance of additional evidence. The article by Fletcher et al. (*1*) reflects this bias by discounting the importance of nonrandomized controlled trial data. Screening data from the eight randomized controlled trials are discussed for women aged 40 and older, with special focus on women in the specific age group 40-49, the period during which screening is most controversial.

Fletcher et al. briefly cite many of the problems that have been raised concerning the validity and applicability of the screening data presented at the workshop. However, the reader is left with the implication that these problems are relatively unimportant, when in fact they severely limit both statistical validity and clinical applicability. In this editorial, we present a much more critical analysis of available screening data and also provide what we believe to be a more balanced overview of current knowledge on the strengths and limitations of screening. Paralleling the article by Fletcher et al. (*1*), we concentrate on the evidence from randomized controlled trials of breast cancer screening in women aged 40-49. The major problems concerning the randomized controlled trial data for this age group include lack of statistical power, other flaws in experimental design, and technical deficiencies in implementation of study protocols. Some specific problems are common to all randomized controlled trials, others to individual randomized controlled trials.

Statistical Power

A randomized controlled trial of breast cancer screening must involve sufficient numbers of women who develop sufficient numbers of breast cancers resulting in sufficient numbers of breast cancer deaths, such that a reduction in the number of deaths by screening can be demonstrated with statistical validity. In designing a trial, planners make a calculation of statistical power that estimates how large the study and control populations should be to demonstrate the anticipated benefit. The smaller the expected benefit, the larger the number of women needed in the trial; conversely, if the number of women in the trial is small, only a very large benefit can be demonstrated. If the number of women is too small to begin with, or if dilution effects from noncompliance (refusal of screening) or control group bias (screening in control group) are not accounted for by increasing the number of women studied, then the ability of the trial to demonstrate a benefit is compromised (*2*).

With the exception of the National Breast Screening Study of Canada (NBSS) (*3,4*), for which only preliminary data are available, the randomized controlled trials were designed to evaluate screening in women aged 40 and older as a single

*See "Notes" section following "References."

Briefing Paper

Research Initiatives
Breast Cancer Screening - Breast Imaging Technologies

National Cancer Institute

August 31, 1993

Issues in Breast Cancer Diagnosis: Optimal Usage and Novel Technologies

The favorable effects of early detection are clearly demonstrated for breast cancer. Through intensive efforts directed toward patient and physician education and the provision of screening and other health care modalities to all populations, there has been a significant increase in the numbers of women who have obtained mammography. In 1987, only 37 percent of women over the age of 40 had ever had a mammogram, and only 17 percent had a mammogram within the preceding year. Analysis of the NCI-based National Health Interview Survey conducted in 1990 reveals that 60 percent of women over 40 have now had at least one mammogram. Approximately 40 percent of those women obtained a mammogram within the preceding year and a total of 52 percent have had a mammogram within two years. A cancer control supplemental study is examining continuing trends and will define the proportion of women who continue to obtain screening mammography at regular intervals. This increase in mammography has, in turn, led to an increase in the detection of early stage breast cancer (when it is most amenable to treatment). From 1983 to 1990, for women over 50, in situ cancers (that is, cancers that are not yet invasive) have increased by more than 150 percent for both African-American and white women, while late-stage disease (Stages III and IV) have decreased by around 20 percent for both races. Viewed another way, in situ cancers have increased from 3 percent to 12 percent of all breast cancers in all women, and the percentage of small tumors (<2 centimeters) has increased from 74 percent in 1983 to 92 percent in 1988. Concomitantly, the initial diagnosis of Stages III/IV has reached a nadir of nine percent for white women, although it is almost double (16.4 percent) for African-American women. Nonetheless, the shift in detection to early stage tumors, especially in situ cancers, is responsible to some degree for the apparent increasing incidence of breast cancer in all age groups.

The value of routine mammography for different subsets of women is an important issue in the optimal delivery of health care. To this end, NCI hosted an International Workshop on Screening of Breast Cancer in February 1993 to address key issues surrounding the routine application of mammography to women of age 40 or more: quality assurance (including false positive and negative results); duration of follow-up; endpoint death review; and patterns of mortality, especially in women under age 50. In reviewing studies in Canada, Sweden and the United States, the conclusions emanating from the workshop validate the routine use of mammography in women ages 50-69, where breast cancer-related mortality is reduced by 30 percent. Such benefit can be appreciated within the first 5-7 years of follow-up. In women ages 40-49, however, no such benefit for overall survival is detected in the first 5-7 years, although there may be some benefit as the observation period continues to 10-12 years. In the Canadian study, women whose tumors were detected by mammography alone (i.e. through routine screening) enjoyed the longest survival. Further, the number of women with an initial positive mammogram was two-fold higher in the 50-69 year age group than the 40-49 year age group. The development of more sensitive technology to discriminate benign from malignant lesions in the denser breast tissue of younger women will be the subject of future international studies.

To date, mammography combined with clinical breast examinations is the only screening technology clearly tied to reductions in breast cancer mortality. Technologic advances as well as quality assurance of mammography are essential if this early detection tool is to reach its full capability. Because high-quality x-ray mammography remains the most effective technology currently known to detect breast cancer, NCI will continue to foster research to further refine this technology with regard to lower dose x-rays, enhanced image quality, image digitization for remote centralized reading by reference mammographers and artificial intelligence, as well as statistical techniques for

computer-assisted image interpretation.

Still, although screening for breast cancer in women over 50 years of age has been shown to reduce mortality from the disease by as much as 30 percent, screening remains less than totally effective, in part because early cancers may not be visualized on mammography. In September 1991, NCI held a workshop entitled "Breast Imaging: State-of-the-Art and Technologies for the Future." Experts from industry, academia and several Federal agencies made recommendations for increasing the effectiveness of breast imaging by: 1) improving the quality of conventional film mammography through institution of, and rigid adherence to, quality assurance standards; and 2) supporting research initiatives aimed at the development of novel imaging technologies, including development of non-ionizing radiation imaging technologies and development of digital mammography. In response to the second of these recommendations, NCI has initiated three international research programs: 1) Novel Non-Ionizing Radiation Technologies for Breast Cancer Imaging; 2) the National Digital Mammography Development Group; and 3) the Radiation Diagnostic Oncology Group V: Stereotactic Tissue Diagnosis for Minimal Breast Lesions. The rapid preclinical development, clinical testing and broad application of these novel technologies is a crucial area of commitment for NCI.

Novel Non-Ionizing Radiation Technologies

Current data indicate that conventional mammography is a mature imaging technology producing high- quality images in the majority of patients. Except by improvements in the uniformity of mammographic quality and interpreter competence throughout the country, a problem that is being addressed by the development of quality assurance standards, most experts believe that fundamental limitations of conventional mammography are unlikely to be overcome. However, several non-ionizing technologies do have the potential for improving the diagnostic sensitivity of breast cancer screening without the potential risk of exposure of healthy breast tissue to radiation. These technologies include magnetic resonance imaging (MRI), ultrasound, electron paramagnetic resonance (EPR), and thermal imaging. In addition, a very low radiation exposure technique, positron emission tomography (PET), has the potential for detecting early breast cancers, although it is more likely to have a role in staging already-diagnosed breast cancers. Through a Program Announcement issued in 1992, NCI solicited grant proposals for the development of non-ionizing technologies and imaging methodologies aimed at improving the diagnosis and characterization of breast cancer. Twenty two applications have been submitted, six have been funded and six additional applications are expected to be funded within the next few years.

MRI may provide an additional tool for early detection of cancers through images produced using biochemical information in the target tissue. Research using MRI is currently in progress at the Lombardi Medical Center, Georgetown University. Dynamic contrast enhanced MRI has been shown to be a promising adjunctive diagnostic tool, when conventional mammography plus physical examination are equivocal, for the differentiation of treatment-induced changes from viable tumor in some women treated with local radiation without mastectomy and for detection of tumors in dense breasts. One potential indication for MRI as a screening tool is for women with breast implants, which are virtually radioopaque to mammography. Novel MRI and related imaging techniques include Magnetization Transfer; diffusion/perfusion imaging; Magnetic Resonance Spectroscopy (a newly developing technology that offers the potential for biochemical analysis of tumor and normal cell metabolic rate, pH, oxygen content and other measures of tumor physiology); and Electron Paramagnetic Resonance (EPR, an imaging technology currently being developed by NCI scientists)

which has the capability of detecting free radicals and will use a circulating stable free-radical compound as a "contrast" imaging material.

Currently, both MRI and ultrasound imaging provide metabolic characterizations of previously identified palpable breast lumps or masses or mammographic abnormalities. Ultrasound may provide important additional diagnostic information when the detected lesions need to be characterized as cystic vs. solid and may be particularly useful in younger women with radiodense breast tissue. Novel ultrasound technologies, such as high-frequency systems, modern pulse echo/color flow and 2- and 3-dimensional imaging, may improve image quality and provide improved anatomic and physiologic information. These techniques may provide structural, biochemical, biophysical and physiologic information about breast cancers and, at the least, will certainly be highly informative research tools. Further clinical studies will be required to define the comparative diagnostic value and cost-effectiveness of MRI and ultrasound. Additional novel imaging techniques including optical, microwave and thermal imaging are currently under development as well. Recent reports indicate that ballistic optical imaging can detect structures as small as 0.2 millimeters in vitro. Further studies are required to develop this technique for in vivo use.

Digital Mammography and Teleradiology

Digital mammography may be the most promising of the new technologies in breast imaging. This technology has the potential to improve accuracy, facilitate image interpretation, simplify image storage as well as retrieval and facilitate image transmission to expert mammography interpreters at a distance from the mammography facility. In conventional mammography, "analog" images are obtained on x-ray film. Digital mammography offers many potential advantages over conventional film mammography: 1) improved image quality and resolution; 2) efficient archiving of images; 3) quick retrieval of previous images from the same patient for comparison with current images to detect suspicious changes; 4) facilitation of computer-assisted image interpretation; 5) electronic transmission of images within and between institutions, allowing improved continuity of care; and 6) expert consultation and remote interpretation of images, which will improve the quality of care generally, but especially in areas underserved by radiologists expert in mammography interpretation. Although conventional film images can be digitized to avail the benefits of data storage and transmission, the full benefit of digital mammography will be realized only with the development of high-resolution digital detection systems for acquiring images with substantially greater information content in terms of ranges of gray level than is possible with x-ray film. This will allow computerized image enhancement and computer-assisted interpretation. Along these lines, a completely digital mammographic system for integration into the Madigan Army Medical Center System is currently being tested. Preliminary findings in ongoing studies of image analysis technology demonstrate promising results in terms of tumor feature identification and enhancement.

To a considerable extent, the development of digital mammography, digital data archiving and teleradiology will be enhanced by collaborations with scientists within the Department of Defense (DoD) and NASA involved with the technology of image enhancement and analysis, artificial intelligence and the development of high-resolution digital detection systems. NCI proposes grant support and encouragement of collaborations and technology transfer among DoD and NASA scientists, NCI-supported scientists, largely in academic centers and small businesses, and industry for development of: 1) digital mammography; 2) software for computerized image enhancement,

interpretation and archiving; and 3) teleradiology. To stimulate research in this area, NCI solicited grant proposals through a Program Announcement to create a National Digital Mammography Development Group (NDMDG). The NDMDG will be supported by the Interactive RO1 mechanism and will consist of six major components: 1) digital mammography; 2) image processing; 3) computer-aided diagnosis; 4) teleradiology; 5) pre- clinical and clinical technology evaluation; and 6) headquarters for the centralized development of experimental design and data processing. The NCI/NASA Working Group for Technology Transfer, an outgrowth of the NIH/NASA interagency agreement initiated in 1992, held a workshop in May 1993 to identify and evaluate prospective technological advances. Of 43 applications that have been submitted in response to the Program Announcement, 13 technologies appear to be novel and promising, with potential breakthroughs in the areas of digital detectors and high-performance, low-cost all-fiberoptic networks for teleradiology. Funding of several investigators is anticipated to begin over the next two years.

Stereotactic Tissue Diagnosis for Minimal Breast Lesions

Image-guided stereotactic needle biopsy of breast lesions is a new technology that has potential advantages over surgical excisional or incisional biopsy. The minimally invasive procedure is rapid and accurate, relatively painless and without cosmetic morbidity and less costly than an open surgical biopsy. NCI is supporting studies using stereotactically-guided needle aspiration biopsy, which allows the three-dimensional localization of an abnormality, so that a diagnostic tissue sample can be removed without more invasive surgery. The technique could significantly reduce the trauma of breast biopsy, increase accuracy and thus permit more frequent biopsy in questionable cases. Indeed, since about 80 percent of breast biopsies are negative for cancer, the use of this procedure would obviate a surgical procedure in the majority of these women. Comparisons of needle biopsies with surgical excision biopsies using hookwire localization of the lesion are underway. Early efforts have concentrated on technical issues. One study, designed to accrue 300 patients, is expected to provide valuable information to improve breast biopsy procedures. Further, in 1993, NCI issued a Request for Applications (RFA) for Cooperative Agreements to form the Radiologic Diagnostic Oncology Group V: Stereotactic Tissue Diagnosis for Minimal Breast Lesions. This group will conduct multi-institutional clinical trials to evaluate prospectively the diagnostic accuracy and cost-effectiveness of stereotactic needle biopsy in comparison with open surgical diagnosis. In response to the RFA, three consortia, each consisting of more than 10 institutions, applied. One consortium and a headquarters institution are expected to be funded by December 1993.

Positron Emission Tomography Diagnostic Oncology Group (PETDOG)

While currently used clinical diagnostic imaging modalities, such as computed tomography (CT) and magnetic resonance imaging (MRI), provide predominantly anatomic information, positron emission tomography (PET) reflects tissue physiology and metabolism. Frequently, CT or MRI fail to differentiate definitively among tumors, infection, edema and necrosis – in effect, the likeliest possibilities when an abnormality is seen in an imaging study in a patient with cancer. A variety of clinical reports over the last several years have suggested that PET can distinguish these possibilities. Preliminary data from one very small study of PET as an adjunct to initial staging in women was recently reported; although the false- negative diagnostic rate was 25 percent (4/16 positive axillae missed), the false-positive rate was zero (12/12 true positives). In 1992, NCI convened a workshop to discuss the current status and potential of PET as a clinical tool in

oncology. The panel concluded that sufficient preliminary data existed to justify prospective clinical trials assessing the ability of PET with 2-[^{18}F]-fluoro-2-deoxy-D-glucose (FDG) as a tracer to stage newly-diagnosed primary breast cancer.

NCI plans to initiate a new Diagnostic Oncology Group through Cooperative Agreements with three or four institutions to evaluate PET prospectively. The goals of this group will be to evaluate PET's ability to detect and characterize tumor extent and metabolic activity, regional lymph node and systemic metastases; to assess tumor response to treatment; and to correlate PET data with other imaging modalities and clinical outcome.

Control-Related Research in Breast Cancer Screening

Other data sources are being established to augment national information systems. Included among these is the international database on breast cancer screening. Funding for collaborative analyses of data available from several of the participating countries is planned for 1995. The data on the organization and performance of breast cancer screening in countries further along in this process than the United States will provide information for policy decisions and initiatives made by the government and private sector.

The risk of breast cancer increases with advancing age and women aged 65 years and over have experienced the major impact of morbidity and mortality from this disease. In 1992, the NCI-funded Panel on Breast Cancer Screening in Older Women concluded that efforts to reach this population are inadequate and developed recommendations for primary care physicians. To this end, NCI and the National Institute on Aging (NIA) are planning to co-sponsor a research initiative to support the development of innovative interventions to increase and improve the early detection of breast cancer in asymptomatic women aged 65 and over. The development of interventions will be predicated on documentation of barriers to mammography referral among these women and their primary care physicians. Cost-effective interventions to increase the adoption and regular compliance with procedures for early detection of breast cancer are also planned.

The results of six major community intervention trials to increase breast cancer screening, particularly mammography, are being used to produce a state-of-the-art manual on "How to Increase Breast Cancer Screening in Your Community." This manual will be distributed to health departments, other community organizations, the American Cancer Society (ACS) and cancer centers across the country. The effectiveness of several other dissemination strategies will be tested through this research as well. The strategies that result in widespread adoption of trial results will be identified for future applications in the general population.

NCI in collaboration with the Food and Drug Administration (FDA) and the Centers for Disease Control and Prevention (CDCP) has developed the National Strategic Plan for the Early Detection and Control of Breast and Cervical Cancers. While the plan recognizes the continuum of activities, its primary focus is on early detection and follow-up of breast and cervical cancers. A key component of the Plan includes quality assurance for breast cancer control that includes recommendation for actions appropriate to a broad array of procedures including: clinical breast examination, screening mammography, fine-needle aspiration and core biopsy.

NCI was central to the planning and execution of a conference on "Long-Term Effects of Exposure to DES," held in 1992. Because of a potential increased risk of breast cancer for these women as

reported at the conference, future educational needs were identified. Currently, there are no breast cancer screening efforts in the United States to target women who were administered DES between 1940 to 1971, despite their 40 percent elevated risk for developing this cancer. NCI issued an RFA in 1993 to design and test a regional breast cancer screening program for DES-exposed women. Members of existing cohorts of DES women as well as members of various DES action groups and/or callers to the Cancer Information Service (CIS) Hotline will be used to identify those eligible for screening.

Finally, the application of molecular markers to screening for predisposition to certain types of cancers is a crucial investigative direction that intersects with all aspects of cancer prevention and control. Briefly, with respect to breast cancer, NCI is planning a study in a cohort of women at high risk for breast cancer on the basis of genetic and epidemiologic characteristics to investigate the association between mammographic findings and biomarker expression. Serial collection of breast nipple aspirates and breast needle aspirates would be correlated with the biopsy specimen when obtained. Biomarkers such as insulin-like growth factor (IGF)-1 and transforming growth factor (TGF)-beta levels, estrogen and progesterone receptors, retinoic acid receptors, HER-2/neu (erbB-2) expression would be evaluated as tools for early detection of breast cancer.

SPECIAL ARTICLE

VARIABILITY IN RADIOLOGISTS' INTERPRETATIONS OF MAMMOGRAMS

JOANN G. ELMORE, M.D., M.P.H., CAROLYN K. WELLS, M.P.H., CAROL H. LEE, M.D., DEBRA H. HOWARD, M.D., AND ALVAN R. FEINSTEIN, M.D.

Abstract *Background.* Despite the proved value of mammography in screening for breast cancer, its efficacy depends on radiologists' interpretations. The variability in such interpretations is not well understood.

Methods. Using a technique of stratified random sampling, we selected 150 mammograms obtained in 1987: 27 from women with histopathologically confirmed breast cancer and 123 from women with no evidence of breast cancer after three years of follow-up examinations. Ten radiologists, who were unaware of the diagnoses and research hypothesis, each interpreted the 150 mammograms. Disagreement was analyzed within pairs of the 10 radiologists, as well as for the group of 150 women as a whole.

Results. The diagnostic consistency between pairs of radiologists was moderate, with a median weighted percentage of agreement of 78 percent (weighted kappa, 0.47). The frequency of the radiologists' recommendations for an immediate workup ranged from 74 to 96 percent for mammograms from the women with cancer and from 11 to 65 percent for films from the women without cancer. A substantial disagreement in management recommendations — in which one radiologist recommended routine follow-up and another recommended a biopsy for the same patient — occurred in 3 percent of the pairwise comparisons but in 25 percent of the comparisons for the group of women as a whole. When two or more radiologists recommended a biopsy for the same patient, a disagreement in the stated location (right or left breast) occurred in 2 percent of the pairwise comparisons among the radiologists but in 9 percent of comparisons for the group of women as a whole. Because some disagreement was likely, given that 10 radiologists read each film, the pairwise comparison is a more conservative estimate of disagreement.

Conclusions. Although mammography is of value in screening women for breast cancer, radiologists can differ, sometimes substantially, in their interpretations of mammograms and in their recommendations for management. Efforts to improve accuracy and reduce variability in interpretation may increase the effectiveness of mammography in detecting early breast cancers. (N Engl J Med 1994;331:1493-9.)

DESPITE extensive efforts to improve the technical aspects of mammography, the extent of radiologists' variability in interpretation has seldom been examined. In a study of xeromammograms, the level of agreement on the diagnosis of benign abnormalities was relatively low,[1] but the level of agreement was higher for classifying mammographic parenchymal patterns.[2] Subsequent studies of film-screen mammograms have shown similar levels of variability,[3,4] which may delay the detection of breast cancer.[5] We conducted a study to determine the extent and possible sources of variability in diagnostic accuracy and subsequent recommendations for mammograms.

METHODS

Case Selection

From mammograms obtained in 1987 at Yale–New Haven Hospital, 150 were randomly selected from three strata of diagnostic findings ("normal," "abnormal, probably benign," and "abnormal, suggestive of cancer") and from the definitive designation of breast cancer or the absence of cancer. The definitive diagnosis was breast cancer if it was histopathologically confirmed within three years after the 1987 mammogram. The three-year period was allowed because there may have been some false negative readings of the 1987 films.[6] A definitive designation of the absence of cancer required both no evidence of breast cancer during the three-year period after 1987 and a designation of normal findings or abnormal, probably benign findings on mammograms obtained in 1990. To avoid bias associated with the workup,[7] a biopsy was not required to verify the absence of cancer.

All the patients were women, and those with benign findings on previous breast biopsies were eligible for inclusion in the study. Patients were ineligible if the 1987 mammogram had not been definitively interpreted. After other exclusions, discussed below, the mammograms for each eligible patient were reviewed for technical quality without knowledge of the patient's identity, the 1987 diagnostic interpretation, or the patient's outcome.

Mammography was performed with a standard film-screen technique on a Senographe 500T Unit (Thompson CGR Medical, Columbia, Md.), with the use of Kodak Ortho M film and MIN-R screens. Mediolateral oblique and craniocaudal views of original films, rather than copies, were available for each breast. The films were considered to be of good technical quality by 1987 standards. To maintain confidentiality, all mammograms were coded by number. The protocol was approved by the Yale–New Haven Medical Center Human Investigation Committee.

Participating Radiologists

Board-certified diagnostic radiologists from community and academic practices were invited to participate in the study if they routinely read mammograms. The radiologists were unaware of the objectives and design of the study and of the number of cases in which cancer was subsequently diagnosed. Each radiologist received a small honorarium.

Data Acquisition

In phase 1 of the research, each of the 10 radiologists independently read the mammograms for the 150 patients. In phase 2, which occurred five months later, the same films were reviewed again, this time in a new, randomly arranged sequence. The radiologists were told that all the patients were female, with no history of previous breast cancer, but they were not told that they would be seeing the same films a second time.

