

Screening Mammography Benefit Controversies Sorting the Evidence

Stephen A. Feig, MD

KEYWORDS

• Screening • Mammography • Benefit • Controversies

KEY POINTS

- Numerous clinical studies have confirmed that screening women age 40 years and older reduces breast cancer mortality by 30% to 50%.
- Several factors including faster breast cancer growth rates and lower breast cancer incidence among younger women, as well as shorter life expectancy and more comorbid conditions among older women, should also be considered in screening guidelines.
- Annual screening beginning at age 40 years and continuing with no upper age limit, as long as a woman has a life expectancy of at least 5 years and no significant comorbid conditions, is currently recommended by the American Cancer Society, the American College of Radiology, and the Society of Breast Imaging.

Screening mammography by virtue of its ability to substantially reduce death rates from the most common type of malignancy among women and the second leading cause of their death from cancer represents one of the major medical achievements of our time. Yet, unlike other medical advancements, the value of screening women age 40 years and older did not become apparent until after many years of clinical trials which began in the 1960s. Lengthy observational follow-up was required, because breast cancer is a chronic disease. During subsequent decades, there have been numerous improvements in technology, beginning with the replacement of direct exposure film mammography by film/screen mammography, the more recent conversion to digital mammography, and the current clinical evaluation of digital tomosynthesis. There have also been improvements in performance of mammography, such as better breast compression paddles and automatic exposure devices, mammographic grids, use of the mediolateral oblique (MLO) view instead of

the straight mediolateral view. Some advances, such as use of 2-view screening (craniocaudal and MLO), instead of a single MLO view alone, screening at an annual rate rather than semiannual intervals, and double reading by 2 radiologists, have still not been universally accepted because of concerns regarding cost-effectiveness.

Screening controversies began in 1975 and continue. Some issues are legitimate, but most have been artificially contrived. No other medical test has been more thoroughly scrutinized and debated over the past 40 years. Keeping informed on these complex issues has been challenging for all physicians, including breast imagers. The public especially deserves empathy, because their information is channeled through the nonmedical media. Thus, the purpose of this article is to assess our current knowledge of screening benefits. Comprehensive reviews of adverse consequences and costs of screening may be found in the author's previous articles in the *Radiologic Clinics of North America*.^{1,2}

Department of Radiological Sciences, University of California Irvine Medical Center, 101 City Drive South, Orange, CA 92869-3298, USA
E-mail address: sfeig@uci.edu

Radiol Clin N Am 52 (2014) 455–480

<http://dx.doi.org/10.1016/j.rcl.2014.02.009>

0033-8389/14/\$ – see front matter © 2014 Elsevier Inc. All rights reserved.

RANDOMIZED TRIALS HAVE PROVEN THAT EARLY DETECTION REDUCES BREAST CANCER DEATH RATES

The ability of widespread screening to detect breast cancers at smaller size and earlier stage than encountered in the general population was first established at the Breast Cancer Detection Demonstration Project (BCDDP), a program that screened 280,000 women throughout the United States with both mammography and physical examination from 1973 to 1981, sponsored by the American Cancer Society (ACS) and National Cancer Institute. In this program, 39% (1375) of the 3548 cancers were found by mammography alone, 7% (257) by clinical examination alone, and 51% (1805) by both mammography and clinical examination.³ The 20-year relative survival rates at the BCDDP were 80.5% (overall), 85% for cancers detected by mammography alone, 82% for cancers detected by physical examination alone, and 74% for cancers detected by both mammography and physical examination.⁴ These rates can be compared with the contemporaneous 20-year survival rate of 53% among US women who were largely not being screened.⁵ Although these results were promising, there are several reasons why improved survival rates among such women who volunteer to be screened do not necessarily establish benefit from screening. They include selection bias, lead time bias, length bias, and interval cancers.⁶ Thus, differences in survival rates may be influenced by factors other than the screening process itself.