In both phase 1 and phase 2, films for 50 patients, shown with the patient's age but no clinical information, were used to assess in-

From the Departments of Internal Medicine (J.G.E., C.K.W., D.H.H., A.R.F.), Diagnostic Radiology (C.H.L.), and Epidemiology and Public Health (A.R.F.), Yale University, New Haven, Conn. Address reprint requests to Dr. Elmore at the Primary Care Center, Yale University School of Medicine, 20 York St., New Haven, CT 06504.

Supported by a grant (RD-346) from the American Cancer Society. Dr. Elmore was a Robert Wood Johnson Clinical Scholar.

Presented in part at the meeting of the American Association of Physicians, Washington, D.C., April 30–May 3, 1993.

traobserver variability. Of the films for the remaining 100 patients, half were shown with a detailed clinical history in phase 1 but with only the patient's age in phase 2; this sequence was reversed for the other half. The detailed clinical history included signs and symptoms (e.g., a palpable lump or nipple discharge), the location of a previous breast biopsy if there had been one, menopausal status, estrogen use, history of other cancers, and family history of breast cancer. We performed a separate analysis of the effect of clinical histories on mammographic interpretation (data not shown).

For each mammogram, the radiologists used a checklist to indicate observations, diagnostic interpretations, and recommendations for management. The checklist was reviewed with the participating radiologists before phase 1, and those who were unfamiliar with the form received brief written instructions. The radiologists were asked to interpret the films as they would in their own clinical practice, with no time limits for the readings.

In descriptive observations, the radiologists could note specific abnormalities (e.g., a mass or calcification) and their location (right or left side, with the position demarcated according to the hours on a standard 12-hour clock). If more than one abnormality was noted in the same patient, the radiologists were asked to indicate which side contained the most serious abnormality (i.e., the most suggestive of cancer). Diagnostic interpretations could be chosen from one of four categories: normal; abnormal, probably benign; abnormal, indeterminate (for an uncertain finding); or abnormal, suggestive of cancer. The radiologists were asked to avoid the category "abnormal, indeterminate" whenever possible.

Management recommendations could be chosen from the following categories: routine mammographic follow-up (i.e., screening with mammography), mammographic follow-up after a short interval (e.g., within six months), additional mammographic views, ultrasound examination, or biopsy.

Data Analysis

The original data were double-entered and verified for electronic coding. Statistical analyses were performed either with electronic hand calculators or with programs in the Statistical Analysis System.[8]

Interobserver variability in interpretations of the mammograms for the 150 patients was appraised from the readings of pairs of radiologists in phases 1 and 2. Intraobserver variability was assessed for the 50 mammograms that were shown in both phases with the age of the patient as the only additional information provided. The indexes of variability used in the paired comparisons were the percentage of agreement and the kappa statistic.[9,10] For ordinal measures, both indexes were weighted with integers representing the number of ordinal categories from perfect agreement.[11,12]

The frequency of agreement was evaluated for each of the four categories of diagnostic interpretation. The accuracy (sensitivity and specificity) of the diagnosis of cancer was determined for films interpreted as "abnormal, suggestive of cancer," as compared with the accuracy for a combination of the other three diagnostic categories. The percentage of patients with recommendations for a workup was evaluated separately for patients with cancer and those without cancer.

For any pair of radiologists reading the same patient's mammogram, a substantial clinical disagreement was considered to exist if the films were interpreted as normal by one radiologist and as abnormal and suggestive of cancer by the other, or if routine mammographic surveillance was recommended by one radiologist and a biopsy was recommended by the other. If both members of a pair recommended an immediate workup (or a biopsy), a substantial disagreement was considered to exist if the most abnormal lesion was located in the right breast by one radiologist and in the left breast by the other. A description of bilateral abnormalities by one radiologist and of a unilateral abnormality by the other radiologist was not counted as a disagreement.

The proportions of substantial clinical disagreements were calculated in two ways: first, within the maximal number of pairwise comparisons, and second, for the overall group of 150 patients (per-patient comparisons). For the pairwise calculation, the numerator was the number of disagreements between the members of a pair of radiologists. The denominator was the maximal number of possible pairwise comparisons for the 10 radiologists: 45 pairs for an individual patient ($10 \times 9 \div 2$), and 6750 pairs (45×150) for all 150 patients. For the per-patient calculations, the numerator was the number of patients about whom at least two radiologists had a substantial disagreement, and the denominator was the total number of patients (150). Since immediate workup and biopsies were not recommended for all patients, disagreement about the location of abnormalities was calculated with denominators consisting of pairwise (or per-patient) comparisons in which at least two radiologists had recommended such a workup.

A simulated screening population was composed of the subgroup of asymptomatic women whose mammograms were accompanied by a clinical history. Regardless of a possible family history of breast cancer or a previous breast biopsy with benign results, these women reported no symptoms related to the breast at the time the mammography was performed. The results in this subgroup were assessed separately for variability, substantial clinical disagreements, and the percentage of patients without cancer for whom an immediate workup was recommended.

Assessment of Sources of Variability

After the phase 2 analysis, the participating radiologists were assembled to review specific mammograms chosen as examples of disagreement and to comment on the variability.

Results

Selection of Cases

The patients were selected from those with mammograms in 1987 from three strata of diagnostic findings: normal (>3000 women), abnormal but probably benign (567), or abnormal, suggestive of cancer (124). Random sampling within these categories resulted in 275 eligible women with normal findings and 207 with abnormal but probably benign findings plus all 124 considered to have abnormalities suggestive of cancer. As is customary in studies of variability among observers, the groups were selected to provide an enriched spectrum rather than the predominance of normal cases in actual screening, which would not be challenging and could also lead to artificially low results for kappa indexes of concordance.[13]

Of the 606 eligible patients, 271 had adequate follow-up (pathological studies or additional mammograms) for the determination of a definitive outcome: 89 were in the original group with normal findings, 110 had abnormal but probably benign findings, and 72 had abnormalities considered suggestive of cancer. From this group, 121 patients were then excluded for the following reasons: a history of breast cancer (24); large breasts, requiring special films or more than one film per breast (24); previous cosmetic breast surgery (8); radiopaque skin marker on the breast (6); films of inadequate technical quality (11); or films not available or only one view available (48). In the final group of 150 patients, the diagnostic interpretations from 1987 (and the subsequent outcome status) were as follows: 54 with normal findings (1 with cancer), 61 with abnormal but probably benign findings (7 with cancer), and 35 with abnormal findings suggestive of cancer (19 with cancer).

Of the 150 patients, 95 (63 percent) were at least 50 years old (range, 33 to 83), and 40 (27 percent) had previously undergone a breast biopsy, with negative results. Breast cancer had been diagnosed in 22 patients within one year after the 1987 mammography, and in 5 more patients during the two ensuing years. For 85 of the 127 patients who had reported no symptoms, the mammograms were read with the clinical history available, and this group of 85 women became the simulated screening subgroup; 9 were subsequently found to have cancer.

Participating Radiologists

Of the 10 participating board-certified radiologists, 7 were in private practice and 3 held full-time academic positions in New York or Connecticut. All 10 had received training in mammography during residency or in continuing medical education courses (or both), and 9 would have qualified to read mammograms in a facility accredited by the American College of Radiology.[14] The group had a median of 7 years of clinical experience reading mammograms (range, 1.5 to 20) and had interpreted a median of 1900 mammograms in the year before the study (range, 200 to 6000). The median percentage of time spent reading mammograms in clinical practice was 23 percent (range, 8 to 50 percent). Seven of the radiologists later said they were not aware that the same patients' mammograms had been shown in phases 1 and 2, two said they had recognized less than 3 percent of the mammograms in phase 2, and one claimed to have recognized about 25 percent of the films.

Observer Variability

Observations, Interpretations, and Recommendations for Management

The observations, diagnostic interpretations, and recommendations for management in phase 1 are shown in Table 1. Among the 27 patients with cancer, a mass was reported in a median of 54 percent (range, 37 to 78 percent) by the 10 radiologists. The diagnostic interpretation of an abnormality suggestive of cancer was reported for 37 to 85 percent of the patients (median, 70), and biopsies were recommended for 33 to 81 percent (median, 65). Figure 1 shows three sets of films for which there was 100 percent agreement on recommended management. Figure 2 shows a set of films from a patient without cancer for which interpretations differed. Interobserver variability in phase 2 closely resembled that in phase 1 and is not presented here.

Agreement on Diagnostic Interpretations and Recommendations

Although all 10 radiologists agreed on the diagnostic interpretation in only 10 cases (7 percent), at least 5 of the 10 radiologists agreed in 129 cases (86 percent). In most instances in which at least five radiologists agreed, the films were interpreted either as normal or as abnormal and suggestive of cancer. The median agreement (Table 2) was 78 percent for interobserver variability in diagnostic interpretations and 85 percent for the recommendation of a biopsy. The corresponding median kappa values were 0.47 for the diagnostic interpretation and 0.49 for the biopsy recommendation. As expected, the corresponding results were higher for intraobserver agreement than for interobserver agreement.

Table 1. Results of 10 Radiologists' Readings of Mammograms from 150 Patients (Phase 1).

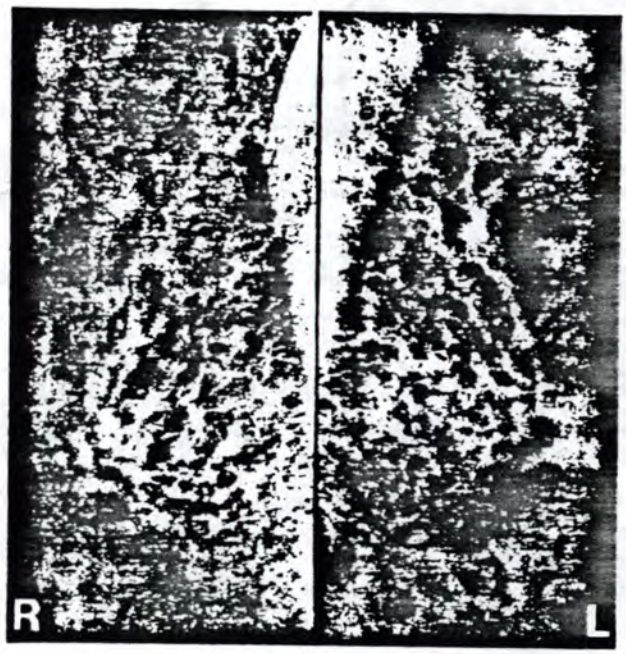
Result	Patients with Cancer (N = 27)	Patients without Cancer (N = 123)
	median percentage (range)	
Observation		
Mass	54 (37–78)	21 (11–34)
Focal asymmetric density	17 (0–52)	17 (3–31)
Calcification	41 (22–48)	21 (14–31)
Diagnostic interpretation		
Normal	7 (0–22)	36 (20–69)
Abnormal, probably benign	9 (4–26)	29 (15–54)
Abnormal, indeterminate	17 (4–33)	21 (7–35)
Abnormal, suggestive of cancer	70 (37–85)	7 (1–15)
Recommendation		
Routine follow-up	11 (4–26)	44 (26–81)
Another mammogram within 6 mo	11 (0–15)	12 (7–41)
Immediate follow-up		
Additional mammographic views	39 (22–70)	36 (6–55)
Ultrasound studies	11 (4–19)	11 (5–24)
Biopsy	65 (33–81)	8 (3–20)

Range of Accuracy

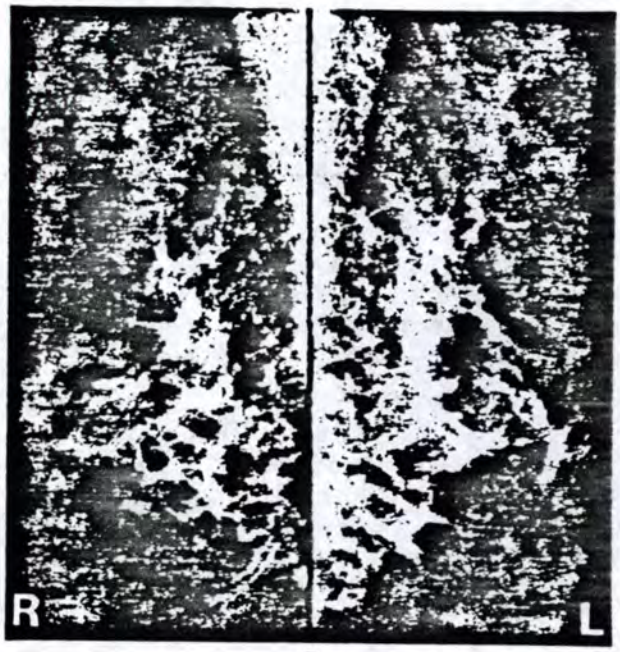
The 10 radiologists differed widely in their recommendations for management (Table 3). For example, the percentage of patients with cancer for whom an immediate workup was recommended was highest (96 percent) for radiologist A and lowest (74 percent) for radiologist J. Radiologist A recommended an immediate workup for almost all the patients with cancer, but also for 64 percent of the patients without cancer. In general, the radiologists who recommended workups most frequently did so for both the patients with cancer and those without cancer. The radiologists varied particularly in requesting additional mammographic views (range for patients without cancer, 6 to 55 percent) but were more consistent in requesting ultrasound examinations and biopsies. The diagnostic interpretations of the 10 radiologists had a median sensitivity of 70 percent (range, 37 to 85 percent) and a median specificity of 94 percent (range, 85 to 99 percent).

Substantial Clinical Disagreements

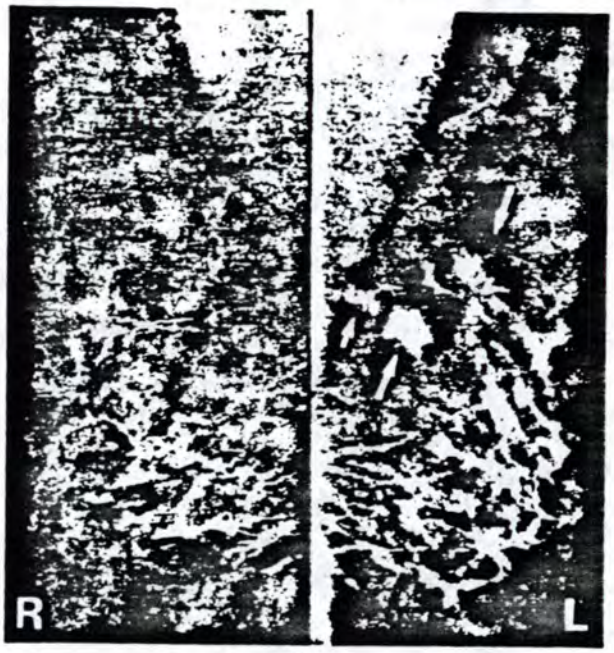
At least one pair of radiologists had a substantial diagnostic disagreement in 2 percent of the pairwise comparisons and for 19 percent of the patients (Table 4). Disagreements regarding biopsy recommendations occurred in 3 percent of the pairwise comparisons and 25 percent of the per-patient comparisons. When at



A



B



C

Figure 1. Examples of Mammograms for Which There Was Complete Agreement on Recommended Management.

Panels A and B show bilateral mediolateral oblique mammograms for two women without cancer, for whom all 10 radiologists recommended routine follow-up. Normal cases can range from predominantly fatty breasts (Panel A) to those with dense parenchyma (Panel B). Panel C shows bilateral mediolateral oblique views from a woman with breast cancer; all 10 radiologists noted spiculated masses with calcifications in the left breast (arrows). All the radiologists thought that these findings were suggestive of cancer and recommended a biopsy. R denotes right breast, and L left breast.

least two radiologists recommended an immediate workup for the same patient, they disagreed on the side (right or left breast) for the workup in 9 percent of pairwise comparisons and 33 percent of per-patient comparisons. In the subgroup of patients for whom biopsies were recommended, the stated side of the abnormality differed in 2 percent of the pairwise comparisons and 9 percent of the per-patient comparisons (four patients).

Assessment of Asymptomatic Patients

In the simulated screening group of 85 asymptomatic patients, the results were similar to the overall results for the 150 patients. For the diagnostic interpretations, the median agreement was 78 percent (kappa index, 0.38), and for the biopsy recommendations, it was 86 percent (kappa index, 0.43). Among the 76 patients without cancer in the simulated screening group, additional mammographic views were recommended for 8 to 67 percent, ultrasound studies for 5 to 17 percent, and biopsies for 4 to 18 percent. The proportion of substantial clinical disagreements in the simulated screening group was also similar to that in the overall group of patients.

Explanations for Variability

Three sources of variability were identified. First, there were differences in visual observation. Some radiologists noted masses where others saw only normal parenchyma, and some missed subtle findings, such as tiny microcalcifications. For example, on being shown a lesion, one radiologist stated, "I just didn't see it." Second, in some cases the same abnormality was perceived differently. Thus, some radiologists regarded

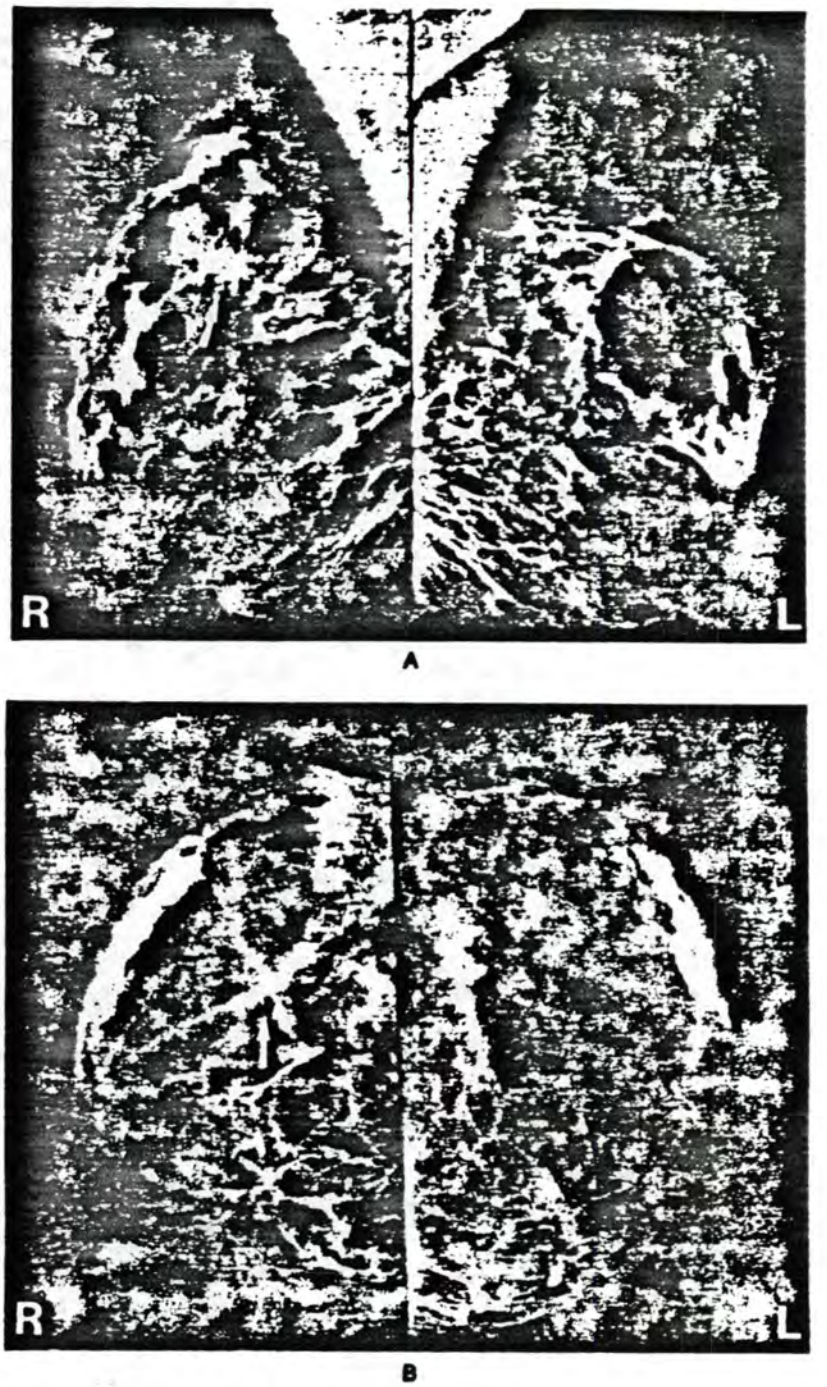


Figure 2. Example of Mammogram for Which There Was Marked Variability in Recommended Management.

Bilateral mediolateral oblique (Panel A) and craniocaudal (Panel B) views are shown. Three radiologists thought the findings were normal, and seven noted an abnormality in the right breast (arrows), which was interpreted as probably benign by two, indeterminate by four, and suggestive of cancer by one. Three radiologists recommended routine follow-up, and seven recommended additional mammographic views; two of the seven also recommended a biopsy. The patient had no evidence of breast cancer during three years of follow-up. R denotes right breast, and L left breast.

an obvious mass as well defined and therefore probably benign, whereas others believed portions of the margin were obscured, making it suggestive of cancer. Third, there were different thresholds of concern about perceived abnormalities. Although some radiologists were willing merely to follow what they regarded as probably benign abnormalities, others sought additional evaluation for almost every perceived abnormality.

Discussion

No study can perfectly mimic the real world. There are some notable differences between the interpretation of mammograms in routine clinical practice and that in our study. The testing situation may have led the radiologists to identify more subtle abnormalities than would otherwise be noted. Previous mammograms were not available for comparison, and a clinical history was not presented for every patient. In addition, the group of patients included many more women with breast cancer or other mammographic abnormalities than would be found in a screening population. Because of this deliberate enrichment, additional workup was recommended in a much larger proportion of patients than the 5 to 14 percent customarily recalled for further evaluation after routine screening.[6,15,16]

The effects of the study situation are difficult to determine. The radiologists may have paid more attention than usual, so that any possible lesion (regardless of its clinical importance) was reported, with a recommendation for further evaluation. On the other hand, knowing these were not films of their actual patients, the radiologists may have taken less care than usual.

Although the usual two views of each breast were evaluated, as they would be during screening, the goal of screening mammography is to separate normal from potentially abnormal mammograms rather than to establish a definitive diagnosis. Therefore, analysis of the radiologists' recommendations for management may be more clinically pertinent than their diagnostic

Table 2. Variability in Diagnostic Interpretations and Biopsy Recommendations (Phase 1).

Interpretation or Recommendation	Interobserver Variability (N = 150 Patients)	Intraobserver Variability (N = 50 Patients)*
	Phase 1	Phase 1 vs. Phase 2
	median (range)	
Diagnostic interpretation		
% Weighted agreement	78 (71–82)	84 (72–89)
Weighted kappa	0.47 (0.31–0.55)	0.57 (0.37–0.71)
Biopsy recommendation		
% Agreement	85 (65–91)	91 (82–98)
Kappa	0.49 (0.20–0.69)	0.71 (0.46–0.91)

*In phase 2, one radiologist did not give a diagnostic interpretation or a biopsy recommendation for one patient.

interpretations. In addition, measures of sensitivity and specificity may not truly reflect the diagnostic accuracy because of the criteria used. For example, a true positive result for sensitivity was defined as a finding designated as abnormal and suggestive of cancer, with a subsequent confirmation of cancer. Accuracy might have been better appraised if true positive results had been defined as recommendations for immediate workups in patients with cancer (range, 74 to 96 percent). Analogously, false positive results could have been defined as recommendations for immediate workups in patients without cancer (range, 11 to 65 percent).

In the absence of an ideal way to calculate clinical disagreements, results were examined for both the pairwise and the per-patient assessments. Because 10 radiologists participated, the per-patient assessments inflate the amount of variability by counting a disagreement for a patient if any one radiologist disagreed with any one of the other nine, even if those nine were in perfect agreement with one another. The pairwise analysis avoids the excesses of the per-patient method but may underestimate the amount of variation, because disagreements were determined with a denominator of 6750 (or 45 × 150), which is the maximal number of possible pairwise agreements. Although reported here for easy understanding, this denominator is nearly twice the number of possible pairwise disagreements, which is 25 per patient and 3750 (or 25 × 150) for the total. In addition, the data were not matched for all analyses according to the specific patient, the side of the abnormality (right or left breast), or the stated location of the abnormality within the breast. For example, if two radiologists each recommended a biopsy for the same patient, the recommendations would be scored as an agreement, even though the abnormality could have been located in the left breast by one radiologist and in the right by the other.

The review after the study indicated that the participating radiologists may have observed the same thing (e.g., a mass in the left breast) but interpreted it differently. They also may have made the same diagnostic interpretation (e.g., an abnormality that was probably benign) but differed in their recommendations for follow-up because of different thresholds of concern. For each radiologist, the threshold may reflect not only personal estimates of the probability of cancer but also subjective factors (such as fiscal, physical, and psychological risks for the patient, as well as the risk of malpractice for the physician). Future efforts to reduce variability must include accuracy as a goal, because accuracy cannot be improved if radiologists agree on a diagnosis that is wrong.

Our results show that radiologists can differ, sometimes substantially, in their mammographic interpretations and recommendations for management. These results should not be regarded as casting doubt on the efficacy of mammography, the value of which has been well documented.[17-23] The variability found in this study, however, indicates that among radiologists who read mammograms there is a wide range of accuracy. Recent efforts to improve the accuracy of mammographic interpretation and to standardize it include the accreditation program of the American College of Radiology,[14] publication of a lexicon for describing mammographic findings,[24,25] decision-making aids,[26] guidelines for mammographic reporting,[27] and a description of auditing procedures.[15,28] A reduction in variability will probably require more consistent criteria for diagnostic interpretation and standards for recommending subsequent evaluation. In addition, examinations of performance will probably be required,

Table 3. Management Recommendations by the 10 Radiologists (Phase 1).

Radiologist	Patients with Cancer (N = 27)				Patients without Cancer (N = 123)			
	Any immediate workup*	Biopsy	Ultrasound	Additional views	Any immediate workup*	Biopsy	Ultrasound	Additional views
	percentage of patients							
A	96	82	11	41	64	20	8	53
B	96	70	11	52	40	7	11	34
C	93	63	19	67	65	9	24	55
D	93	78	11	22	57	13	15	42
E	89	70	15	26	41	14	14	24
F	85	48	19	70	45	3	11	43
G	85	67	7	30	44	5	11	38
H	81	52	15	37	38	10	14	25
I	78	48	4	41	30	3	7	25
J	74	33	7	37	11	3	5	6

*Any immediate workup was defined as a recommendation to obtain additional mammographic views, ultrasound studies, or a biopsy.

Table 4. Substantial Disagreements in Diagnostic Interpretations and Recommendations for Management.

Interpretation or Recommendation	Example of Disagreement		Pairwise Comparison*		Per-Patient Comparison†	
	Reading 1	Reading 2	Interobserver	Intraobserver	Interobserver	Intraobserver
			no. with disagreement/total no. (%)			
Diagnostic interpretation	Normal	Abnormal, suggestive of cancer	130/6750 (2)	5/499 (1)	28/150 (19)	4/50 (8)
Management recommendation	Routine follow-up	Biopsy	234/6750 (3)	5/499 (1)	37/150 (25)	4/50 (8)
Side for recommended workup‡	Left breast	Right breast	209/2268 (9)	19/247 (8)	41/123 (33)	16/45 (36)
Side for recommended biopsy‡	Left breast	Right breast	18/745 (2)	4/122 (3)	4/46 (9)	4/28 (14)

*The numerator is the number of pairwise comparisons involving disagreements. The denominator is the maximal number of possible pairwise comparisons. In phase 2, one radiologist did not give a diagnostic interpretation or a biopsy recommendation for one patient.

†The numerator is the number of patients for whom there were disagreements. The denominator is the total number of patients.

‡Disagreements were assessed in the subgroup of patients for whom more than one radiologist recommended a biopsy or an immediate workup (additional mammographic views, ultrasound studies, or a biopsy). The numerator is the number of comparisons (pairwise or per patient) involving a disagreement in the stated location of an abnormality. The denominator is the maximal number of comparisons (pairwise or per patient) when a side was specified by at least two radiologists.

rather than merely the dissemination of criteria; moreover, collaborative efforts, self-auditing procedures, and specialized education will be needed.

We are indebted to Drs. R.I. Horwitz and J.F. Jekel for their thoughtful comments and to the 10 radiologists who participated in the study and who prefer to remain anonymous.

References

1. Boyd NF, Wolfson C, Moskowitz M, et al. Observer variation in the interpretation of xeromammograms. J Natl Cancer Inst 1982;68:357-63.
2. *Idem*. Observer variation in the classification of mammographic parenchymal patterns. J Chronic Dis 1986;39:465-72.
3. Vineis P, Sinistrero G, Temporelli A, et al. Inter-observer variability in the interpretation of mammograms. Tumori 1988;74:275-9.
4. Ciccone G, Vineis P, Frigerio A, Segnan N. Inter-observer and intra-observer variability of mammogram interpretation: a field study. Eur J Cancer 1992;28A:1054-8.
5. Baines CJ, McFarlane DV, Miller AB. The role of the reference radiologist: estimates of inter-observer agreement and potential delay in cancer detection in the national breast screening study. Invest Radiol 1990;25:971-6.
6. Bird RE. Low-cost screening mammography: report on finances and review of 21,716 consecutive cases. Radiology 1989;171:87-90.
7. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978;299:926-30.
8. (SAS) Version 6.04. Cary, N.C.: SAS Institute, 1987.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
10. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley, 1981.
11. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;70:213-20.
12. Kramer MS, Feinstein AR. Clinical biostatistics. LIV. The biostatistics of concordance. Clin Pharmacol Ther 1981;29:111-23.
13. Feinstein AR, Cicchetti DV. High agreement but low kappa. I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543-9.
14. McLelland R, Hendrick RE, Zinninger MD, Wilcox PA. The American College of Radiology Mammography Accreditation Program. AJR Am J Roentgenol 1991;157:473-9.
15. Sickles EA, Ominsky SH, Sollitto RA, Galvin HB, Monticciolo DL. Medical audit of a rapid-throughput mammography screening practice: methodology and results of 27,114 examinations. Radiology 1990;175:323-7.
16. Dershaw DD, Liberman L, Lippin BS. Mobile mammographic screening of self-referred women: results of 22,540 screenings. Radiology 1992;184:415-9.
17. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. J Natl Cancer Inst 1982;69:349-55.
18. Collette HJ, Day NE, Rombach JJ, deWaard F. Evaluation of screening for breast cancer in a nonrandomised study (the DOM project) by means of a case-control study. Lancet 1984;1:1224-6.
19. Verbeek AL, Hendriks JH, Holland R, Mravunac M, Sturmans F, Day NE. Reduction of breast cancer mortality through mass screening with modern mammography: first results of the Nijmegen project, 1975-1981. Lancet 1984;1:1222-4.
20. Tabár L, Fagerberg CJ, Gad A, et al. Reduction in mortality from breast cancer after mass screening with mammography. Lancet 1985;1:829-32.
21. Palli D, Del Turco MR, Buiatti E, et al. A case-control study of the efficacy of a non-randomized breast cancer screening program in Florence (Italy). Int J Cancer 1986;38:501-4.
22. Andersson I, Aspegren K, Janzon L, et al. Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. BMJ 1988;297:943-8.
23. UK Trial of Early Detection of Breast Cancer Group. First results on mortality reduction in the UK Trial of Early Detection of Breast Cancer. Lancet 1988;2:411-6.
24. D'Orsi CJ, Kopans DB. Mammographic feature analysis. Semin Roentgenol 1993;28:204-30.
25. Breast imaging — reporting and data system (BI-RADS). Reston, Va.: American College of Radiology, 1993.
26. Swets JA, Getty DJ, Pickett RM, D'Orsi CJ, Seltzer SE, McNeil BJ. Enhancing and evaluating diagnostic accuracy. Med Decis Making 1991;11:9-18.
27. Kopans DB. Standardized mammography reporting. Radiol Clin North Am 1992;30:257-64.
28. Sickles EA. Quality assurance: how to audit your own mammography practice. Radiol Clin North Am 1992;30:265-75.

THE ACCURACY OF MAMMOGRAPHIC INTERPRETATION

IT should come as no surprise that there is a range of skills and expertise among physicians involved in similar activities. This is true in every endeavor, and it is true for radiologists who interpret mammograms. The findings reported in this issue of the *Journal* by Elmore et al.[1] are therefore not unexpected. The apparent magnitude of the variation among radiologists, however, is in part a result of the trial design itself, which amplified the rates of disagreement. Of the 150 women with mammograms reviewed by the 10 radiologists studied by this group, 54 (36 percent) had benign lesions, which is a much larger percentage than would be expected in practice. The average prevalence of breast cancer is 6 to 8 per 1000 women. The study included 27 cancers, for a rate of 180 per 1000, or more than 10 times the rate in an actual screening program, yet the radiologists were instructed to simulate their own practice. In addition, almost one third of the cancers were sufficiently ambiguous in their original presentation that one was missed and seven others were misinterpreted on the original readings in 1987.

The interpretive categories used by Elmore et al. oversimplified the task facing radiologists, producing more disagreement. For example, whether one radiologist interprets a film with a benign calcifying fibroadenoma as "normal" while another classifies it as "abnormal, probably benign" is trivial and of no clinical consequence, yet it amplifies the apparent variability in interpretation.

The authors provide no indication of the histologic types of cancer included in the study. Since calcifications were included, some of these lesions were probably ductal carcinomas in situ. There is disagreement about the importance of such lesions, and their management is also an area in which experts do not completely agree. The actual importance of the variation in the recommendations for managing calcifications is uncertain, and how much this variation contributed to the overall disagreement cannot be assessed. The authors provide no information about the inclusion of circumscribed cancers. Only 2 to 5 percent of circumscribed masses prove to be malignant. One also hopes that lobular carcinoma in situ was not included among the cancers, since it is always found serendipitously and is best considered a lesion associated with a high risk for the development of cancer rather than a cancer itself.

Even though mammography is effective only as a screening test, the study was designed primarily to evaluate diagnostic accuracy, despite the effort to select a subgroup to simulate screening. As demonstrated repeatedly, mammography is not an accurate diagnostic test, because there is a considerable overlap in the appearances of benign and malignant lesions on mammograms, and the procedure therefore cannot be used to rule out breast cancer. Clinicians are regularly reminded not to rely on a mammogram in determining whether a palpable abnormality is benign or malignant. The inclusion of mammograms from women with palpable abnormalities is another problem, since negative mammograms in these cases are irrelevant. Such abnormalities would have to be evaluated clinically regardless of the results of mammography.[2]

The fact that mammography cannot be used to rule out cancer does not negate its importance in detecting tumors at an early stage, but it is rarely possible to diagnose an abnormality (i.e., determine its histologic characteristics) on a screening study. It is possible to estimate the probability of cancer, given the abnormalities found by mammography. There is, however, no agreement about the appropriate threshold for intervention. A positive predictive value of 15 to 25 percent for biopsies of lesions detected by clinical breast examination has long been considered acceptable for intervention.[3] It should be reasonable for this rate to be considered acceptable for mammography as well. Should one ignore a lesion with a 10 percent probability of being malignant? Should the threshold be 50 percent? If the woman involved is consulted, she may wish to undergo a biopsy even if there is only a 1 percent chance of cancer. She certainly should be given the option. The variety of interpretations in the study by Elmore et al. reflects variations in the thresholds for intervention used by the 10 radiologists.

It is disturbing that the interpreters failed to identify some patients with cancer. Without any way of assessing how conspicuous the lesions were in the cases chosen for review, it is difficult to determine the meaning of the results. Almost 30 percent of the cancers included in the study were incorrectly categorized when the films were originally interpreted in 1987 (the "worst" of the 10 study radiologists actually had a higher rate of sensitivity). Since the expected false negative rate for mammography is 10 to 15 percent, the case mix included an unusually high percentage of subtle cancers.

One of the radiologists later admitted, "I just didn't see it." Studies have shown that there is a perceptual threshold that causes everyone occasionally to overlook an important abnormality. This is true no matter how experienced the observer. It has been shown for several types of lesions, including lung nodules,[4] fractures, and polyps, as well as abnormalities on mammograms, even though the same observer can easily see the abnormality in retrospect. This phenomenon is similar to the experience of searching for something and looking directly at it (such as a car key) but not seeing it even though it is in plain view. Since observers differ in what they miss, it may be preferable for two observers to review each mammogram. In one study, an additional 5 percent of cancers were discovered by double reading.[5] The Swedish screening programs, considered among the best in the world, have shown that, even among highly expert radiologists, double reading increases the rate of cancer detection by 15 percent.[6,7]

This study reaffirms the basic problems of observational interpretation. Since the characteristics of cancers are well known to radiologists, the interpretation of a given abnormality, assuming it was seen by the ob-

server, is based on his or her degree of concern about the finding. A high rate of cancer detection can be achieved only at the expense of a high false positive rate. Conversely, if the goal, as is increasingly emphasized today, is to reduce the false positive rate, then some cancers will be overlooked.[8] The results of the current study clearly demonstrate this problem. The radiologist who diagnosed the most cancers had a high false positive rate, and the radiologist who diagnosed the fewest cancers had a low false positive rate. A better way to assess the success of a screening program is to evaluate its sensitivity, as well as the size and stage of the detected cancers.

The National Cancer Institute has suggested that mammography is of no benefit in screening women between the ages of 40 and 49 years for breast cancer. Yet the National Cancer Institute continues to recommend clinical breast examination,[9] even though its sensitivity is much lower than that of mammography, the findings of a higher percentage of biopsies performed on the basis of clinical examination prove to be benign, and tumors detected by clinical examination are often larger and at a later stage than those detected by mammography.[10] Nevertheless, the skills of the average physician who performs clinical breast examination have been ignored. The variation in the performance of clinical breast examination is probably much greater than the variation in the interpretation of mammograms.

The radiologic community has been working for many years to improve all aspects of breast-cancer detection and diagnosis and has made considerable improvements. As Elmore et al. point out, in 1987 the American College of Radiology developed the Mammography Accreditation Program, which establishes standards and provides credentials for interpreting mammograms. The American College of Radiology conducts numerous training programs throughout the country, has developed a system to standardize the language of image interpretation, and has worked closely with the Food and Drug Administration to implement the Mammography Quality Standards Act. Together with the Centers for Disease Control and Prevention, the American College of Radiology has developed a curriculum for training in the detection and diagnosis of breast cancer. Radiologists have worked hard to improve the quality of health care services associated with the detection of breast cancer. Improving the quality of mammographic interpretation continues to be a critical element of this effort.

Harvard Medical School
Boston, MA 02115

DANIEL B. KOPANS, M.D.

REFERENCES

1. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. N Engl J Med 1994;331:1493-9.
2. Kopans DB. Breast imaging and the standard of care for the symptomatic patient. Radiology 1993;187:608-11.
3. Spivey GH, Perry BW, Clark VA, Coulson AH, Coulson WF. Predicting the risk of cancer at the time of breast biopsy: variation in the benign to malignant ratio. Am Surg 1982;48:326-32.
4. Revesz G, Kundel HL. Psychophysical studies of detection errors in chest radiology. Radiology 1977;123:559-62.
5. Bird RE. Professional quality assurance for mammography screening programs. Radiology 1990;177:587.
6. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol Clin North Am 1992;30:187-210.
7. Thurfjell EL, Lernevall KA, Taube AA. Benefit of independent double reading in a population-based mammography screening program. Radiology 1994;191:241-4.
8. D'Orsi CJ. To follow or not to follow, that is the question. Radiology 1992; 184:306.
9. Kopans DB, Feig SA. The Canadian National Breast Screening Study: a critical review. AJR Am J Roentgenol 1993;161:755-60.
10. Bassett LW, Liu TH, Giuliano AE, Gold RH. The prevalence of carcinoma in palpable vs impalpable, mammographically detected lesions. AJR Am J Roentgenol 1991;157:21-4.

CLINICAL IMPLICATIONS OF BASIC RESEARCH

SUSCEPTIBILITY GENES FOR BREAST CANCER

In an era in which the importance of a scientific advance often appears to be measured by the amount of media coverage it generates, the level of reporting of the recent identification of *BRCA1*, a susceptibility gene for breast cancer on chromosome 17 and the first such gene to be isolated, suggests that this discovery is the most exciting breakthrough in modern times. Earlier news coverage of the "race" to find *BRCA1* was noteworthy for the amount of media attention focused on something that had yet to be accomplished. This general interest in *BRCA1* is illustrated by the fact that the announcement that the gene had finally been isolated was the lead story on the *NBC Nightly News* three weeks before the report appeared in *Science*,[1] breaking the prepublication embargo imposed by the journal's editors. In this case, the use of the media as a meter of importance may be justified, given the clinical and scientific ramifications of this finding. In addition, questions asked by reporters in the wake of this spectacular scientific accomplishment focus on the very issues faced by scientists and clinicians alike: What does the identification of *BRCA1* mean for women with or at risk of breast cancer? When will presymptomatic testing for breast cancer become clinically feasible, and what are the risks and benefits of such testing? What should we tell women with inherited mutations in *BRCA1*? What should we tell their families? Will the isolation of *BRCA1* advance our understanding of the causes of breast and ovarian cancer?

Positional cloning, the approach used to isolate *BRCA1*, is a method of finding disease-related genes on the basis of their approximate chromosomal location. This approach is based on the genetic analysis of families with inherited forms of the disease under study, in this case breast cancer. As a result, families likely to harbor germ-line mutations in *BRCA1* and family members likely to be carriers of the mutations were identified long before *BRCA1* was isolated. Study of these families provided data on the incidence and clinical characteristics of breast cancer in putative carriers of *BRCA1* mutations, and the members of these families were the first in the world to be offered predictive testing for breast and ovarian cancer. On the basis of findings in these families, it appears that women with germ-line *BRCA1* mutations have an astonishing 85 percent lifetime risk of breast cancer; breast cancer will develop in more than half of these women before the age of 50 years.[2] Thus, although inherited *BRCA1* mutations may be responsible for less than 2 percent of the cases of breast cancer diagnosed after the age of 70 years, as many as 30 percent of women given a diagnosis of breast cancer before the age of 45 may be carriers of these mutations. These studies[2] further suggest that mutations in *BRCA1* are responsible for approximately half of all cases of inherited breast cancer.

BRCA2, the second susceptibility gene for breast cancer to be recognized,[3] has recently been localized to a portion of chromosome 13, but it remains to be isolated. Mutations in *BRCA2* are thought to account for approximately 70 percent of cases of inherited breast cancer that are not due to *BRCA1* mutations and are associated with an increased risk of breast cancer in men. There does not appear to be any increased risk of male breast cancer associated with germ-line mutations in *BRCA1*, but the risk of ovarian cancer in women with germ-line *BRCA1* mutations is substantial. The lifetime risk of ovarian cancer associated with *BRCA1* mutations is more variable than the risk of breast cancer but ranges from 20 to 50 percent. In addition, carriers of *BRCA1* mutations may be at increased risk of colon and prostate cancer.[4] The histologic or biochemical characteristics of the tumors resulting from *BRCA1* mutations are not obviously different from those of sporadic tumors, nor does there appear to be a difference in the stage-for-stage prognosis.

The *BRCA1* DNA and protein sequences made available by the recent isolation of this gene provide additional information.[1] *BRCA1* is an extremely large gene, with exons spread over approximately 100,000 base pairs of genomic DNA. The *BRCA1* messenger RNA (mRNA) is 7.8 kb in length, and the protein consists of 1863 amino acids. The initial reports describe eight different mutations[1,5]; six are point mutations, one is a deletion of 11 base pairs, and one appears to be a mutation outside the coding region that results in the complete loss of *BRCA1* mRNA. Many other mutations are likely to be described in the coming months, as additional investigators gain access to the reagents necessary to detect mutations in samples from persons unrelated to the initial study families. The *BRCA1* protein sequence contains a zinc-finger motif often found in transcription factors (which regulate the transcription of DNA into mRNA) but otherwise appears unrelated to any previously described proteins.