Selection bias refers to the possibility that women who volunteer for screening differ from those who do not volunteer in ways that may alter the outcome of their disease, such as health status and behavioral factors. Therefore, survival rates in screened and nonscreened women may be influenced by factors other than the screening process itself.

Lead time bias implies that screening may affect the date of detection but not necessarily the date of death from breast cancer. Let us suppose that a woman who has never been screened finds her breast cancer serendipitously in 2009. She dies from her disease 5 years later, in 2014. If this same woman had been screened, her cancer might have been detected by mammography in the year 2005. Although small, the cancer detected in this woman by mammography might have dissemination beyond the breast. Despite screening, the woman dies from her disease in the year 2014. Because of screening, she is said to have survived for 9 years instead of 5 years. Therefore, the seemingly 4-year improvement in survival may not be real.

Length-biased sampling postulates that cancers detected at screening contain a disproportionate number of less aggressive cancers. Their growth rates are so slow that in the absence of screening, they might never reach sufficient size to surface clinically. Even if undetected, such indolent cancers might never result in death.

Possibly, the favorable survival rates for screen-detected cancers might be negated by lower survival rates for faster-growing interval cancers, which are undetected by mammography and surface clinically between screenings.

Considering these potential biases, benefit from screening cannot be proved by observation of improved survival rates. Rather, such proof requires prospective comparison of breast cancer death rates among a study group of women offered screening and a control group of women not offered screening in a randomized clinical trial (RCT).⁶ Apart from the offer to be screened, these groups should not differ in any other substantial way. Therefore, a statistically significant difference in breast cancer deaths between the groups on follow-up represents incontrovertible proof of benefit from the screening. Observation of lower mortality for the screened group in a well-designed and well-conducted RCT is not affected by selection bias, lead time bias, length bias, or interval cancers.

RESULTS OF RCTS

Seven population-based trials of breast cancer screening by mammography alone or in combination with physical examination have been conducted. They are as follows: (1) the Health Insurance Plan of Greater New York (HIP) trial,⁷ (2) the Swedish Two-County trial consisting of Kopparberg and Ostergotland counties,^{8–10} (3) the Malmö (Sweden) Mammographic Screening trial,^{11–15} (4) the Stockholm (Sweden) trial,^{14–17} (5) the Gothenburg (Sweden) Breast Screening trial,^{14,15,18–20} (6) the Edinburgh (Scotland) trial,^{21,22} and (7) the UK Age trial.²³ In a population-based RCT, study and control groups are randomly selected from a predefined population. There has also been 1 non-population-based RCT, the National Breast Screening Study of Canada (NBSSC).^{24–26} In a non-population-based RCT, study and control groups are randomly selected from women who volunteer to participate.

Protocols and results for women of all ages at entry into these 8 RCTs are shown in **Table 1**. Mortality reduction is equal to 1 minus the relative risk (RR) of dying from breast cancer in the study group women versus the control group. The HIP trial, the

Table 1
Randomized trials of mammography screening: protocols and results

Trial (y)	Age at Entry (y)	Number of Views	Frequency of Mammography (mo)	Rounds (n)	CBE	Follow-Up (y)	Relative Risk (95% Confidence Interval)	Mortality Reduction (%)
HIP trial (1963–1969)	40–64	2	12	4	Annual	18	0.78 (0.61–0.97)	22
Malmö, Sweden (1976–1986)	46–69	1–2	18–24	5	None	20	0.78 (0.65–0.95)	22
Two-County Swedish (1979–1988)	40–74	1	23–33	4	None	30	0.68 (0.54–0.80)	32
Edinburgh, Scotland (1979–1988)	45–64	1–2	24	4	Annual	14	0.78 (0.62–0.97)	22
NBSSC (1980–1987)	40–49	2	12	5	None	13	0.97 (0.78–1.33)	0.3
NBSSC-1								
NBSSC-2	50–59	2	12	5	Yes	13	1.02 (0.78–1.33)	–2
Stockholm, Sweden (1981–1988)	40–64	1	28	2	None	16	0.90 (0.63–1.28)	10
Gothenburg, Sweden (1982–1988)	40–59	2	18	4	None	14	0.79 (0.58–1.08)	21
UK Age trial (1991–2005)	39–41	1–2	12	8	None	10	0.83 (0.66–1.04)	17