Examination of a panel of 46 unselected samples of breast and ovarian tumors (i.e., samples from patients with disease not known to be familial) provided surprising information: four previously unsuspected germ-line mutations were discovered, but not one somatic mutation was detected.[5] The absence of noninherited *BRCA1* mutations in these unselected tumors may be due to difficulty in detecting mutations in tumors admixed with normal tissue or may suggest that acquired *BRCA1* mutations do not play an important part in the development of sporadic tumors. These data are particularly difficult to reconcile with observations of chromosomal alterations flanking *BRCA1* in sporadic tumors and have raised the possibility of a second, unidentified gene in the region that may be involved in the development of noninherited breast cancer. Thus, although predictions that *BRCA1* would be a classic tumor-suppressor gene and would be important in the de-

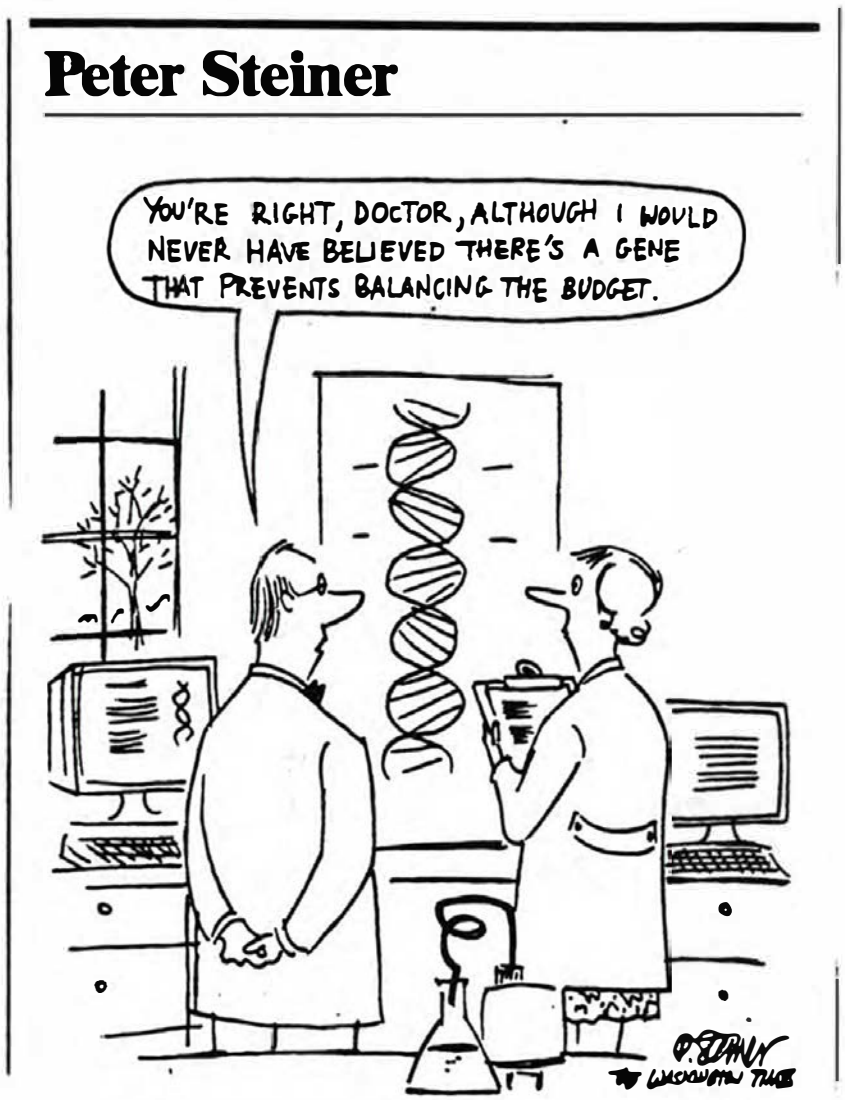
velopment of as many as 50 percent of cases of noninherited breast cancer were well grounded, they may not be borne out.

In answer to reporters' questions about why the isolation of *BRCA1* is so important, one can say that this is not the end of the race but the beginning of opportunities in many exciting areas of investigation. Scientists now have the tools to study this gene, the function of which appears critical in preventing the most common cancer in American women. We also have the ability to identify persons carrying inherited *BRCA1* mutations, estimated to be 1 in 200 to 1 in 400 people in the United States, without the need to study large, severely affected families. What will eventually follow is a fundamental understanding of the function of *BRCA1* and ways to exploit that knowledge in efforts to diagnose, treat, and ultimately prevent breast cancer. We now need the means to make screening for mutations in this large gene accessible and accurate, a consensus on how best to provide women with this information, and the patience to ensure that these critical initial goals are accomplished before testing for germ-line *BRCA1* mutations becomes a part of everyday medical practice.

University of Pennsylvania
Medical Center
Philadelphia, PA 19104-6306 BARBARA L. WEBER, M.D.

REFERENCES

1. Miki .., Swensen J, Shattuck-Eidens D, et al. Isolation of BRCA1, the 17q-linked breast and ovarian cancer susceptibility gene. Science 1994;266:66-71.
2. Easton DF, Bishop DT, Ford D, Crockford GP. Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. Am J Hum Genet 1993;52:678-701.
3. Wooster R, Neuhausen SL, Mangion Y, et al. Localization of a breast cancer susceptibility gene, *BRCA2*, to chromosome 13q12-13. Science 1994;265:2088-90.
4. Ford D, Easton DF, Bishop DT, Narod SA, Goldgar DE. Breast Cancer Linkage Consortium. Risks of cancer in *BRCA1*-mutation carriers. Lancet 1994;343:692-5.
5. Futreal PA, Liu Q, Shattuck-Eidens D, et al. BRCA1 mutations in primary breast and ovarian cancers. Science 1994;266:120-2.



JOURNAL
OF THE NATIONAL CANCER INSTITUTE

Date: November 29, 1994

MEMORANDUM FOR: Science Writers and Editors on the Journal Press List

From: Julianne Chappell, Managing Editor (301) 496-6975

Subject: Highlights from December 7 Issue, No. 23
(EMBARGOED FOR RELEASE 6 P.M. EST DECEMBER 6)

By-Product of Chemical in Vegetables Both Mimics and Counters Estrogen Effects on Breast Tumor Cells

A chemical by-product of a compound found in broccoli, cabbage, Brussels sprouts, and cauliflower can both inhibit breast tumor cell growth by acting as an anti-estrogen and promote growth by mimicking the effects of estrogen, according to a study reported in the December 7 Journal of the National Cancer Institute.

Hong Liu, Ph.D., Texas A&M University led a laboratory study to measure the antiestrogenic and estrogenic effects of ICZ (indolo[3,2-b]carbazole) on human breast cancer cells. ICZ is formed when I3C (indole-3-carbinol), a compound found in vegetables of the Brassica genus, is placed in acidic solution (as is found in the stomach). I3C and ICZ have been shown to bind to the aryl hydrocarbon (Ah) intracellular receptor, inducing activity of P-450 enzymes, which have a major role in the metabolism of many environmental carcinogens, drugs, and hormones, including estrogen. I3C also inhibits mammary tumor development in rodents.

In a series of experiments, ICZ induced P-450 activity in the human breast cancer cells, and elicited a number of anti-

--more--

estrogenic responses, including inhibition of estrogen-induced cell proliferation. ICZ also exhibited a weakly estrogenic effect both in the cells and in binding to the cells' estrogen receptors. Liu et al. suggest that ICZ and other Ah receptor binding compounds inhibit estrogen activity through an interaction between the Ah and estrogen receptor systems.

The authors believe suggestions that environmental estrogens may contribute to increased breast cancer incidence in women and reproductive tract disorders in men fail to account for the existence and possible counteractive effects of both dietary and environmental estrogens and antiestrogens. More information is needed, they say, on dietary levels, potencies, and interactive effects of estrogenic and antiestrogenic chemicals. ICZ, they note, is a unique compound binding to both Ah and estrogen receptors, thereby contributing to the dietary intake of both estrogens and antiestrogens.

In an accompanying editorial, Barry R. Goldin, Ph.D., Tufts University School of Medicine, emphasizes that the major research challenge in preventing or treating estrogen related cancers is to discover compounds with much higher estrogen inhibitory than estrogenic activity that can reach the necessary concentrations in target body tissues to exert a beneficial health effect. This can be achieved in part with synthetic compounds such as tamoxifen, he notes, but use of naturally occurring substances could eliminate side effects and permit safe, long-term preventive usage.

Contact: Gene Charleton, Texas A&M University, (409) 845-4644. Editorial: Dan Quinn, (617) 956-5705.

--30--

Note: This memo to reporters is from the Journal staff and is not an official release of the National Cancer Institute; nor does it reflect NCI policy. In addition, unless otherwise stated, all articles and items published in the Journal reflect the individual views of the authors and not the official points of view held by NCI, any other component of the U.S. government, or the organizations with which the authors are affiliated. Neither NCI nor any other component of the U.S. government assumes any responsibility for the completeness of the articles or other items or the accuracy of the conclusions reached therein.

The Invisible Epidemic

Chronic-fatigue syndrome is a debilitating disease diagnosed far more often in women. Why is it being ignored? BY HILLARY JOHNSON

WILHELMINA JENKINS WAS WORKING AS A GOVERNment physicist and a teacher at Howard University when she suddenly fell ill with a disease absolutely no one was able to diagnose. Jenkins, then 33, had many symptoms, including profound exhaustion. But the most troubling was a rapid loss of mental clarity.

For five years, Jenkins tried to work around the problem by taking a less pressured teaching job and then quitting regular work altogether to do independent research, but her decline continued. Finally, in 1988, a psychiatrist sent her to an internist, who determined that she suffered from what had recently been named chronic-fatigue syndrome (CFS), a neurological, metabolic and immunological disease that can disable victims for years (see box, page 71). Early on, sufferers typically experience a range of symptoms—from an ongoing, low-grade fever and lymph-node pain to night sweats and temporary amnesia—but the persistent hallmarks are a near-paralytic weakness and what patients often describe as a constant mental fog.

For Jenkins, as for many patients, the emotional toll has been nearly as great as the physical debilitation. "I felt I had been robbed," she says. "The things that I had been told could never be taken away from me—my education, my intelligence—had been taken away. Now there are days when I cannot get through a comic strip. The emphasis on fatigue in this disease is overplayed—the cognitive problems are what break you."

Before 1984, few doctors had ever encountered such an illness; by the end of the decade, the phenomenon was common. Precise measurements of the epidemic's breadth have yet to be made, but some estimates suggest that a million or more people are sick, to varying degrees, with this incurable malady. Women are diagnosed with CFS at least twice and perhaps three times as often as men. Among the rich and famous who were open about their struggle with CFS were Cher and Gilda Radner, who suffered from the disease during the year preceding her ovarian-cancer diagnosis (preliminary research suggests that CFS may lead to cancer in some sufferers). A number of other public figures are known by insiders to suffer from CFS but are unwilling to reveal their condition.

Although one wag has called the disease "girl AIDS," CFS afflicts men and children as well as women; three and a half years ago, Jenkins's daughter, Kamilah Neal, then 14, was diagnosed with it. (Several CFS researchers have observed that, not infrequently, the disease spreads within families.) Nevertheless, measured by years of productivity lost rather

than by mortality, CFS may well pose the most serious health threat to women in their prime.

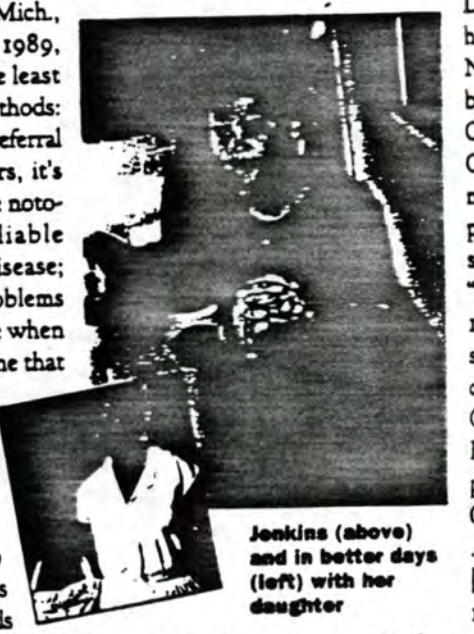
During the late '80s, TV and print journalists were on familiar terms with the disease, usually calling it the "yuppie flu" because it seemed to strike mostly upper-middle-class whites. By 1990, as a result of some intriguing scientific findings, the disease had become a popular news topic. Yet today CFS has fallen into a kind of Bermuda Triangle of misplaced epidemics. Powerful forces have conspired to render it invisible—namely, doctors confounded by the disease and federal health officials who, burdened with projects they deemed more important, hoped the brouhaha would simply blow over. Indeed, that a public-health issue of this magnitude has been overlooked—in effect, "disappeared"—is one of the most remarkable chapters in 20th-century medicine.

The Data Shortage

Key to understanding any new disease is assessing its breadth. Yet close to 10 years after a highly publicized outbreak in Incline Village, Nev., no one is certain how many people in the U.S. have CFS or how it is spread. Responsibility for this crucial task rests with the Atlanta-based Centers for Disease Control and Prevention (CDC). The agency's first brush with the disease was its perfunctory and unrevealing investigation of the 1985 Nevada outbreak. A few influential scientists and doctors soon advanced the notion that CFS was a psychiatric disorder. How else to explain why healthy, productive members of society would suddenly drop out, claiming an array of dramatic-sounding symptoms?

Congress harbors a distaste for interfering in the scientific activities of federal health agencies like the CDC, on the reasonable grounds that science can't be legislated. Bowing to pressure from patients, though, it shed its neutral role in the late 1980s, much as it had done with AIDS, and began demanding more and better CFS research. Each year for the last seven, Congress has asked the Atlanta agency to perform national surveillance, appropriating ever-larger sums of money—from $400,000 in '88 to $4.7 million this year. The cash-starved agency has funneled the money into its various labs and scientific branches in the name of CFS research, but its epidemiologists have yet to perform reliable national surveillance.

Assuming that the disease was extremely rare, if it existed at all, the agency's epidemiologists launched a limited "passive-surveillance" program in Atlanta, Reno, Wichita, Kans., and Grand Rapids, Mich., in September 1989, using one of the least trustworthy methods: a doctor-based referral system. Doctors, it's been noted, are notoriously unreliable reporters of disease; imagine the problems that might arise when the disease is one that few doctors understand well enough to diagnose and that many have been led to believe exists only in the minds of malingerers, bored housewives and excitable members of the press. In addition, reported victims had to be subjected to hours of psychological evaluation and have their cases approved by a panel of researchers—several of whom had little practical experience with CFS—using the CDC's definition of the disease, which has been extensively criticized, even within the agency. (A revised version, which elevates neurocognitive symptoms from a minor to a defining characteristic of the disease, is expected to be published in the *Annals of Internal Medicine* next month.) Walter Gunn, the CDC's principal CFS investigator until his retirement in 1991, frequently attended CFS support-group meetings in Atlanta. "There were inevitably 100 to 200 people there," Gunn says. "I always asked how many were enrolled in our surveillance program. Usually, just a few raised their hands."



Jenkins (above) and in better days (left) with her daughter

A year ago, a CDC report stated that there were 3,000 to 10,000 people in the U.S. with chronic-fatigue syndrome, or an average of 4.5 victims for every 100,000 adults (the agency recently changed the figure to four to 10 cases per 100,000). The low assessment stunned students of the disease, including Dr. Paul Cheney, an internal-medicine specialist in Charlotte, N.C., who, with his then-partner, Dr. Daniel Peterson, had brought the Nevada outbreak to the CDC's attention. Cheney, who now sees CFS patients exclusively, notes that "at four per 100,000, there should be 40 cases in the Charlotte area. But I have 300 patients from Charlotte alone. . . . I would say [the CDC] is off 10- to 100-fold."

In several independent studies, epidemiologists have found vastly higher rates of disease by surveying populations, a more accurate means of measuring disease prevalence. Researchers at Harvard Medical School and the University of Washington in Seattle, studying the incidence of CFS among members of a Seattle HMO, recently arrived at an estimate of 98 to 267 cases per 100,000 people, up to 90 times the CDC's initial estimate for the general population. Such a rate suggests that CFS is more prevalent than a well-known disease like multiple sclerosis. Investigators who have looked at populations in which CFS rates were thought to be even higher have had their suspicions confirmed. Epidemiologist Sandra Daugherty of the University of Nevada surveyed the populations of Incline Village and neighboring Crystal Bay and found

that more than four of every 100 people had fallen ill with CFS in the previous seven years—a vastly higher number than the government's. An ongoing study of Chicago nurses being done by DePaul University clinical sociologist Leonard Jason offers a second model. "The rates that we are coming up with are approximately 680 per 100,000 nurses," says Jason.

The big numbers in the Nevada prevalence study suggest "time-space clustering," a signal that the disease may be transmissible, as patient activists and some CFS experts have long believed. The Chicago study supports another theory, held by Gunn and others, that there is an unusually high degree of CFS among medical professionals. Many experts believe their vulnerability may result from exposure to sick people, some of whom are CFS sufferers yet to be diagnosed.

According to Dr. Keiji Fukuda, who is assigned to the disease at the CDC, agency staffers are lately admitting that their passive-surveillance system "undercounted" patients. Now, says Fukuda, "we're going to do it better" by using an active system to look for cases that involves dialing telephone numbers at random in selected cities.

Institutionalized Indifference

In 1988, the CDC officially began to identify the malady as chronic-fatigue syndrome. Besides making the disease sound harmless, the name has had the subtle effect of inspiring hostility toward the victim. Gunn, a psychologist with cross-training in epidemiology, is well versed in standardized systems of evaluating the psychological import of words. "*Chronic-fatigue syndrome* has a real negative impact," he says. "The word *chronic* is associated with chronic complainers, chronic whiners. And *fatigue* is even worse." No wonder CFS has frequently been viewed in professional and lay circles alike as a "volitional" disability.

To some degree, all chronic maladies generate indifference: "It's a cultural ethos," explains University of Illinois psychiatrist and CFS researcher Judith Richman. "Americans are freaked out about death, but disability is invisible. If you're half dead, no one cares."

As a result of their low status within the medical establishment, CFS patients have almost no safety net. As with other chronic illnesses, the disease has been considered a preexisting condition that usually precludes health-insurance coverage. But unlike other chronically ill people, most CFS sufferers have had tremendous difficulty winning disability support from the Social Security Administration, which only recently established a code for the disease. Because they lack the medical and social supports usually accorded to the seriously ill, CFS sufferers often end up going to court to settle cases involving bankruptcy, divorce, child custody, employers' refusal to pay disability and even involuntary confinement in psychiatric hospitals.

Not surprisingly, many CFS sufferers are angry. "The government's just not taking it seriously," says Wilhelmina Jenkins. "There has been no effort, for example, to reach out to the public and explain how very serious this disease is. The loss to this country of people whose skills could be helpful is simply not understood."

Closing the Credibility Gap

If epidemiologists in Atlanta have shortchanged the public on documenting this pressing matter, the scientists at the nation's premier medical-research agency, the National Institutes of Health (NIH) in Bethesda, Md., have hardly done better in investigating the origins of the disease. Since 1987, the NIH's National Institute of Allergy and Infectious Diseases (NIAID) has made a few grants, but in amounts many researchers consider so small as to be mere tokens. In 1991, for instance, the institute funded three CFS research centers at universities. Typically, such centers are awarded multimillion-dollar grants; the CFS centers averaged $450,000 apiece. "You would have to ascend from Atlantis to be able to do the level of work that they thought you could do for that kind of money," says immunologist Nancy Klimas, an AIDS and CFS specialist at the University of Miami.

An "intramural," or in-house, program exists to study the disease at NIAID. For a decade its agenda has been dominated by Dr. Stephen Straus, chief of the Laboratory of Clinical Investigation and an influential architect of the "psychoneurotic" theory. In 1988, he wrote in the *Journal of Allergy and Clinical Immunology* that an appraisal of CFS patients "often uncovers histories of unachievable ambition, poor coping skills and somatic complaints." (One sufferer quipped in response, "I'd like to be in the room when you tell that to Cher.") Straus included no supporting data for his comments. In a 1989 article, he suggested, on the basis of his evaluation of 28 people, that CFS sufferers were psychologically different long before they ever fell ill. Straus's CFS research has cost taxpayers an average of $800,000 for each of the past five years. (He declined to comment for this article.) These days, NIAID administrators say they believe that the physiology of the disease also merits investigation. Says Dr. John La Montagne, who directs the division of microbiology and infectious diseases at NIAID, "This is not a trivial disease—cer-

Money can do many things, but it cannot expunge scientific bias

DO YOU HAVE CHRONIC-FATIGUE SYNDROME?

In spite of its name, fatigue alone will not justify a diagnosis of chronic-fatigue syndrome, nor can the disease be cured by sleep. Most often, CFS begins suddenly. Patients and inexperienced doctors alike often confuse early symptoms with those of the flu or mononucleosis. Profound exhaustion is accompanied by flu-like symptoms such as a sore throat, headaches and muscle and joint pain. In time, patients develop neurological problems like short-term memory loss, difficulty finding words, confusion and balance disturbances. CFS expert Dr. Paul Cheney estimates that for those who are ill longer than four years, the recovery rate is anywhere from 5% to 20%. For more information, contact the CFIDS Association at 800-442-3437. —M.J.

tainly not in the minds of patients."

In the fall of 1990, an immunologist from Philadelphia's Wistar Institute, the nation's oldest independent scientific-research center, unhinged Straus's psychoneurotic theory at a neurology conference in Kyoto. Elaine DeFreitas described her discovery of viral gene fragments in CFS patients, the first time anyone had linked a potentially new virus specifically with the disease. American reporters were galvanized. Formerly a subject of hilarity, the disease now had an enormous horror quotient: It might be "catching." Major papers reported the story; the malady's baptism by media was complete when a CBS reporter pronounced CFS a real disease capable of turning "an active person into a virtual invalid overnight." *Newsweek*'s editors put the story on the November 12 cover, calling CFS a "gray plague," which could be read as a play on either the politically incorrect term "gay plague" or the Black Death. Indeed, "gray plague" seemed an apt synonym for a disease that, unlike the 14th-century scourge, cast its victims into unending illness without killing them. The magazine estimated that there were two to five million Americans suffering from CFS. That issue of *Newsweek* was the hottest-selling of the year.

For a while it seemed that CFS might come to be viewed as the public-health catastrophe its victims and their advocates in medicine believed it was. University of California at San Francisco microbiologist Jay A. Levy, one of the discoverers of HIV, announced that he had created a test to help diagnose CFS based on an immune-system irregularity. The scientist's stature ensured wide coverage of his finding. Unfortunately, Levy's test proved to be less than universal in its ability to identify sufferers, a fate that has haunted many of the discoveries surrounding CFS. In 1993, virologists at the CDC revealed that they had been unable to replicate DeFreitas's work. The embattled scientist, now a professor at the University of Miami Medical School, and her clinician-collaborators, Charlotte's Paul Cheney and Harvard pediatric-CFS expert David Bell, insist that the government failed to employ DeFreitas's more costly and time-consuming protocols; government scientists counter that their own methods were sufficient.

Despite the inconclusiveness of DeFreitas's work, it was one of a series of events that heated up the research climate. The NIH's Straus, for one, has appeared to be engaged in a game of scientific catch-up. Earlier this year, Straus's research group reported finding immunological aberrations in CFS patients. The NIH issued a press release as if this were a breakthrough, but, in fact, many CFS investigators had for years been documenting abnormalities, such as aberrant levels of several kinds of white blood cells.

> "Americans are freaked out about death, but disability is invisible. If you're half dead, no one cares."

In the absence of further high-profile discoveries, awareness of the disease has receded, even if the disease itself hasn't. Although Cheney reports that he continues to see "a lot" of new cases, which he defines as people who have been ill for less than a year, 1987 is the year most frequently named as the one in which CFS sufferers fell ill. It may be that the epidemic peaked in the late 1980s, but, according to Cheney, if it took 10 years to reach a peak, it could take three times that long to level off. (Cases have now been identified dating back to the late '70s, and some theorize that CFS has occurred in isolated outbreaks for decades.) "The exponential growth has lessened," says Cheney, "but the *number* of people getting sick is still growing."