Data from Refs. ^{7,9,11,12,17,19,20,22,23,25,26}

first RCT ever conducted, found a 22% reduction in breast cancer deaths (RR = 0.78).⁷

The Two-County Swedish trial was the first to show a statistically significant benefit from screening by mammography alone. The latest 30-year follow-up for this trial found a 32% reduction in breast cancer deaths among women aged 40 to 74 years at entry.⁹ In the Edinburgh trial, screening by annual physical examination and biennial mammography resulted in a statistically significant 22% decrease in breast cancer deaths among women aged 45 to 64 years at entry.^{21,22}

The Gothenburg Breast Screening trial had a 21% reduction in deaths from breast cancer among women aged 40 to 59 years at entry into screening, a finding that had marginal statistical significance.^{14,15,18–20} The Malmö Mammographic Screening trial found a significant 22% reduction in breast cancer deaths among women who began screening between ages 45 and 60 years.^{11–15} The Stockholm trial described a nonsignificant 10% reduction in breast cancer deaths among women screened between 40 and 60 years of age, which was not statistically significant.^{14–17}

Combined results from a 15.8-year follow-up of women aged 38 to 75 years at entry into 4 Swedish trials (Malmö, Ostergotland, Stockholm, and Gothenburg) showed a statistically significant 21% reduction (confidence interval = 0.70–0.89) in breast cancer mortality with screening.^{14,15}

NBSSC failed to show any benefit for mammography screening in women aged 50 to 59 years.^{24,25} In that trial, women undergoing annual mammography and physical examination were compared with those being screened by physical examination alone. Possible explanations for the variance between NBSSC results and those of the 7 other randomized trials include poor technical quality of mammography improper study design, and control group contamination with advanced cancers.^{27–36}

Of the 8 randomized screening trials, 7 showed evidence of benefit from screening.³⁷ Breast cancer mortality reduction was statistically significant in each of 4 trials (HIP, Swedish Two-County, and Edinburgh) and in combined results from the Stockholm, Malmö, Ostergotland, and Gothenburg trials, and marginally significant in the Gothenburg trial. Only 1 trial, the NBSSC, found no evidence of benefit.

Origins of the Controversy Regarding Women Aged 40 to 49 Years

Initial reports from the HIP trial found a difference in breast cancer death rates between study and control groups for women 50 years of age and

older at entry that was apparent by year 4.⁷ However, a difference for women aged 40 to 49 years did not emerge until 7 to 8 years of follow-up. By 18 years of follow-up, the reduction in breast cancer deaths among study women aged 40 to 49 years at entry was 23%, the same as that for women aged 50 to 64 years at entry. Yet, even by that time, benefit for younger women was still not statistically significant according to the original investigators.⁷ This lack of statistical significance was a consequence of the smaller number of younger women enrolled and the lower breast cancer incidence. Nevertheless, the apparent lack of statistically significant benefit led to controversy regarding screening of women in their 40s.^{37–39}

However, the HIP trial was designed to determine the efficacy of screening a single group of all women aged 40 to 65 years rather than the efficacy of screening separate age groups. Attempts to subdivide the study group reduced statistical power. The observation that results for younger women lacked statistical significance was often cited in the subsequent screening debate. The fact that the data for women aged 50 to 59 years as well as 60 years and older at entry, when analyzed separately, also lacked statistical significance was largely ignored.^{7,40} Subsequently, Chu and colleagues,⁴¹ using a different method of analysis, found statistically significant mortality reductions of 24% for women aged 40 to 49 years and 21% for those aged 50 to 64 years at entry into the HIP trial. Despite the report by Chu and colleagues, some observers were still not convinced that screening would benefit women in their 40s for several reasons. First, the reduction in the breast cancer death rates in the trials for younger women did not appear until several years after appearance of the reduction for women older than 50 years. Second, results for younger women were not statistically significant for any other individual trial until 1997.