The Misery Index

Despite the government's stance and the disease's name, CFS has one of the highest morbidity rates (degrees of suffering) of any illness, rivaling the physical misery of advanced cardiac disease and cancer. Dr. Philip Peterson, head of the infectious-diseases department at Minneapolis's county medical center, opened a CFS research clinic there in 1988. The average age of the clinic's patients then was 38, and 80% were women. "Roughly half the patients could walk only three blocks or less," Peterson says. He and his collaborators explored the "functional severity," or degree of disability CFS imposes, using the Medical Outcome Study, in which a score of 100 is "best health." They compared their patients' scores with those of healthy people and people who had suffered a heart attack or who had rheumatoid arthritis. Healthy people scored an average of 75, heart-attack patients in the mid-40s and victims of rheumatoid arthritis slightly higher. Clinic patients scored, on average, 16.5, a level never before measured on the Medical Outcome scale. University of Minnesota medical professor Nicole Lurie had to redraw the scale to fit the clinic patients onto a diagram Peterson used for a formal paper on the subject. A 1991 study, published in the *Journal of Clinical Psychiatry*, compared the severity of CFS with that of multiple sclerosis, lupus and Lyme disease and found CFS to involve significantly more debilitating fatigue.

A Grassroots Struggle

The "disappearing" of CFS as a public-health issue can be blamed on a phenomenon even more profound than an antipathy to chronic illness or the impoverishment of science at government agencies. If ever there was a disease that required listening to patients and carefully observing them, this is it. Yet an increasing dependence on medical technology has served to erode and devalue the methods that served physicians for centuries. The art of cognitive medicine lost ground when blood tests, MRI scans and sonograms arrived to provide the diagnosis. In an age of high-tech medicine, the voices of doctors who recognize the singularity of this (Continued on page 74)

ing with treated me like a piece of furniture. And he'll say, 'No, that's your imagination.' It just doesn't bother him, and it bothers the hell out of me."

Fortunately, the Dolans take enormous pride in each other's accomplishments and acknowledge that their collaboration is essential to their individual success. That's not always the case. Consider this horror story, told by Pepper Schwartz, a professor of sociology at the University of Washington and author of *Peer Marriage: How Love Between Equals Really Works:* "I recall one couple who worked as an academic team, a very intense writing team, for years. But she felt he was getting more credit for their ideas than she was, and so she started aggressively promoting herself. Then he wanted to have kids, and she thought of that as a way for him to get further advantage in their intellectual partnership. Ultimately they divorced and went to court for the rights to their collaborations. The judge wound up dividing rights between the two of them. It was not pleasant."

"The secret," says Nelton of this incident, "is to compete with the outside world, not with each other. Men have a harder time with that than women." That's no surprise. Women in America tend to be trained to be more collaborative, while men are rewarded for greater competitiveness.

In the final analysis, most of the problems of couples working together can be solved with equal parts respect and communication. Though it may appear that people who spend most of their waking hours together would do little *but* communicate, that's often not the case. Judie and Craig Pattberg run a business in which they provide free sample goods and services to people who have just moved into their community. In exchange, the businesses that provide the freebies have a shot at new customers. Judie was in charge of visiting the new neighbors, while Craig dealt with the businesses. "She was not comfortable with the packet she was taking and wanted me to get more businesses in to make it better," says Craig. "But I didn't realize how passionate she was—our difference of opinion felt like kind of a game. The situation finally came to a head, and her tears were what got me to come around. Once we actually sat down and discussed it, it was easy to rectify."

Sometimes a time to communicate must be penciled in. Some couples clear the last few hours of the week; others have a working lunch each day. Bill and Diane Black have been co-owners of a database-management service since 1989. Among other things, the business maintains the yearly registration data on 30,000 members for United States Masters Swimming. "We attend an annual convention that runs from Wednesday evening to Sunday morning," Bill explains. "Before the convention, we spend three days traveling and talking with each other without the children, the phone or other distractions. We work out what we want to do with the business and what problems need to be resolved, generally reinforcing the relationship. We can then return to the routine refreshed and invigorated for the new year."

Even with all the communication skills and mutual respect in the world, there are going to be battles. But for us, and for many other couples in partnership, working together is the only way to live. With a sense of humor and common goals, most problems can be overcome . . . eventually.

"Every time my husband and I write a book together, I swear I'll never do it again," says Sharon Crawford. They just finished their eighth book, *Voodoo UNIX.*

"There are times I'm glad we work side by side, and there are times I want to throw him in the river," adds Kathleen Zimmerman, who runs a dental laboratory with her husband, Rob. "It has its ups and downs, but it's mostly ups."

"It's not good or bad, working with your spouse—it's good and bad," says Potash. "There may be a fight when something goes seriously wrong. But the flip side is that you make love on the conference table after the meeting went really well."

Now *there's* a fringe benefit IBM can't offer. ■

The Greenes' last piece for WORKING WOMAN *was "The Best-Paid Women in Corporate America" (January).*

HEALTH

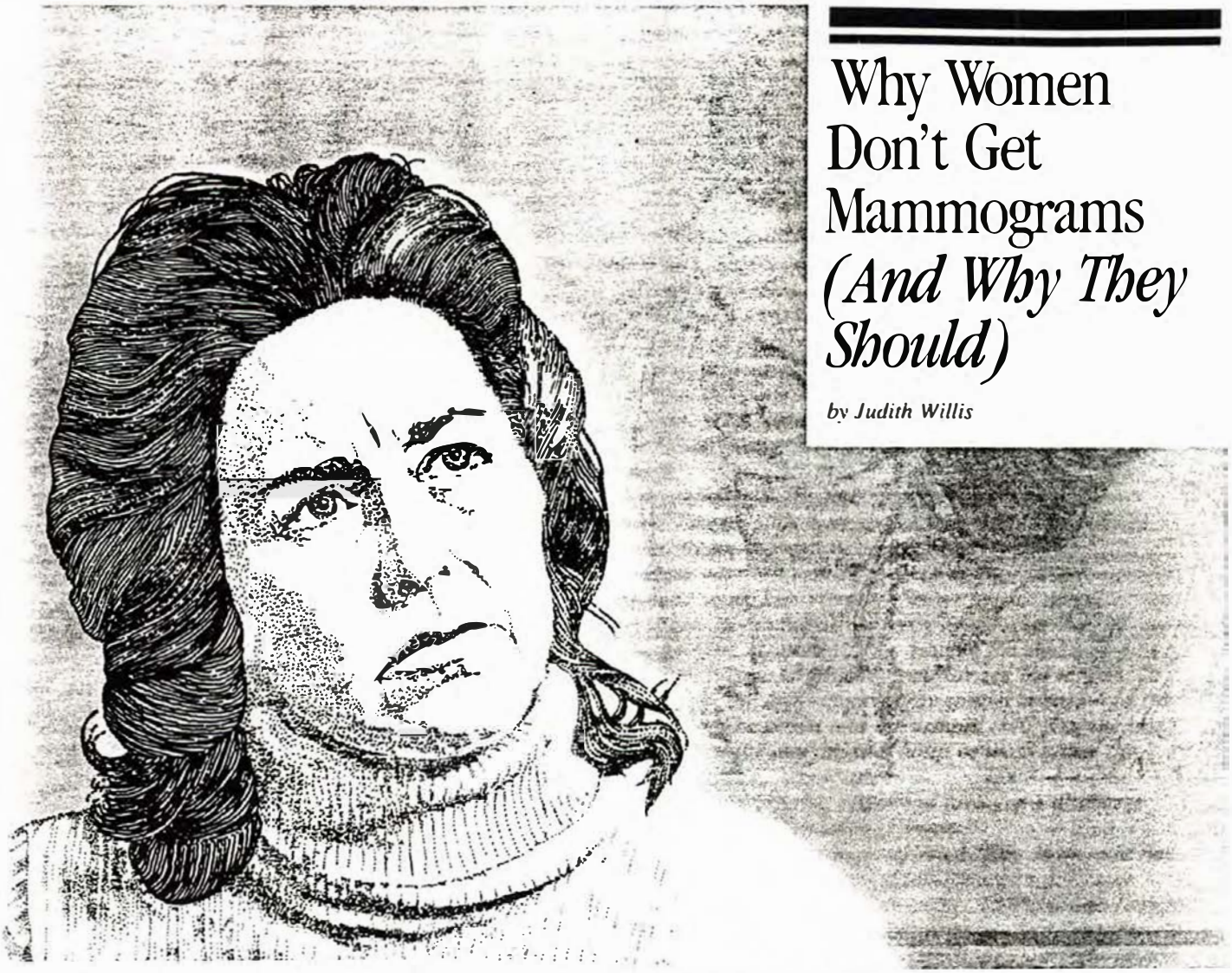
(Continued from page 72) disease by observing their patients have been overpowered. Today, the NIH scientist or the influential university department head carries all the authority. In such a climate, a disease has no reality unless a diagnostic test exists, a pathogen is clearly linked to its transmission—or the patient dies.

In the meantime, the struggle for recognition endures at the grassroots level: patients appealing to legislators whose subcommittees appropriate money to the CDC and the NIH. For each of the last three years, Congress has increased CFS spending by nominal amounts, ranging from $1.1 million to $2.2 million. But money is no longer the only issue, notes health lobbyist Tom Sheridan, who was engaged in 1992 by the Charlotte-based CFIDS Association, the largest patient organization in the country (the acronym represents the name that patient-activists prefer: chronic-fatigue immune-dysfunction syndrome). Money can do many things, but it cannot expunge scientific bias.

"This situation is unique," notes Sheridan, who also represents a number of AIDS-patient organizations. "I have told my [CFS] clients, I can go to Congress and get more money for the NIH and the CDC. But before we ask, let's get an infrastructure in government that is *helpful* to you rather than hurtful."

In order to advance the discovery process, both the CDC and the NIH need to fold into their ranks scientists who come to the issue with open minds and fresh ideas, and to include more patients and patient advocates in their decision-making bodies, as they do with other major diseases. NIH committees reviewing outside scientists' grant proposals need to be composed of true peers—doctors and scientists who either have hands-on experience with CFS or are well-educated in its clinical and laboratory hallmarks. Above all, health officials need to start listening to patients and to those rare observant doctors who have been listening for years. ■

Hillary Johnson is the author of the forthcoming Osler's Web: Inside the Labyrinth of the Chronic Fatigue Syndrome Epidemic.



Why Women Don't Get Mammograms *(And Why They Should)*

by Judith Willis

Why aren't women getting mammograms?

Major health organizations concerned with breast cancer recommend annual mammography screening for women over 50, and less frequent screening for those between 40 and 50. But only an estimated 5 percent to 15 percent of these women are getting this X-ray examination of the breasts. This is particularly disconcerting because about 1 out of 10 women gets breast cancer, and the best chance of detecting the disease early – while it's most curable – is through mammography.

There may be several reasons why women fail to get mammograms. Some may be unaware that mammography screening, in addition to physical examination, is so widely recommended. Others may be reluctant because their insurance doesn't pay for it as a screening procedure. (Most insurers don't cover mammography for screening, but do cover it for diagnosis, once an abnormality is suspected.) Still other women may be afraid that the radiation exposure may actually increase their risk of cancer, or they may have heard that the procedure is painful. Yet, a look at the facts can help reassure women that the benefits of a properly performed mammogram outweigh the risks in groups for which it is recommended.

Mammography is presently the best method of detecting breast cancer at its earliest stages – before the cancer has spread and often before the tumor can be felt. Detecting the cancer this early means a higher chance of survival and perhaps less drastic surgery. One study estimates that for women 50 and older, physical examination alone has reduced deaths due to breast cancer by 18 percent, but physical exam combined with annual mammography screening reduced breast cancer mortality by 56 percent.

A REPRINT FROM FDA CONSUMER MAGAZINE

REPRINTED FROM MAY 1987 (REVISED OCTOBER 1988) **DHHS Publication No. (FDA) 90-1137**

DEPARTMENT OF HEALTH AND HUMAN SERVICES, Public Health Service,
Food and Drug Administration, Office of Public Affairs, 5600 Fishers Lane, Rockville, Md. 20857

☆U.S. GOVERNMENT PRINTING OFFICE 1990-261-206/20012

Mammography is presently the best method of detecting breast cancer at its earliest stages—before the cancer has spread and often before the tumor can be felt.

While it is true that radiation can increase the risk of cancer, radiation exposure from mammography equipment has been reduced by as much as 50 percent in recent years (see "Progress Against Breast Cancer" in the May 1986 *FDA Consumer*). An even more reassuring fact is that the risk of radiation-induced breast cancer is lowest among women who received the exposure later in life, as they would with mammography. (The risk is greatest when the exposure occurs during puberty.) So for older women, for whom the radiation risk is at its lowest and the chance of detecting breast cancer is highest, the benefits of a properly performed mammogram far outweigh the risks.

Yet women should be aware that all mammography facilities may not offer the same quality and skill in performing mammograms. Fortunately, there are some things a woman can know and some questions she can ask to assure herself that she is getting the best and safest medical care in a mammography exam.

It is important that the facility's personnel be properly trained and that up-to-date equipment be used so that the images produced will be of the highest quality. The better the quality of the image, the better the chance of detecting cancer if it is present. High-quality images also help avoid unnecessary levels of radiation exposure.

Here are some questions for a woman to ask about a facility before getting a mammogram:

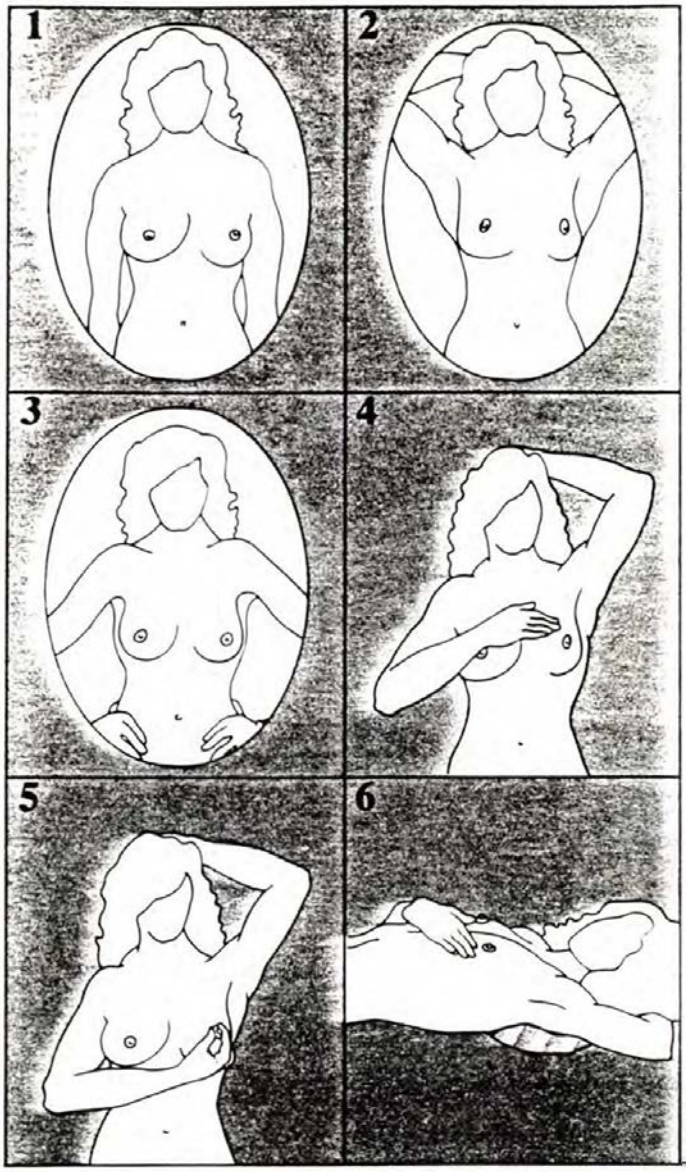
1. Do your X-ray technologist and radiologist have specific training in mammography?

Technologists perform the examination, properly positioning the patient, operating the equipment, and developing the film. Radiologists are medical doctors who interpret the results of the mammogram. Special training is essential for obtaining the best images with the least possible radiation and for interpreting those images properly.

2. Do you use "dedicated" mammography equipment?

In technological parlance, "dedicated" means that the equipment is designed specifically for mammography, providing the best mammograms with as little radiation as possible.

Breast Self-Examination (BSE)



Breast self-examination should be done once a month so you become familiar with the usual appearance and feel of your breasts. Familiarity makes it easier to notice any changes in the breast from one month to another. Early discovery of a change from what is "normal" is the main idea behind BSE. If you menstruate, the best time to do BSE is two or three days after your period ends, when your breasts are least likely to be tender or swollen. If you no longer menstruate, pick a day, such as the first day of the month, to remind yourself it is time to do BSE.

1. Stand before a mirror. Inspect both breasts for anything unusual, such as any discharge from the nipples, puckering, dimpling, or scaling of the skin.

The next two steps are designed to emphasize any change in the shape or contour of your breast. As you do them you should be able to feel your chest muscles tighten.

2. Watching closely in the mirror, clasp hands behind your head and press hands forward.

3. Next, press hands firmly on hips and bow slightly toward your mirror as you pull your shoulders and elbows forward.

Some women do the next part of the exam in the shower. Fingers glide over soapy skin, making it easy to concentrate on the texture underneath.

4. Raise your left arm. Use three or four fingers of your right hand to explore your left breast firmly, carefully, and thoroughly. Beginning at the outer edge, press the flat part of your fingers in small circles, moving the circles slowly around the breast. Gradually work toward the nipple. Be sure to cover the entire breast. Pay special attention to the area between the breast and the armpit, including the armpit itself. Feel for any unusual lump or mass under the skin.

5. Gently squeeze the nipple and look for a discharge. Repeat the exam on your right breast.

6. Steps 4 and 5 should be repeated lying down. Lie flat on your back, left arm over your head and a pillow or folded towel under your left shoulder. This position flattens the breast and makes it easier to examine. Use the same circular motion described earlier. Repeat on your right breast.

3. How often is your mammography equipment inspected and calibrated?

The answer should be at least *annually*. A look at what goes on during a mammography examination may also help allay some fears. The patient usually stands during the procedure, and the technologist positions the breast on a horizontal support above the X-ray film. Compression of the breast by a device called a "paddle" is necessary for proper positioning. The paddle is about the size of this page. With some equipment, the technologist uses the paddle manually to compress the breast. The paddle stays in position while the technologist goes behind an X-ray shield and takes the exposure. After the exposure, the technologist returns to the patient and releases the compression. However, with some newer equipment, the technologist compresses the breast only partially before going behind the machine where there is a button that allows automatic completion of the compression. After the X-ray is taken, the compression is immediately released automatically, minimizing the compression time and the discomfort. The newer compression devices may be more uncomfortable than the older-style compression by a balloon-type device, but the newer devices provide higher quality images.

Although there may be some discomfort while the breast is compressed, there should not be a great deal of pain. Proper compression is crucial to a high-quality mammogram with the lowest radiation dose possible. Women who experience increased tenderness of the breast before their menstrual periods may want to schedule the exam for after their periods.

Experts generally agree that mammography screening should be performed annually after the age of 50, and that it should be done every one to two years between the ages of 40 and 50. Some experts also recommend a single "baseline" mammogram (to which later mammograms can be compared) between ages 35 and 40.

A woman should ask her physician which course is best for her, and should remember that these recommendations apply to *screening* when *no* abnormalities have been found during a physical examination. (A woman in whom an abnormality has been identified or is suspected should follow her physician's advice regarding diagnostic mammographic examination – the age guidelines do not apply here.)

Women should also examine their breasts monthly for any abnormalities. (See diagram.)

Fear or ignorance need not keep a woman from being screened for breast cancer by mammography. With the proper knowledge and caution she can limit the risks and reap the benefits of this relatively painless, quick and safe procedure that could save her life.

Judith Willis is editor of the FDA Drug Bulletin, *a publication for health professionals.*

Current Clips

NATIONAL CANCER INSTITUTE

Office of Cancer Communications

March 11, 1994 10pp

Newsday. March 9, 1994 p. 16

Age Matters

Lawmakers rip mammogram rule

By Ridgely Ochs
STAFF WRITER

Washington — The head of the prestigious National Cancer Institute came under fire yesterday from House members for changing the guidelines on mammography for women ages 40-49.

In a hearing of the House subcommittee that oversees the cancer institute, chairman Edolphus Towns (D-Brooklyn) called the change "wholly irresponsible," said it was based on "biased and flawed" research, and questioned whether the institute had felt pressure to cut costs in light of Clinton's health-care proposal.

Since 1987, the institute, the American Cancer Society and numerous other medical organizations had recommended that all women ages 40-49 get a mammogram every year or every other year. In December, the institute withdrew its recommendation for regular mammograms for women under 50.

Towns said the institute based its decision on European and Canadian studies that included few nonwhite women. Dr. Kathleen Cantwell of the National Medical Association, which represents black doctors, testified that the change was likely to deter younger black women from getting mammograms, even though they have a higher incidence of breast cancer and a much higher mortality rate than young whites.

Members of the Congressional Women's Caucus testifying at the hearing castigated institute director Samuel Broder for the change, saying it was part of a pattern of neglect regarding women's health issues and research.

Rep. Pat Schroeder (D-Colo.) said the guideline change "threw a real wrench into the works," confusing American women about the value of mammography, one of the few reliable options available to catch the cancer early.

The Women's Caucus said it was joining a House resolution introduced by Towns to include mammography screening for women between ages 40-49, as well as older women, in any comprehensive benefits package Congress approves in revamping the nation's health care system.

Broder, however, repeatedly denied that the cost of such coverage played any role in the institute's decision. The decision, he said, was based on the fact that researchers were divided on whether mammography was beneficial for younger women and on the fact that the clinical trials weren't definitive in showing a clear reduction in mortality for that age group.

"We need to stick to the facts as we know them," Broder said. "We need either consensus or clinical trials, but we lack both."

03. 11. 94 11:09 AM P01

Newsday File / Jim Cummins

Changes in mammogram rules for women ages 40-49 have put the National Cancer Institute in the hot seat.

The institute revised its guidelines despite a 14-1 vote by its advisory board to keep them unchanged, and despite continued support for the previous guidelines from a number of medical organizations, including the American Medical Association and the American Cancer Society.

The change was based on an international workshop chaired by Dr. Suzanne Fletcher which reviewed trials on mammography and concluded it produced no significant decrease in mortality for women under 50.

Towns pointed out that Fletcher and others on the workshop panel had publicly said before the meeting that they did not favor mammography for younger women.

Broder denied there was any potential conflict of interest.

workers to support its aging population. It is also a question of national identity—of whether the country can shed its ethnically rooted concepts of German citizenship in favor of a more republican notion. The episode showed the downside of Kohl's imperviousness to, and occasional contempt for, criticism. Kohl would often rather speak to his chronically insecure people in comforting, self-justifying terms than issue moral challenges. "Germany is not an immigration country," he said, parroting a popular but unreal line.

Yet as Kohl and his bridge generation age, and postwar Germans move into power, competing views about German interests and identity will sharpen, possibly resulting in a reconfiguration of the German ideological and party structure with which Kohl is so familiar. The likely next leaders of both the CDU and SPD are further to the right and left, respectively, than the parties' current candidates. Germany is not immune to the trend toward voter discontent and party realignment in all the Western democracies. And since reunification, a new pool of voters—East Germans—has been injected into the system. They have weak attachments to the old parties of West Germany, and many of them are reeling from the changes imposed on them by the end of communism.

Hence the phenomena of the year in German politics: the rise of the former East German Communists as an anti-Bonn *Ossi* protest party; and the decline in support for the FDP, which might not get the 5 percent needed to join parliament and form another coalition with the CDU. If the FDP does collapse, Kohl could be cheated of his expected triumph, and forced to lead a "Grand Coalition" with the SPD. It could even mean a "Red-Green" government of the left, supported, if not joined, by the former East German Communists. Kohl is doing his best to exploit this: but his red scare only seems to underscore his lack of any real fresh ideas.

Germany's political landscape needs a rethinking, of which Kohl is perhaps incapable, and which Kohl's perpetuation in power at least partially postpones. Indeed, German party politics has wilted a bit in the shadow of the *Eichenschrank*. His CDU is united but, in terms of fresh talent, hollow—thanks to Kohl's efforts to marginalize potential rivals. The SPD, Greens and former Communists all share in the crisis of the Western left. The FDP is at the brink of extinction. None is a particularly promising alternative to Kohl. Indeed, Germany could do much worse than four more years of the fat man. But someday it will have to try to do better. •

A chilling case study in health care reform.

HOW SAFE ARE YOUR BREASTS?

By Malcolm Gladwell

At one point during his testimony before a congressional hearing on breast cancer in January, Samuel Broder, director of the National Cancer Institute, was asked what seemed to be a simple question.

"Dr. Broder," said Henry Waxman, the powerful head of the House Health and Environment Subcommittee, "assume you have a patient who is a woman 45 years old.... She has read about breast cancer, is concerned about whether to get a mammogram. In your professional judgment, would you recommend [one]?"

Broder's response takes up two and one-half pages of the hearing transcript. He talks about the "sanctity" of the doctor-patient relationship, about the necessity of explaining the risks of false positive or false negative results, about letting the patient participate in the process. His answer, when it finally emerges, is less an answer than a kind of painful circumlocution.

"I would say, however," Broder said, "that—to keep the discussion focused on what we are saying today—it would be wrong for the health care provider to attempt to induce the woman to obtain mammography or to say the basis of the decision, whatever it might be, is informed by the fact that mammography in that age group has been shown to save lives."

This was an extraordinary moment. For a generation, mammography has been touted as the only effective weapon against breast cancer, the best way to find tumors while they are still small enough to treat. In the mid-1980s the National Cancer Institute itself became the principal cheerleader for mammography, gathering every medical organization in the country together to urge that all women older than age 40 receive regular breast x-rays. Yet here, under oath before a congressional subcommittee, the government's chief cancer expert could not bring himself to say "yes" to what was once the simplest of questions. What happened?

MALCOLM GLADWELL is the New York bureau chief for *The Washington Post.*

The answer is one of the most depressing tales of our new era of cost-conscious medicine. Last December, after two years of study, the NCI decided to revise its guidelines on mammography. The only women who should get regular mammograms, the institute now said, are those aged 50 and over. For women in their 40s—30,000 of whom will be diagnosed with breast cancer this year—the institute's experts declared that they could find no evidence that routine mammography resulted in a "statistically significant reduction in mortality."

This decision was supposed to be based on a hard reading of the available medical data. But in reality it was far from that, and this is where the trouble begins. The NCI looked at only some of the available data; when judgments were made, they were made narrowly and tendentiously. Some studies were misinterpreted, others ignored. It was science gone awry, and the consequences for the health of American women are obvious. Given the NCI's enormous influence on the way medicine is practiced, the new guidelines mean that fewer younger women will ask for mammograms, fewer doctors will order them and fewer insurers will pay for them. It means the war on breast cancer has been scaled back.

But the implications go much further than this. When the NCI announced its decision, it was quickly endorsed by the Clinton administration and hailed by health economists as an example of how medicine can learn to hold the line against unnecessary health care. This is the new age of cost-conscious medicine, when medical advances are celebrated as much for saving money as for saving lives.

There is nothing wrong with this, of course. But the unanswered questions about this new climate of constraint are whether it will compromise the ability of the medical industry to defend the public health, and whether, in the end, the cost-cutters will know the difference between what is necessary and what is unnecessary. The battle over who should receive a breast x-ray was the first real test. The results were not encouraging.

The idea behind mammography is one that informs all cancer prevention: the earlier a tumor is detected and removed, the better the chance of defeating the cancer entirely. Since a lump that can be felt by a woman or her physician can be several years old, radiologists in the 1960s began x-raying breasts, to screen for tumors while they were still in their infancy.

As logical as this strategy sounds, it was by no means clear—at least in the beginning—that it would actually save lives. Screening tests for cancer work only if they are accurate and detect tumors early on. The pap smears given to women for cervical cancer, which are credited with saving thousands of lives, clearly do this. By contrast, the test for prostate cancer has never been shown to save any lives. Why? One hypothesis is that even though prostate exams pick up small tumors, they still don't find them early enough to make a difference in treatment. It is also possible that prostate exams are simply discovering what scientists call "silent tumors," tumors that just sit there and, if left undetected, never would have amounted to anything. If mammograms only picked up cancers that had already inflicted their damage, then, encouraging women to have them would be a waste of money. And if all they did was pick up silent tumors, they would only create a lot of unnecessary worry and needless surgery.

There was a second consideration beyond whether mammography saves lives. What if mammography worked, but only among women of a certain age? This was not an idle concern. As women get older, their breasts become less and less dense. Breast tissue is converted into fatty tissue, which makes it much easier to look for tumors with an x-ray. Older women also get breast cancer far more frequently than younger women; national rates for women in their 60s are as much as three times as high as the rates for women in their 40s. Put simply, mammography might work brilliantly in older women while proving a waste of effort and money for those ten or fifteen years younger.

With these questions in mind, scientists over the past thirty years have conducted studies on the effectiveness of mammography, directly comparing the experience of women who have received regular mammograms with those who have only had their breasts examined by hand during clinical examinations. Eight of these studies have now been completed, and the data show a remarkably consistent pattern: for women over 50, mammography saves lives. For those under 50, the evidence is equivocal. Two of the studies showed that women who were not given mammograms lived slightly longer than those who were, although the difference was too small to be statistically significant. One showed no difference. The remaining five showed an advantage for mammography, but the studies were not large enough to show whether the edge was real or merely the result of chance.

When the National Cancer Institute met in 1987 to formulate a policy on mammography, it recommended annual mammograms for women over 50 and, in deference to the uncertainty for younger women, one- to two-year exams for those in their 40s. Then, in November 1992, the results of the Canadian National Breast Screening Study were published. This was the largest mammography study ever, and the first to specifically address the question of whether mammography worked for women aged 40 to 49. When it also showed no statistically significant advantage for mammography, the cancer institute reconvened its experts. After months of deliberation, the institute revised its guidelines.

It all seemed so straightforward and logical, medicine at its sober and exacting best. It was not.

The first step in understanding where the NCI went wrong in its assessment of mammography lies in an understanding of mammography itself.

A mammogram is, like any x-ray, nothing more than a picture. But it is a picture in which the tiniest detail matters. To detect a break in, say, a thigh bone requires only an x-ray that can distinguish the bone in question from the surrounding muscle and tissue. But for a mammogram to be successful, it needs to create an image that distinguishes among blood vessels, glands, fatty tissue and breast tissue, all while allowing the radiologist to pick out a tumor that may appear on the film no larger than a grain of sand.

Radiologists generally agree that for much of mammography's thirty-year history, neither the x-ray machines themselves nor the techniques used in taking the pictures were up to this task. In the beginning, radiologists used the same x-ray machines for breasts as they did for chests and legs, crude machines with high levels of radiation. Even as late as the end of the 1970s, mammography was still a relatively unsubtle diagnostic technique. It could easily pick out incipient tumors in the relative translucent breasts of older women. Presented with the denser breast of a younger woman, however, it more often than not produced what looked like a badly overexposed negative.

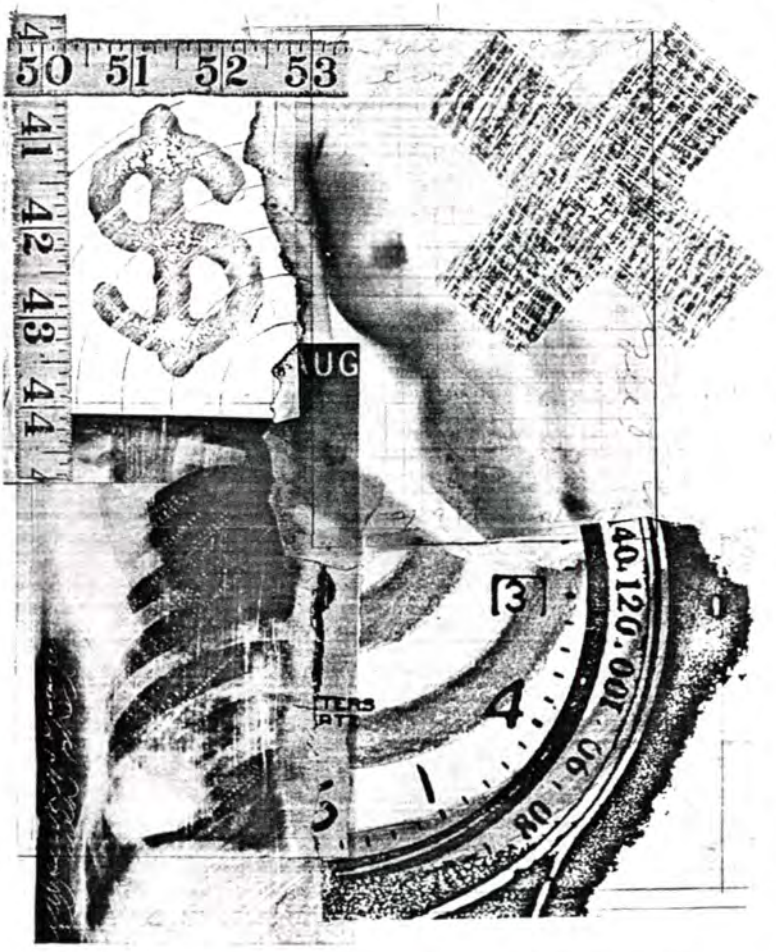


ILLUSTRATION BY TOM PRESTON FOR THE NEW REPUBLIC

During the past fifteen years, however, mammography has undergone a transformation. Machines are now available with x-ray tubes specially designed to create a beam that can penetrate younger breasts. Lead grids are now placed between the breast and the photographic film to soak up radiation "scatter" and deliver a sharper picture. Photographic companies now have special films and processing chemicals that produce a higher contrast, more precise picture.

Radiologists have also recently discovered the answers to several imaging problems. How, for example, should the x-ray camera be positioned? Once it was thought the best results came from taking one picture of the breast from the side. Others thought it made more sense to take a side shot as well as a picture from the top of the breast looking down. Now research has shown that radiologists can find 10 percent more tumors by combining the top shot with what is called a medial lateral oblique view, where the camera is tilted down at the breast at a 45 degree angle.

Or consider a second problem: How often should a woman get a mammogram? In the beginning, radiologists experimented with intervals ranging from a year to two and one-half years. Now, thanks to recent research, the medical community has been able to devise the optimal interval. In women in their 50s, a mammogram can usually detect a tumor anywhere between 3.5 and 4.5 years before it becomes a palpable lump. If an older women gets a mammogram every year-and-a-half or every two years, in other words, the x-ray will probably catch most tumors well before they reach the dangerous stage. In women in their 40s, however, tumors appear to grow much faster, and a mammogram typically provides only an eighteen- to thirty-month early warning. In this group, anything other than an annual mammogram will give cancers a chance to grow and spread before they can be picked up.

These improvements have made a clear difference in the ability of radiologists to find tumors in younger women. Consider the results of a breast cancer screening program conducted in Holland, beginning in 1975. This was classic old-style mammography—one view only (first side, then medial lateral oblique), x-rays once every two years—and with older women it worked well. Generally, radiologists are happy if they can find a tumor before it exceeds ten millimeters in length, and in those women older than 50, almost half the tumors picked up by x-ray were smaller than that. Radiologists also are optimistic about a woman's survival if they can find a tumor while it is still in what is known as stage one—before it has spread. Seventy percent of the older women's tumors were caught at stage one.

Younger women were another story. Using the old

mammography protocol, the Dutch were able to pick up just 19 percent of tumors while they were still under ten millimeters, and only 24 percent that were at stage one, a vastly inferior outcome.

Compare this to the results of two studies conducted over the past three years in San Francisco and British Columbia, both of which used state-of-the-art mammography, two pictures (top and medial lateral oblique) and annual x-rays. Following this protocol, both groups of researchers found no difference at all in the size and stage of tumors found in women under 50 and those over 50.

It should seem obvious, then, that it would make little sense to evaluate how well mammography works on younger women by using data from a time when mammography did not work well on younger women. Yet that is exactly what happened. Of the eight trials used by the National Cancer Institute to reach its revised assessment, one began in 1963, two began in 1976, one began in 1977 and three were started in the early 1980s. Two of the eight took one x-ray per examination. Many did not use the medial lateral oblique position. A majority did not use a lead grid. Two performed mammograms every eighteen months, a third did them every two years, a fourth offered mammography only every twenty-eight months and a fifth every thirty-three months.

Only the eighth and last of these studies—the Canadian trial—was conducted recently using the optimal two views with annual x-rays. But even here the mammography technology was so out-of-date and the training of the technicians so poor that almost half of the mammograms performed during the first four years of the study were later judged to be technically unsatisfactory by outside reviewers. In fact, two radiologists called in to oversee the quality of the program quit in disgust.

It is worth noting, that the experts who decided that mammography did not work in younger women were not, by and large, themselves expert in the use of mammography. They were statisticians, internists and research scientists, and the allegation has been made by radiologists who specialize in breast x-rays that these outside specialists simply did not appreciate fully the technology they were supposed to be evaluating. In the words of one radiologist, the outsiders thought a mammogram was something where "you put a nickel in and you got a Coke out."

There is a second, deeper problem with the NCI's decision, and that has to do with what statisticians refer to as the question of "power." Imagine, for example, that you wanted to find out whether a new kind of headache pill worked better than aspirin. If you gave it to 10,000 people, and for the sake of comparison gave aspirin to another 10,000 people, you could be reasonably sure that the study's results accurately reflected the merits of the two pills. The key is not just the number of people in the study but the sheer number of what are called "events": if, say, the average person averages one headache a month, then a one-year trial would result in your pill being compared to aspirin a whopping 120,000 times.

On the other hand, suppose you were interested only in your pill's effectiveness against a rare kind of migraine, which strikes only one person out of 5,000 every year. In this case, a trial of 10,000 people would provide only two events a year. Even if your pill beat aspirin hands down on both those occasions, you would have no way of knowing whether your pill was actually better or whether its superiority was simply the result of chance. As a general rule, the more events a study has, the greater its power to give an accurate answer.

This is the problem that has plagued researchers trying to gauge the effectiveness of mammography in younger women. In breast cancer trials, researchers follow two groups of women for a number of years, the first of whom are receiving regular mammograms and the second of whom are getting only clinical breast examinations. When there are significantly more events—in this case, deaths from breast cancer—in one group than the other, researchers can then conclude whether, and by how much, mammography works.

In older women, this is not as difficult. They get breast cancer two to three times more often than women under 50, and the disease tends to kill them more quickly. If you combine the result of a number of smaller tests, you can generate a powerful conclusion. But trying to evaluate mammography in women under 50 is a bit like trying to test the imaginary migraine pill. Because younger women get the disease less frequently and live so much longer when they do, most of the time there simply aren't enough events to reach a conclusion. If a study of younger women is going to have any power at all, it is going to have to be huge.

How huge? Harvard University radiologist Daniel Kopans and two colleagues recently calculated that for a study of younger women to accurately measure the same kind of benefit shown in older women it would have to have 480,000 women. But, if you add together all the women under 50 who have ever been part of a mammography trial, the total comes to less than 200,000. It simply hasn't been financially or logistically feasible for anyone to mount a trial of half a million women.

Here is the real explanation for the cancer institute's complaint that no study of younger women to date has shown a "statistically significant reduction in mortality." It does not mean that the benefits of mammography in younger women are too small to measure. It certainly does not mean that mammography doesn't work in younger women (although that is precisely what one senior Clinton administration health official, Devra Lee Davis, claimed in a bizarre editorial in the *Journal of the American Medical Association* earlier

Executive Committee Officers
Patricia Schroeder, *Co-chair*
Olympia Snowe, *Co-chair*
Cardiss Collins, *Secretary*
Marcy Kaptur, *Treasurer*

Congressional Caucus for Women's Issues

Congress of the United States
Washington, D.C. 20515

2471 Rayburn Building
(202) 225-6740

Lesley Primmer
Executive Director

For Release
Tuesday, March 8, 1994

Contact: Susan Wood
(202) 225-6740
Andrea Camp (Schroeder)
(202) 225-4431
Nicholas Graham (Snowe)
(202) 225-6306

NEWS RELEASE

CONGRESSWOMEN CALL FOR BROAD COVERAGE OF SCREENING MAMMOGRAMS IN HEALTH CARE REFORM

Today, the Congressional Caucus for Women's Issues called for broader coverage of screening mammograms in any health care reform package that comes before the Congress. With a number of committees about to take up various health care reform measures, the Caucus made clear that screening mammograms must be part of any basic benefits package.

"With all the conflicting recommendations that are coming from a number of highly respected sources, women need to know that screening mammograms will be covered," said Caucus Co-Chair Patricia Schroeder (D-Co). "Sending women the message that some mammograms may not be covered sends the message that mammograms are not important -- and that's a dangerous message."

"Women under 50 have been left once again in the now familiar situation that the research is not clear and the experts are still debating," Caucus Co-Chair Rep. Olympia Snowe (R-ME) pointed out. "We unfortunately have only mammograms in our arsenal against breast cancer -- which will strike one out of every eight women over a lifetime -- so until we have something more to offer these women, we can't take this away."

The position established by the Caucus calls for complete coverage (no copayments or deductibles) of mammograms for women over 50 every one to two years, where screening mammography has been shown to reduce mortality by 30 percent. Coverage should be provided

Congressional Caucus for Women's Issues

EMBARGOED UNTIL 10:00 AM, TUESDAY, MARCH 8

Caucus Position on Coverage of Screening Mammography in any Basic Benefits Package

- For women **age 50 or older**, where mammography has proven to bring about a 30 percent reduction in mortality:

  * **Plans should provide free screening mammograms, every one to two years.**

- For women **age 40-49**, where mammography has a less certain impact in reducing mortality, but does help in earlier detection:

  * **Plans should cover screening mammograms every two years, with applicable co-pays and deductibles for those reform plans that include cost-sharing.**

  * **All cost-sharing should be based on a sliding scale, so that it does not constitute a barrier to low-income women receiving care.**

- For women of any age:

  * **Plans should provide free screening mammograms for any woman at high risk for breast cancer.**

  * **Plans should cover annual clinical breast exams for women over 40.**

Released March 8, 1994

2471 Rayburn Building, Washington, D.C. 20515 • 202-225-6740

Congressional Caucus for Women's Issues

EMBARGOED UNTIL 10:00 AM, TUESDAY, MARCH 8

Caucus Position on Coverage of Screening Mammography in any Basic Benefits Package

- For women age 50 or older, where mammography has proven to bring about a 30 percent reduction in mortality:

  * **Plans should provide free screening mammograms, every one to two years.**

- For women age 40-49, where mammography has a less certain impact in reducing mortality, but does help in earlier detection:

  * **Plans should cover screening mammograms every two years, with applicable co-pays and deductibles for those reform plans that include cost-sharing.**

  * **All cost-sharing should be based on a sliding scale, so that it does not constitute a barrier to low-income women receiving care.**

- For women of any age:

  * **Plans should provide free screening mammograms for any woman at high risk for breast cancer.**

  * **Plans should cover annual clinical breast exams for women over 40.**

Released March 8, 1994

103D CONGRESS
2D SESSION

H. RES. 368

Expressing the sense of the House of Representatives with respect to the inclusion in any comprehensive benefits package under health care reform of mammography screenings for women under the age of 50.

IN THE HOUSE OF REPRESENTATIVES

FEBRUARY 23, 1994

Mr. TOWNS (for himself, Mrs. COLLINS of Illinois, Ms. VELÁZQUEZ, and Mrs. VUCANOVICH) submitted the following resolution; which was referred jointly to the Committees on Energy and Commerce and Ways and Means

---

RESOLUTION

Expressing the sense of the House of Representatives with respect to the inclusion in any comprehensive benefits package under health care reform of mammography screenings for women under the age of 50.

Whereas breast cancer is a substantial public health problem in the United States, and the incidence and prevalence of this cancer is increasing;

Whereas breast cancer is the number one cause of cancer death in women between the ages of 15 and 54, particularly in African-American and Hispanic women;

Whereas, in 1992 alone, 40,000 cases of breast cancer were diagnosed in women under 50, of which 28,900 cases were diagnosed in women between the ages of 40 and 49;

Whereas the risk of breast cancer increases significantly at age 40;

Whereas, in 1992, the difference between the incidence of breast cancer in women in their forties and women in their fifties amounted to a difference of only 1 percent of all breast cancer cases;

Whereas there has been a 2.5 percent increase in the breast cancer mortality rate of premenopausal African-American women, and experts believe that the increase is significantly influenced by the lack of access to early detection and health care;

Whereas the 5-year survival rates for low-income women with breast cancer are 9 percent lower than the rates for upper-income women;

Whereas the National Cancer Advisory Board of the National Cancer Institute voted 14 to 1 that the Institute's guidelines recommending mammography screenings for women between the ages of 40 and 49 should remain in place because the science is inconclusive to support any changes at this time;

Whereas mammography is among the most effective and widely available methods of early detection;

Whereas the Special Commission on Breast Cancer (formed by the President's Cancer Panel) found that earlier detection increases the likelihood of reducing mortality;

Whereas such Special Commission concluded that the failure of health care providers to recommend mammography is a major cause of poor utilization of mammography screening;

Whereas the Special Commission concluded that health care reform must ensure the removal of financial barriers to access to early detection and screening; and

Whereas the Special Commission recommended that further studies be conducted to determine the mortality benefit of screenings in women between the ages of 40 and 49: Now, therefore, be it

Resolved, That it is the sense of the House of Representatives that—

(1) any comprehensive benefits package under health care reform should include mammography screenings—

(A) for women between the ages of 40 and 49 (as well as for women over the age of 49); and

(B) for women under the age of 40 who are at high risk for breast cancer or whose physician determines that it is medically necessary or appropriate;

(2) a copayment should be imposed for mammography screening under such benefits package, except for low-income families; and

(3) although it is currently prudent for such package to include mammography screenings for women between the ages of 40 and 49, the Public Health Service, in conjunction with national and international cancer research centers, should imme-

diately undertake a randomized clinical trial to determine the effectiveness and benefits of mammography and other emerging screening technologies for such women.