The controversy intensified in 1992, with publication of the 7-year follow-up report from the NBSSC.²⁴ This study found no evidence of benefit among women aged 40 to 49 years who were offered 5 annual screenings by mammography and physical examination. There are several explanations for these disappointing results. First, the technical quality of mammography was poor.^{27–29} During most of the trial, more than 50% of the mammograms were poor or completely unacceptable by an independent expert panel, even as assessed by the standards of the day. Second, there are indications that the randomization process through which women were assigned to study and control groups was undermined.^{30–35}

All women were given a physical examination before randomization to the study trial. This protocol may have allowed preferential allocation of women with palpable masses (later-stage breast cancers) to the study group.^{29–36} As a likely consequence, an excess of late-stage breast cancers and breast cancer deaths was found in the study group compared with the control group throughout the trial.

Mortality Reduction Among Women Screened in Their 40s

Beginning in 1993, several successive meta-analyses of combined data for multiple RCTs were performed to accrue more women-years of follow-up than possible from any 1 RCT alone. However, the earliest meta-analyses, published in 1993 and 1995, suggested little if any benefit from screening women younger than 50 years.^{42–44}

Subsequent meta-analyses published by Smart and colleagues⁴⁵ in 1995 and the Falun Meeting Committee⁴⁶ in 1996 included later follow-up data. These studies showed a statistically significant mortality reduction of 24%, for women aged 40 to 49 years at entry into the 7 population-based RCTs (Table 2). They also found a 15% to 16% mortality reduction, which barely missed statistical significance when the NBSSC, a non-population-based RCT, was also included. A meta-analysis of these trials, published by Hendrick and colleagues⁴⁷ in 1997, found statistically significant mortality reductions among women invited to undergo screening in their 40s: 18% for all 8 RCTs and 29% for the 5 Swedish RCTs (see Table 2). Thus, with increasing length of follow-up, successive meta-analyses have shown progressively greater and statistically significant mortality reductions for women who began

screening between 40 and 49 years of age. Regardless of whether NBSSC results are included or excluded, meta-analyses for screening women aged 40 to 49 years show statistically significant benefit. Subsequently, 2 other individual RCTs besides the HIP were each able to show benefit for women aged 40 to 49 years (Table 3). Bjurstam and colleagues^{18–20} reported a statistically significant 45% mortality reduction for women aged 39 to 49 years at randomization in the Gothenburg trial. Andersson and Janzon¹³ reported a statistically significant 35% breast cancer mortality reduction for women in the Malmö trial who began screening mammography at age 45 to 49 years. A randomized trial of women aged 39 to 41 years at entry in the UK study showed a statistically nonsignificant 17% mortality reduction for those invited and 24% mortality reduction for those screened.¹³ Deficiencies that restricted results in the study included use of 1 rather than 2 mammographic views on incidence screens, which the investigators knew missed 20% to 25% of cancers, failure to biopsy clustered calcifications (causing them to miss additional small cancers), low recall rates and insufficient women-years of follow-up, a low 68% compliance rate, considerable contamination of the unscreened control population with private facilities offering mammograms, and mortality follow-up limited to 10 years.^{48,49} Later analysis is likely to show more benefit. We also now have data from many service screening studies, showing statistically significant mortality reduction from screening women aged 40 to 49 years, equal to that found for screening older women. For example, in Ostergotland and Dalarna counties in Sweden, Tabar and colleagues⁵⁰ found breast cancer mortality reductions of 48% for women aged 40 to 49 years and 44% for all women aged 40 to 69 years. No such decline was seen in the 30-year to 39-year age group, in which none was offered screening. In British Columbia, Coldman and colleagues⁵¹ found that breast cancer deaths were reduced 39% among women aged 40 to 49 years versus 40% for all women aged 40 to 79 years. In the 2 northern Swedish counties of Vasternorrland and Norrbotten, breast cancer deaths were reduced 35% to 38% for women aged 40 to 49 years versus 30% for all women aged 40 to 74 years.