O

Clinton Presidential Records
Digital Records Marker

This is not a presidential record. This is used as an administrative marker by the William J. Clinton Presidential Library Staff.

This marker identifies the place of an oversized document(s). Given our digitization capabilities, we are sometimes unable to adequately scan such large documents. The title from the original document is indicated below.

Oversized documents have not been scanned in their entirety for the purpose of digitization. To see the documents please contact or visit the Clinton Presidential Library's Research Room.

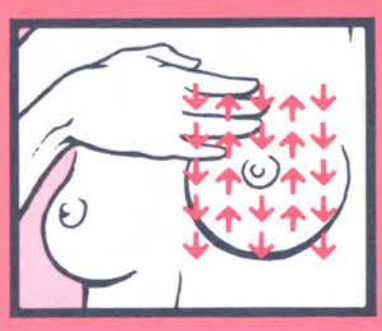
Document Title: **HOW TO DO THE BREAST SELF-EXAM**
AMERICAN CANCER SOCIETY

HOW TO DO THE BREAST SELF-EXAM

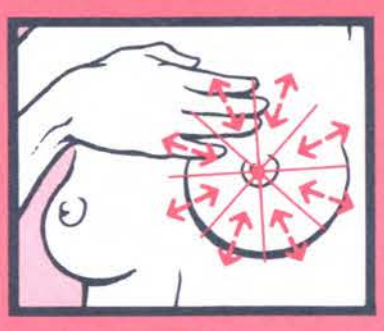
1 Put a pillow under your right shoulder and place your right arm behind your head.

2 Using the top third of your three left middle fingers, press firmly to feel for lumps in your right breast.

3 Move your fingers around the breast in either the Circle (a), the Up and Down Line (b), or the Wedge (c) method. Make sure you cover the entire breast area.



B



C

4 Place the pillow under your left shoulder. Now examine your left breast using the same method with your three right middle fingers.

The American Cancer Society

REVIEW

Statutory Requirements for Disclosure of Breast Cancer Treatment Alternatives

*Susan G. Nayfield, Gregory C. Bongiovanni, Marianne Haenlein Alciati, Ruth A. Fischer, Lawrence Bergner**

Therapeutic options for breast cancer, particularly for early-stage disease, and increased patient participation in medical decision-making have oriented state legislatures toward ensuring that women with breast cancer have adequate information about treatment alternatives. Currently, 18 states have enacted statutes regarding physician disclosure of treatment alternatives to breast cancer patients. This paper reviews these statutes in the context of the requirements imposed on the physician as health care provider and the content of medical information presented to the patient as a consequence of the laws. State statutes were identified through the National Cancer Institute's State Cancer Legislative Database, and the statutory requirements were analyzed. For statutes requiring development of a written summary of treatment alternatives, the most recent summary was obtained through the responsible state agency, and informational content was analyzed for relevance to treatment decisions in early-stage disease. As a group, these laws address informed consent for treatment, physician behavior within the patient–physician relationship, and the medical information upon which treatment decisions are based. Individual statutes vary in the scope of the issues addressed, particularly in the responsibility placed on physicians, and treatment option summaries developed in response to this legislation vary widely in content and scope. Despite broad implications of these statutes in oncology practice, little is known about their effects on breast cancer care. Additional research is needed to define the impact of these statutes on breast cancer care, as such legislation is considered by other states for this and other diseases. [J Natl Cancer Inst 86:1202-1208, 1994]

Recent advances in the treatment of breast cancer have provided realistic therapeutic options for breast cancer patients, particularly for those with early-stage disease (*1*). This opportunity for choice, in an age of increased patient participation in medical decision-making, has oriented state legislatures toward ensuring that women have adequate information about breast cancer treatment alternatives. To date, 18 states have enacted legislation regarding physician disclosure of treatment alternatives to breast cancer patients. These laws are of particular interest to medical practice because they prescriptively mandate physician behavior within the patient–physician relationship and potentially define the medical information upon which treatment decisions are based.

This paper reviews the requirements imposed upon the physician in his/her role as health care provider and summarizes the specific medical information presented to the patient as a consequence of the laws. In addition, the implications of the statutes for patterns and processes of cancer care are discussed and the need for research on the impact of these laws on oncology practice is highlighted.

The State Cancer Legislative Database, a computer database of state cancer legislation established under the Public Health Agency Section of the National Cancer Institute (NCI), was used to identify states with breast cancer disclosure statutes. For statutes requiring development of a written summary of treatment alternatives, the most recent summary was obtained from the responsible state agency for content analysis used for this review.

Statutes for Physician Disclosure of Breast Cancer Treatment Options

The earliest state statute pertaining to informed consent and/or disclosure of treatment alternatives to breast cancer patients was enacted in 1979, and the most recent was enacted in 1989 (Table 1); 10 of the statutes have been amended subsequent to enactment.

**Affiliations of authors:* S. G. Nayfield (Community Oncology and Rehabilitation Branch), M. H. Alciati (Public Health Applications Research Branch), Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, Md.

G. C. Bongiovanni, Vanderbilt University School of Law, Nashville, Tenn.

R. A. Fischer, The MayaTech Corporation, Silver Spring, Md.

L. Bergner, National Opinion Research Center, Washington, D.C.

Correspondence to: Susan G. Nayfield, M.D., M.Sc., National Institutes of Health, Executive Plaza North, Rm. 300, Bethesda, MD 20892.

See "Notes" section following "References."

Table 1. State legislation on disclosure of breast cancer treatment options

State	Year of enactment	Statutory citation
California	1980*	Cal Health & Safety Code §1704.5
Florida	1984*	Fla Stat §458.324
Georgia	1984*	Ga Code Ann §43-34-21
Hawaii	1983	Haw Rev Stat §671-3
Illinois	1989*	Ill Rev Stat ch 127, §55.49
Kansas	1984	Kan Stat Ann §65-2836
Kentucky	1984	Ky Rev Stat Ann §311.935
Maine	1989	Me Rev Stat Ann tit 24, §2905-A
Maryland	1986	Md Health-Gen Code Ann §20-113
Massachusetts	1979	Mass Gen L ch 111, §70E
Michigan	1986*	Mich Comp Laws §333.17013
Minnesota	1982	Minn Stat §144.651
New Jersey	1984*	NJ Stat Ann §45.9-22.2
New Mexico	1984†	NM Ilj Mem 4, 36th Leg, 2d Sess
New York	1985*	NY Pub Health Law §2404
Pennsylvania	1984*	Pa Cons Stat §5641
Texas	1989	Tex Health & Safety Code Ann §86.002
Virginia	1984*	Va Code Ann §54.1-2971

*Amended subsequent to year of enactment.
†By resolution rather than by statute.

The Physician's Role

While all statutes address the participation of the breast cancer patient in making decisions about her treatment, they differ in the manner and extent to which they direct the physician's conduct in communicating information (Table 2). Two statutes (those of Pennsylvania and Virginia) require only that the patient execute a preoperative consent form authorizing a breast biopsy and/or such additional "operations or procedures" as may be "necessary." The statutorily defined Pennsylvania consent form notably includes the following statement: "I have been informed of the current medically acceptable alternatives to radical mastectomy."

Minnesota and Massachusetts laws, codified within their states' statutory bill of patients' rights, indicate the patient's right to information on breast cancer treatment alternatives but do not specify a format for the physician's response to that right. The Minnesota law holds the physician responsible for providing "full information" on "all alternative methods of treatment of which [he/she] is knowledgeable . . . and the risks associated with each" but also acknowledges the individual's right to refuse this information. The Massachusetts law broadly states the breast cancer patient's entitlement to "complete information" but does not direct the physician more specifically.

While 14 statutes require the development and publication of a standardized summary on medically acceptable treatment alternatives for breast cancer, the laws vary in the physician's obligation to utilize these brochures. Four of the 14 states place no explicit duty on the physician to distribute the brochure: The Illinois statute states only that the physician "shall make the summaries available to the public"; the New Mexico resolution and Georgia statute "urge," but do not require, physicians to use the summary; and the Texas statute indicates that "a physician may distribute the summary to a patient when the physician determines . . . that it is in the best interest of the patient."

However, 10 of the states require the physician to give the written summary to his/her patient with breast cancer. Five states (California, Kansas, Kentucky, New Jersey, and New York) unequivocally direct the physician to provide the brochure to his/her patients. Three other states (Maryland,

Table 2. Statutory requirements for physician disclosure of breast cancer treatment options

State	Pre-operative consent form only	Patients' bill of rights to alternative treatment information	Treatment alternatives summary						Documentation		
			No explicit duty to provide	Direct physican to provide	Impose affirmative duty to educate	Physician must express willingness to discuss	Provide both in written form and orally	Procedures by Board of Medical Examiners	Patient chart	Patient signature	Professional discipline
California				•					•		•
Florida					•	•			•		
Georgia			•								
Hawaii								•			
Illinois			•								
Kansas				•						•	•
Kentucky				•							
Maine							•			•	•
Maryland					•					•	•
Massachusetts		•									•
Michigan					•					•	•
Minnesota		•									
New Jersey				•						•	
New Mexico			•								
New York				•							
Pennsylvania	•										•
Texas			•								
Virginia	•										

Michigan, and Florida) impose an affirmative duty on the physician to educate his/her patients about treatment alternatives. In Maryland and Michigan, this requirement may be satisfied by the standardized brochure. Florida law places an additional legal duty upon the physician to "express his willingness to discuss [the brochure] with the patient," while Maine law mandates that the physician provide the patient with information on treatment alternatives "both orally and in writing." The Hawaii statute defers the decision on what information the physician must give and how it must be given to the State Board of Medical Examiners.

Some statutes require documentation of physician compliance with the law by an annotation by the physician in the patient's chart (California and Florida) or by the patient's signature acknowledging that she has received the required information (Kansas, Maine, Maryland, Michigan, and New Jersey). Furthermore, various statutes make provision for professional discipline or note the availability of civil suit in the event of physician noncompliance (California, Kansas, Maine, Maryland, Massachusetts, Michigan, and Pennsylvania).

The Standardized Written Summaries

In 12 of the 14 states, as required by statute, a standardized written summary has been developed and printed in brochure or booklet form[1,2] (Table 3). All 12 summaries carefully note that their function is to provide information and understanding to supplement, not replace, the patient's communication with her physician. However, the summaries vary in the manner in which they encourage the patient's participation in medical decision-making. Some summaries present medical facts with a minimum of exhortation, while others sound a strong and direct note: "Because there are alternative approaches, and choices to be made among these, you must be well-informed and involved in the decision-making. It is your body and you have the most to gain or lose by how it is treated" (Hawaii's brochure).

As required by statute, each state summary reviews standard medical and surgical approaches to breast cancer therapy (Table 4). All 12 summaries focus on a general discussion of acceptable treatments for breast cancer (surgery, radiation therapy, chemotherapy, and hormone therapy). While most brochures state or imply that treatment depends on stage of disease at diagnosis, few explain that decisions among medically equivalent or similar treatments generally apply only to patients with disease limited to the breast and axillary lymph nodes. As described in the following sections, information in the brochures varies in relevance to treatment decisions facing women with early-stage breast cancer.

Breast-conserving surgery versus mastectomy. Treatment approaches to operable primary breast cancer have changed dramatically since the late 1970s, when modified radical mastectomy replaced more extensive surgery as standard therapy and two-step procedures (diagnostic biopsy separate from definitive surgery) became recommended practice (*2,3*). More recently, breast-conserving surgery with postoperative radiotherapy has been recognized as equivalent in medical outcome to that achieved with mastectomy for early-stage disease (*4-9*).

State summaries vary in their information content regarding different surgical approaches to primary treatment. Nine of the 12 mention the option of two-step procedures and comment that this approach is preferred over one-step procedures (Table 4). Illinois' summary cautions, "If you choose the one-step procedure, you will need to make alot [*sic*] of decisions about treatment before you know for sure that you have breast cancer."

Most summaries describe primary surgical procedures separately, discussing potential advantages and disadvantages of radical mastectomy, modified radical mastectomy, total (simple) mastectomy, and partial mastectomy or lumpectomy. At the very least, each of the summaries describes differences between mastectomy and breast-conserving procedures. Most summaries mention axillary lymph node dissection in association with specific surgical procedures, and several explain that information from the axillary lymph nodes will contribute to decisions about further therapy.

In general, the summaries do not discuss clinical research findings or recommend one form of treatment over another. It is notable, however, that five of the state summaries (those from California, Kansas, Maryland, New York, and Texas) do include the research findings of medical equivalence between lumpectomy followed by radiation therapy and mastectomy for early-stage breast cancer. A typical example (New York's brochure)

Table 3. State summaries regarding disclosure of breast cancer treatment options

State	Most recent edition	Title of summary; source
California	1991	*Breast Cancer Treatment: Summary of Effective Methods, Risks, Advantages, Disadvantages;* California Department of Health Services, Sacramento
Florida	1984	*Breast Lumps: A Guide to Understanding Breast Problems and Breast Surgery;* Florida Department of Health and Rehabilitative Services, Tallahassee (Spanish version available)
Georgia	1984	*About Breast Cancer;* Georgia Composite State Board of Medical Examiners, Atlanta
Hawaii	1988	*Breast Cancer Treatment Alternatives: A Woman's Choice;* Hawaii Medical Association, Honolulu
Illinois	1993	*Breast Cancer: Your Right to Know;* Illinois Department of Public Health, Springfield (Spanish version available)
Kansas	1992	*Facts about Breast Disease and Your Options;* Kansas Board of Healing Arts, Topeka
Kentucky	1991	*Choices: Treatments for Breast Cancer;* Commonwealth of Kentucky Cabinet for Human Resources, Frankfort
Maryland	1989	*If You Have Breast Cancer: Treatments That Are Available;* Maryland Department of Health and Mental Hygiene, Baltimore
Michigan	1990	*Breast Cancer: What You Should Know Before Treatment;* Michigan Department of Licensing and Regulation, Lansing
New Mexico	1984	*Women and Breast Cancer: A Guide to Treatment Decisions;* State of New Mexico Health and Environment Department, Albuquerque
New York	1988	*Breast Cancer;* New York State Health Department, Albany (Spanish version available)
Texas	1991	*Breast Cancer Treatment: What You Need To Know;* Texas Department of Health, Austin

Table 4. Informational content of standard summaries developed in response to state statutes

States requiring development of a summary	Summary developed	Illustration included	Breast-conserving surgery versus mastectomy			Chemotherapy and hormone therapy			Breast reconstruction	Clinical trials
			Discuss two-step surgical procedure	Include research findings of medical equivalence*	Indicate some patients may not be candidates	Mention chemotherapy and hormone therapy	Discuss chemotherapy in adjuvant setting	Discuss hormone therapy in adjuvant setting		
California	•		•	•	•	•	•	•	•	•
Florida	•	•	•			•	•		•	
Georgia	•	•	•		•	•			•	
Hawaii	•	•				•	•		•	•
Illinois	•		•			•			•	
Kansas	•	•	•	•		•			•	
Kentucky	•	•	•			•	•	•	•	•
Maine	While the Maine statute requires that written information be provided to the patient, it permits use of brochures from recognized professional organizations (National Cancer Institute, American Cancer Society, American College of Surgeons) in lieu of a newly developed summary.									
Maryland	•	•	•	•	•	•	•	•	•	•
Michigan	•	•	•			•	•	•	•	•
New Jersey	The New Jersey mandate for development of an informational booklet was enacted as a 1993 amendment to the original statute. The informational booklet is not yet available.									
New Mexico	•	•			•	•	•		•	
New York	•	•	•	•	•	•	•		•	•
Texas	•	•		•		•			•	•

*Equivalence between lumpectomy followed by radiation therapy and mastectomy for early-stage breast cancer.

is as follows: "Several studies show that, for relatively small breast cancers, lumpectomy or segmental mastectomy followed by radiation therapy may result in the same cure rate as removal of the breast." However, information in two summaries fails to reflect current knowledge regarding breast-conserving treatment. Georgia's summary, developed in 1984, describes breast-conserving surgery as "a new procedure for breast cancer which is under investigation to see if it is as safe and effective as the modified radical mastectomy." New Mexico's brochure, also originating in 1984, cautions that "these procedures for breast cancer are relatively new and the results are currently being studied" and concludes that "current medical practice favors the modified radical mastectomy."

In most instances, the summaries relate that treatment depends on the size, location, and type of breast cancer. However, only five summaries (those from California, Georgia, Maryland, New Mexico, and New York) explain specifically that breast conservation may not be an option for all women.

Adjuvant therapy. Since the late 1970s, adjuvant chemotherapy or hormone therapy has become standard practice for the treatment of breast cancer patients with positive axillary lymph nodes (*10,11*). Although more recent research suggests that adjuvant therapy may have an impact on treatment outcomes for lymph node-negative patients (*12*), the 1990 National Institutes of Health (NIH) Consensus Development Conference (*9*) recommended that, for these patients, "the decision to use adjuvant treatment should come after a thorough discussion with the patient about the likely risk of recurrence without adjuvant therapy, the expected reduction in risk with adjuvant therapy, toxic effects of therapy, and the impact of therapy on quality of life."

All summaries mention chemotherapy and hormone therapy as treatments used for breast cancer, and most include a brief description of common (acute) side effects. Many summaries discuss the concept of adjuvant therapy. For example, Maryland's pamphlet explains, "Additional breast cancer treatments may follow surgery; this additional treatment is called adjuvant therapy. Adjuvant therapy may be prescribed for patients to boost the effects of their surgical treatment by destroying any cancer cells that remain in the body. Additional treatments are only given to those women who may be helped by them. Whether or not a woman can be helped by adjuvant therapy will depend on a number of factors. . . ." California's brochure describes adjuvant therapy as "supplement[ing] the initial surgical or radiation treatment when it is likely or possible that the cancer will recur" and explains that "adjuvant therapy may increase the effectiveness of surgery or radiation therapy and reduce or delay the recurrence of breast cancer." However, four of 12 summaries do not discuss the use of chemotherapy in the adjuvant setting, and eight do not mention adjuvant hormone therapy.

Breast reconstruction. For many women who have had mastectomy for early-stage breast cancer, reconstructive surgery is an important part of the recovery and adjustment process. While all summaries mention breast reconstruction, they vary in the degree of detail in the discussion. For example, California's pamphlet includes information on external and internal prostheses and on procedures for reconstructive surgery, whereas Hawaii's pamphlet states simply that "reconstruction of the breast at a later date may be desired."

Participation in clinical research. The breakthroughs in breast cancer care that have led to treatment options have been made through patient participation in clinical trials, and con-

tinued clinical research is necessary to ensure future progress. The summaries vary in discussion of clinical research. Seven of the summaries (those from California, Hawaii, Kentucky, Maryland, Michigan, New York, and Texas) inform their reader that clinical trials (designated variously as "research studies" or "investigative treatments") are under way and suggest that she inquire into them.

Discussion

The enactment of statutes on disclosure of breast cancer treatment options did not come without controversy. Massachusetts passed the first piece of legislation in 1979, which was a broad statement of a patient's right to "complete information" about her treatment options. The state's Medical Society launched an attempt to repeal the statute as soon as the legislature reconvened, citing the law's undue interference in the patient–physician relationship and arguing that all treatment alternatives are not alike (*13*). Elsewhere, physicians raised concerns about the reasonableness of passing a law with regard to one organ and one disease (*14*) and expressed doubts about the possibility of producing a brochure or booklet that could appropriately serve the needs of all patients (*15*). The development of a standardized summary that satisfied everyone proved to be a difficult task in some cases; California's brochure was not completed and mailed to physicians until 2 years after the effective date of the law (*16*). Some states, such as Ohio, have considered disclosure legislation but have never enacted it (*17*). Federal legislation requiring the development and dissemination of a standardized summary on treatment options in states receiving federal health care funding has been introduced before the U.S. Congress as recently as 1991 (*18*); however, it has not been enacted.

For all the controversy, reports and surveys imply a positive reception by physicians once breast cancer disclosure legislation has been implemented and "lived with" (*19,20*). While the statutes have not been a significant source of malpractice litigation (*21*), it has been suggested that they set a precedent for further (and more problematic) legislative incursions into the patient–physician relationship (*22*). Despite this concern, systematic assessment of the consequences of these laws on physician practices and patient care has not been undertaken. Specific questions arise as to the effects of the statutes on process and outcomes of care for breast cancer patients in the context of current knowledge about patient–physician interaction and medical decision-making.

1) What is the impact of this legislation on physician attitudes toward and practices in obtaining informed consent?

Considerable skepticism exists about how well the goals of informed consent are realized in practice (*23-26*). Evidence suggests that physicians do not necessarily view the discussion of treatment alternatives as an integral component of informed consent. In one study observing clinicians' communication with patients about a proposed procedure, alternative treatments were discussed less often than were the rationale for and benefits and risks of the proposed procedure (*27*). In a survey in which physicians were asked, "What does informed consent mean to you?", only 14% of respondents mentioned treatment alternatives, and only 9% linked informed consent to the patient making a choice or stating a preference (*28*). These attitudes may be particularly problematic in the treatment of early-stage breast cancer, where options are available for many patients, and it remains to be determined whether legislation requiring disclosure of treatment alternatives can influence these attitudes or practices.

The manner in which a physician presents breast cancer treatment options may also influence a patient's ability to make a truly informed choice. This is of concern because a physician's preference for specific treatment options may not be based on scientific evidence of efficacy (*29-31*). Twenty-two percent of Colorado surgeons surveyed in 1991 stated their disbelief that breast-conserving surgery with radiotherapy and modified radical mastectomy have equivalent medical outcomes (*32*), despite published results of large clinical trials to this effect (*4-6*). While the remaining 78% of surgeons in this study acknowledged equivalence of medical outcomes with breast-conserving treatment and mastectomy, 44% of these surgeons biased their presentation to the patient by identifying mastectomy as the "gold standard" and/or by stating that cosmesis after radiotherapy is not good. Whether objective information in the summaries can modulate the effects of physician bias and thus enhance the informed consent process requires further study.

For the states in which statutes do not require individual physician action (and for those that require action but have no provision for enforcement), legislation regarding disclosure of treatment options may represent little more than guidelines for physician behavior. In these situations, the statutes may be regarded as the decision-making counterpart to the clinical practice (treatment) guidelines for breast cancer management that have existed since the late 1970s (*3*). Studies (*33,34*) disappointingly suggest that these guidelines have limited effect on the adoption of new treatment technologies. This finding emphasizes the importance of evaluating the impact of legislation on disclosure of treatment options on physician behavior and patterns of breast cancer care.

2) What is the impact of these statutes on the actual decision-making process?

A growing body of literature on the medical decision-making process in cancer care has emerged during the last decade. Interview studies (*35,36*) of women deciding about breast cancer treatment have identified empirical indicators of decision behavior that can categorize style of decision-making. "Deferrers," approximately 40% of patients, accept the treatment option recommended by the physician with little hesitation or reservation. For these women, information about technical aspects or risks of therapy is often viewed as threatening and is avoided whenever possible. "Delayers," about 45% of patients, vacillate among options until they can recognize a difference among the possible choices. The anxiety and conflict involved in the decision-making process for these women make them particularly susceptible to influence by family or physician, and they can often be persuaded to accept a course of treatment just to end the decision conflict. "Deliberators," 15% of patients, view complete information on alternatives as an absolute necessity for decision-making and become distressed when accurate information cannot be obtained from reliable sources. The marked differences in informational needs among these decision-making

types are consistent with observations from other research (*37,38*) and raise the possibility that legislation on disclosure of treatment options may complicate the decision-making process, rather than enhance it, by imposing a decision-making style that may be inappropriate for a majority of breast cancer patients.

Other research (*37,38*) has shown that, while most women prefer active participation in initial decision-making regarding surgery, as many as half may require "shared" decision-making with a spouse or friend closely involved in the process. Important sources of information consistently identified by women in the latter study (*38*) were predominantly "people" sources (physician by 82% of patients, family by 41%, and friends by 36%). Clinic handouts were rated among the three most important sources by only 36% of patients, clinic videotapes by 18%, and scientific journals by 14% (*38*). These findings suggest that providing information in the format of standard summaries may not be consistent with the preferences or needs of a majority of patients.

3) What is the impact on the patient, in terms of psychosocial outcomes?

A statement issued by the NIH Consensus Development Conference (*9*) on the treatment of early-stage breast cancer advises that breast-conserving treatment is preferable to mastectomy because it preserves the breast but cautions that important considerations in selecting therapy include cosmetic results, psychosocial issues, and patient preferences for treatment methods. A prospective study (*39*) of participants in a randomized clinical trial comparing breast-conserving surgery to mastectomy found greater anxiety, more sexual problems, and greater dissatisfaction with body image among mastectomy patients than among patients randomly assigned to receive breast-conserving treatment. However, few differences in psychosocial adaptation were identified between groups of patients who, when appropriate, were given a *choice* of therapy (*40-42*). These findings suggest that a breast cancer patient's opportunity to make her own choice among treatment alternatives, *if she wishes to do so*, may play an important role in her healthy adjustment to the disease. Neither breast-conserving surgery nor mastectomy necessarily results in superior psychosocial outcomes; what seems to matter is the individual's ability to choose between them freely. In this context, statutes on breast cancer treatment disclosure may facilitate identification of clinical options, discussion of associated risks and benefits, and choice among treatment alternatives. However, it remains to be tested whether these statutes have an impact on psychological outcomes of treatment by ensuring choice among treatment options.

4) What is the impact of legislation for disclosure of treatment options on patterns of care for breast cancer patients?

Geographic variations in breast cancer care, particularly in the adoption of breast-conserving therapy for early-stage disease, have received much attention lately. Data from the Surveillance, Epidemiology, and End Results (SEER) Program[3] of the NCI demonstrated marked variability in use of breast-conserving procedures among patients with early-stage breast cancer for different SEER registry areas; utilization rates ranged from 19.6% in Iowa to 41.5% in Seattle for cases diagnosed in 1985-1986 (*43*). Regional variations in rates of breast-conserving surgery were also observed for Medicare recipients with early-stage breast cancer diagnosed in 1986; rates were highest in the Middle Atlantic region (20.0%) and lowest in the West South Central (7.3%) and East South Central (5.9%) regions (*44*). The possible contribution of disclosure statutes to geographic variations in breast-conserving surgery has yet to be investigated. The influence of disclosure statutes on postsurgical care, e.g., adjuvant therapy and breast reconstruction, is another area for health services research.

Studies of the impact of disclosure statutes on patterns of care must also consider appropriateness of care. For example, the outcome of breast-conserving surgery is comparable to that of total mastectomy only when it is followed by radiotherapy. However, the use of radiotherapy after breast-conserving surgery varies widely from one geographic area to another, ranging from 80.7% in Seattle to 60.3% in Iowa for 1985-1986 (*43*). Additional areas for future investigation are 1) the relationship between informed decision-making and appropriate medical care and 2) the potential impact of pamphlet content on appropriateness of care.

A final consideration is the mechanisms by which statutes for disclosure of treatment options could influence patterns and outcomes of breast cancer care. As discussed above, the statutes potentially improve physician performance in obtaining informed consent and empower patients to become active participants in the decision-making process. However, it is possible that the statutes have a positive effect on breast cancer care through indirect influence, i.e., that the public and professional attention (and often controversy) surrounding enactment of the statutes focuses on the medical decision-making process and heightens awareness of advances in breast cancer treatment, irrespective of physician compliance with statutory requirements.

Conclusions

While breast cancer disclosure statutes vary in the requirements placed upon the treating physician, they share the common goal of enhancing patient participation in the medical decision-making process when there are reasonable alternatives for effective treatment. These statutes represent an unusual policy response in that they address the patient–physician relationship in the context of a particular disease and an untested public health intervention in that formal evaluation of their impact on the process and outcomes of breast cancer care has not been undertaken.

These considerations are especially relevant because state legislatures have recently begun to consider similar disclosure laws with regard to prostate cancer. In 1990, California enacted a law requiring the development of a standardized written summary on treatment alternatives for this disease (*45*); a 1993 amendment requires posting a notice to inform patients that the summary is to be provided (*46*). The Massachusetts general legislature entertained a similar bill for prostate cancer during its 1993 session (*47*). Research on the effects of this legislative approach is of primary importance in evaluating the potential for statutes on disclosure of treatment options to change the health care process and to improve both psychosocial and medical outcomes of cancer care.

References

(1) Sanders CM: Current management of breast cancer. Br J Hosp Med 50:588-590, 592-593, 1993

(2) Vana J, Bedwani R, Mettlin C, et al: Trends in diagnosis and management of breast cancer in the U.S.: from a survey of the American College of Surgeons. Cancer 48:1043-1052, 1981

(3) National Institutes of Health: National Institutes of Health Consensus Development Conference Statement. The Treatment of Primary Breast Cancer: Management of Local Disease. Bethesda, Md: NIH, June 5, 1979

(4) Veronesi U, Saccozzi R, Del Vecchio M, et al: Comparing radical mastectomy with quadrantectomy, axillary dissection, and radiotherapy in patients with small cancers of the breast. N Engl J Med 305:6-11, 1981

(5) Sarrazin D, Le M, Rouesse J, et al: Conservative treatment versus mastectomy in breast cancer tumors with macroscopic diameter of 20 millimeters or less. The experience of the Institut Gustave-Roussy. Cancer 53:1209-1213, 1984

(6) Fisher B, Bauer M, Margolese R, et al: Five-year results of a randomized clinical trial comparing total mastectomy and segmental mastectomy with or without radiation in the treatment of breast cancer [Medline comment: scientific misconduct—data to be reanalyzed]. N Engl J Med 312:665-673, 1985

(7) Harris JR, Hellman S, Kinne DW: Limited surgery and radiotherapy for early breast cancer. N Engl J Med 313:1365-1368, 1985

(8) Harris JR, Hellman S, Canellos GP, et al: Cancer of the breast. *In* Cancer: Principles and Practice of Oncology, 2nd ed (DeVita VT Jr, Hellman S, Rosenberg SA, eds). Philadelphia: Lippincott, 1985, pp 1119-1177

(9) National Institutes of Health: National Institutes of Health Consensus Development Conference Statement. Treatment of Early Stage Breast Cancer. Bethesda, Md: NIH, June 18-21, 1990

(10) National Institutes of Health: National Institutes of Health Consensus Development Conference Statement. Adjuvant Chemotherapy of Breast Cancer. Bethesda, Md: NIH, July 14-16, 1980

(11) National Institutes of Health: National Institutes of Health Consensus Development Conference Statement. Adjuvant Chemotherapy for Breast Cancer. Bethesda, Md: NIH, September 9-10, 1985

(12) National Cancer Institute: Clinical Alert From the National Cancer Institute. Bethesda, Md: NCI, May 18, 1988

(13) Annas GJ: Breast cancer: the treatment of choice. Hastings Cent Rep 10:27-29, 1980

(14) Barton S: The subject was breast cancer and informed consent. Va Med 109:744-747, 1982

(15) Carter SK: The California breast cancer law and government-mandated patient education. CA Cancer J Clin 32:173-176, 1982

(16) Lewis HE: Informed consent for breast cancer patients. Iowa Med 75:59-63, 1985

(17) Porter S: The breast cancer controversy—information by legislation. Ohio State Med J 81:642-645, 1985

(18) HR 382, 102d Cong, 1st Sess 1991

(19) Doolittle S: Breast cancer information: New York State [news]. Am J Public Health 77:874, 1987

(20) Burnett CB: Effect of breast cancer disclosure legislation on physician disclosure practice. Diss Abstr Int [A] 51:662, 1990

(21) Henderson IC, Danner D: Legal pitfalls in the diagnosis and management of breast cancer. Hematol Oncol Clin North Am 3:823-842, 1989

(22) Sugarman J, Powers M: How the doctor got gagged. The disintegrating right of privacy in the physician–patient relationship. JAMA 266:3323-3327, 1991

(23) Cassileth BR, Zupkis RV, Sutton-Smith K, et al: Informed consent—why are its goals imperfectly realized? N Engl J Med 302:896-900, 1980

(24) Lidz CW, Meisel A, Osterweis M, et al: Barriers to informed consent. Ann Intern Med 99:539-543, 1983

(25) Mazur DJ: Why the goals of informed consent are not realized: treatise on informed consent for the primary care physician. J Gen Intern Med 3:370-380, 1988

(26) Wear S: Informed Consent. Patient Autonomy and Physician Beneficence Within Clinical Medicine. Boston: Kluwer Acad Publ, 1993, pp 45-46

(27) Wu WC, Pearlman RA: Consent in medical decision making: the role of communication. J Gen Intern Med 3:9-14, 1988

(28) President's Commission for the Study of Ethical Problems in Medicine and Biomedical Behavioral Research: Making Health Care Decisions. The Ethical and Legal Implications of Informed Consent in the Patient–Practitioner Relationship. Vol 1: Report. Washington, DC: US Govt Print Off, 1983-413-987/1329, 1982, pp 17-18

(29) Liberati A, Patterson WB, Biener L, et al: Determinants of physicians' preferences for alternative treatments in women with early breast cancer. Tumori 73:601-609, 1987

(30) Belanger D, Moore M, Tannock I: How American oncologists treat breast cancer: an assessment of the influence of clinical trials. J Clin Oncol 9:7-16, 1991

(31) Deber RB, Thompson GG: Who still prefers aggressive surgery for breast cancer? Implications for clinical applications of clinical trials. Arch Intern Med 147:1543-1547, 1987

(32) Tarbox BB, Rockwood JK, Abernathy CM: Are modified radical mastectomies done for T1 breast cancers because of surgeon's advice or patient's choice? Am J Surg 164:417-420, 1992

(33) Kanouse DE, Winkler JD, Kosecoff J, et al: Changing Medical Practice Through Technology Assessment. An Evaluation of the NIH Consensus Development Program. Ann Arbor, Mich: Health Administration Press, 1989

(34) Ford LG, Hunter CP, Diehr P, et al: Effects of patient management guidelines on physician practice patterns: the Community Hospital Oncology Program experience. J Clin Oncol 5:504-511, 1987

(35) Pierce PF: Women's experience of choice: confronting options for treatment of breast cancer, chap 21. *In* Breast Cancer, Collaborative Management (Harness JK, Oberman HA, Lichter AS, et al, eds). Chelsea, Mich: Lewis Publ, 1988

(36) Pierce PF: Deciding on breast cancer treatment: a description of decision behavior. Nurs Res 42:22-28, 1993

(37) Owens R, Ashcroft J, Leinster S, et al: Informal decision analysis with breast cancer patients: an aid to psychological preparation for surgery. J Psychosoc Oncol 5:23-33, 1987

(38) Ward S, Heidrich S, Wolberg W: Factors women take into account when deciding upon type of surgery for breast cancer. Cancer Nurs 12:344-351, 1989

(39) Schain WS, Findlay P, d'Angelo T, et al: A prospective psychosocial assessment of breast cancer patients receiving mastectomy or radiation therapy in a randomized clinical trial. Proc ASCO 4:248, 1985

(40) Wolberg WH, Romsaas EP, Tanner MA, et al: Psychosexual adaptation to breast cancer surgery. Cancer 63:1645-1655, 1989

(41) Ashcroft JJ, Leinster SJ, Slade PD: Breast cancer—patient choice of treatment: preliminary communication. J R Soc Med 78:43-46, 1985

(42) Morris J, Ingham R: Choice of surgery for early breast cancer: psychosocial considerations. Soc Sci Med 27:1257-1262, 1988

(43) Farrow DC, Hunt WC, Samet JM: Geographic variation in the treatment of localized breast cancer. N Engl J Med 326:1097-1101, 1992

(44) Nattinger AB, Gottlieb MS, Veum J, et al: Geographic variation in the use of breast-conserving treatment for breast cancer. N Engl J Med 326:1102-1107, 1992

(45) Cal Health & Safety Code §1704.7, 1990

(46) Cal Health & Safety Code §1704.75, 1993

(47) HB 3186, 178th Mass Gen Leg, Reg Sess 1993

Notes

[1]The New Jersey mandate for development of an informational booklet was enacted as a 1993 amendment to the original statute. The informational booklet is not yet available.

[2]While the Maine statute requires that written information be provided to the patient, it permits use of brochures from recognized professional organizations in lieu of a newly developed summary.

[3]*Editor's note:* SEER is a set of geographically defined, population-based central tumor registries in the United States, operated by local nonprofit organizations under contract to the NCI. Each registry annually submits its cases to the NCI on a computer tape. These computer tapes are then edited by the NCI and made available for analysis.

Manuscript received January 10, 1994; revised May 26, 1994; accepted June 1, 1994.

The New York State Department of Health's
BREAST CANCER TREATMENT QUALITY ADVISORY PANEL

announces a

Public Forum:
Consumer Input on Barriers to High Quality Treatment

The Advisory Panel would like your comments on the following topics:

- What factors should the Panel consider in determining the quality of breast cancer treatment? What can the Panel and the State Health Department do to help promote the highest quality of breast cancer treatment?
- How does breast cancer treatment impact your life? What can providers do to ease that impact?
- What are the barriers to getting high quality breast cancer treatment and how can the Panel help minimize them?
- What types of information does someone need to make decisions about breast cancer treatment?

WHERE:	New York City 5 Penn Plaza Room 302	Long Island Holiday Inn of Plainview South Room	Syracuse ONCENTER 800 So. State St., Meeting Room 4
WHEN:	January 17, 1995 10:00 am to 3:00 pm	February 28, 1995 10:00 am to 3:00 pm	March 8, 1995 10:00 am to 3:00 pm

NEW YORK STATE'S NEW BREAST CANCER TREATMENT QUALITY INITIATIVE

Every year in New York State about 12,000 new cases of breast cancer are diagnosed. The Department of Health's Office of Quality Improvement is working in collaboration with women and clinicians to help assure that each one of these women, and all women with breast cancer, receive the best possible care.

In 1994, New York State's Department of Health convened a panel that includes some of the leading breast cancer treatment clinicians in the state from a wide range of specialties, nurses, public health officials and women who have had breast cancer to advise the state on breast cancer treatment quality and its improvement. The Breast Cancer Treatment Quality Panel includes assistance of technical advisors from the health care industry and cancer tumor registries.

Input is needed from women in a number of important areas concerning treatment and its impact. The Panel is particularly interested in your perspective on barriers to quality treatment decisions and how best to assess and use information on quality of patient care, factoring in quality of life concerns in treatment and outcome measurement. Advice on the types of information consumers would like to have available is also important to the panel.

We urge individual consumers, women's groups and others to consider speaking or submitting testimony at one of the three public forums that the Panel is sponsoring. We are interested in learning about the unique problems and needs of various sectors of New York State.

FORMAT: Preregistration is required. You may submit written comments either at time of registration or on the day of the meeting. Speakers will be allotted five minutes each. Written comments may be any length.

Time slots are limited. Please respond today by writing, faxing or calling.

Office of Quality Improvement, New York State Department of Health, ESP Corning Tower, Room 1223, Albany, New York 12237, ATTN: Breast Cancer Public Forum or **FAX (518) 474-1202.**

Written comments from those unable to attend are welcome. For further information, please contact **Andrew Stern at (518) 474-5090.**

Clinton Presidential Records
Digital Records Marker

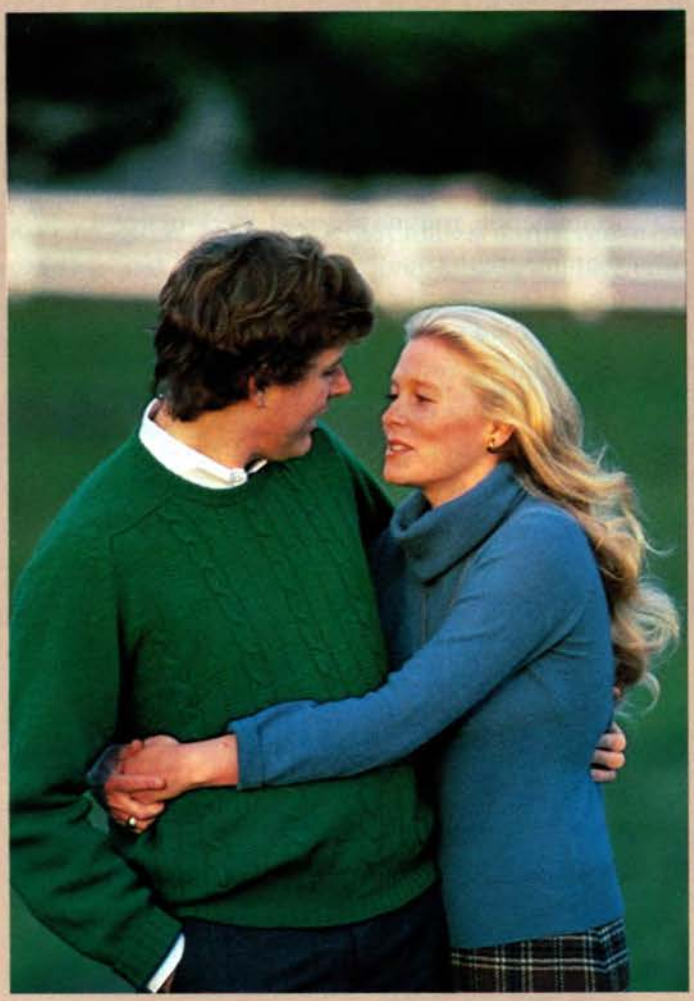
This is not a presidential record. This is used as an administrative marker by the William J. Clinton Presidential Library Staff.

This marker identifies the place of a publication.

Publications have not been scanned in their entirety for the purpose of digitization. To see the full publication please search online or visit the Clinton Presidential Library's Research Room.

Now, Breast Cancer Has Virtually Nowhere To Hide.

BREAST CARE
The Good News



By W.R. Spence, M.D.

Clinton Presidential Records
Digital Records Marker

This is not a presidential record. This is used as an administrative marker by the William J. Clinton Presidential Library Staff.

This marker identifies the place of a publication.

Publications have not been scanned in their entirety for the purpose of digitization. To see the full publication please search online or visit the Clinton Presidential Library's Research Room.

New Frontiers in Breast Cancer Imaging and Early Detection

From Missiles to Mammograms
A Capitol Hill Briefing

October 11, 1994
8:30am-12:30pm
U.S. Capitol Building,
Room HC 5

Imaging Exhibits
10:00am-6:00pm
Rayburn House
Office Building
Foyer

Sponsored by the PHS Office on Women's Health and the National Cancer Institute in collaboration with the Congressional Caucus for Women's Issues

For more information, contact
Dr. Susan Blumenthal
Deputy Assistant Secretary for Health (Women's Health)
202-690-7650

Clinton Presidential Records
Digital Records Marker

This is not a presidential record. This is used as an administrative marker by the William J. Clinton Presidential Library Staff.

This marker identifies the place of a publication.

Publications have not been scanned in their entirety for the purpose of digitization. To see the full publication please search online or visit the Clinton Presidential Library's Research Room.

The Older You Get, the more You Need a Mammogram

American Cancer Society

The older you get, the more you need a mammogram.

AMERICAN CANCER SOCIETY®

THERE'S NOTHING MIGHTIER THAN THE SWORD.

Clinton Presidential Records
Digital Records Marker

This is not a presidential record. This is used as an administrative marker by the William J. Clinton Presidential Library Staff.

This marker identifies the place of a publication.

Publications have not been scanned in their entirety for the purpose of digitization. To see the full publication please search online or visit the Clinton Presidential Library's Research Room.

Questions and Answers About Choosing a Mammography Facility

National Institutes of Health

National Cancer Institute

Questions and Answers About Choosing a Mammography Facility

Quality and cost are important to consider when choosing a mammography facility.

One sign of quality is certification by the American College of Radiology's (ACR) voluntary accreditation program. Facilities accredited by ACR have approved equipment, personnel, and procedures. They have doctors and other staff members who are specially trained to perform and read breast x-rays. They have the equipment and procedures to provide high-quality mammograms with the lowest possible radiation exposure. Look for an ACR-accreditation sticker on your mammography machine.

To find an ACR approved facility near you or to see if a facility is accredited, call the National Cancer Institute's Cancer Information Service (CIS) at 1-800-4-CANCER (1-800-422-6237).

Some high quality mammography facilities may not be accredited through the ACR voluntary program. If you are considering using such a facility, you should be sure that your mammogram will be taken with appropriate equipment by a properly trained x-ray technologist. Call and ask the staff at the facility these five questions before you make an appointment. A qualified staff can answer these questions easily.

NATIONAL INSTITUTES OF HEALTH
National Cancer Institute

Clinton Presidential Records
Digital Records Marker

This is not a presidential record. This is used as an administrative marker by the William J. Clinton Presidential Library Staff.

This marker identifies the place of a publication.

Publications have not been scanned in their entirety for the purpose of digitization. To see the full publication please search online or visit the Clinton Presidential Library's Research Room.

HEALTHY
BETH ISRAEL
16TH STREET
FIRST AVE
NEIGHBOR
SM