Table 2
Meta-analyses of RCTs of mammography showing statistically significant mortality reduction for women aged 40 to 49 years

Trials	Follow-Up (y)	Mortality Reduction (%)
All 8 trials ^a	10.5–18.0	15
7 trials ^b	7.0–18.0	24
5 Swedish trials ^c	11.4–15.2	29

^a HIP, 5 Swedish trials, Edinburgh trial. UK Age trial and NBSSC-1.

^b All trials except NBSSC-1.

^c Malmö Mammographic Screening trial; Swedish Two-County trial, Kopparberg and Ostergotland; Stockholm trial; and Gothenburg Breast Screening trial.

Data from Refs.^{45–47}

Screening Women 75 Years of Age and Older

The question of mammographic screening for elderly women is clinically relevant, because there are more than 10 million women aged 75 years and older in the United States. The average life expectancy for women at age 75 years is 13 years.⁵²

Table 3
Follow-up of RCTs of mammography showing statistically significant breast cancer mortality reduction for women aged 40 to 49 years

Trial (y)	Age at Entry (y)	Number of Views	Frequency of Mammography (mo)	Clinical Breast Examination	Follow-Up (y)	Mortality Reduction (%)
HIP (1963–1969)	40–49	2	12	Annual	18.0	24
Malmö Mammographic Screening Program (1976–1990)	45–49	1–2	18–24	None	12.7	36
Gothenburg Breast Screening trial (1982–1988)	39–49	2	18	None	12.0	45

Data from Refs.^{13,18,41}

Women with good general health have a longer than average life expectancy. Thus, it is reasonable to expect that elderly women might benefit from screening.⁵³ Reduction in breast cancer mortality among women aged 50 years and older becomes apparent within 4 years of entry into RCTs. Therefore, for many older women with screening-detected breast cancer, death from another illness does not occur before they experience the benefit from screening.⁵⁴

Benefit from screening women 75 years older has not been proved, because this age group was not included in any RCT.⁵⁵ Nevertheless, there is no biological reason why early detection should not be effective for these women. The detection sensitivity of mammography is higher in elderly women because of their generally more fatty breast composition.^{56,57} However, because of the lower life expectancy in older women, there is a potential for diagnosis of tumors that would not have given rise to symptoms during their remaining lifetime. In addition, screening is warranted only if the woman is suitable for appropriate therapy in the event of a cancer diagnosis. Taking these considerations, especially comorbidity issues, it seems reasonable that screening mammography should be performed in women aged 75 years having generally good health and life expectancy of 5 years or longer.^{58,59}

Why Do Randomized Trials Underestimate the Benefit from Screening?

There are at least 6 reasons why results from all the early RCTs have underestimated the benefit to an individual woman undergoing screening with current advanced mammography technology:

- Mammographic image quality below current standards

- Use of only 1 mammographic view per breast in some RCTs
- Noncompliance of some study group women
- Contamination of the control group
- Excessively long screening intervals (more than annual)
- Inadequate number of screening rounds

First, there have been many technical improvements in mammographic technique since the early 1980s, when these trials were conducted. These innovations in mammographic equipment, screen-film systems, and processing as well as replacement of analog (film) mammography by digital mammography allow images to have better sharpness, exposure, and contrast.^{60–65} Better image quality facilitates early detection of breast cancer.^{62–65}

Second, women in the RCTs were mostly screened with 1 view per breast. The current standard, 2 views per breast examination has been shown to detect 3% to 20% more cancers than found using an MLO view alone.^{66–71} Of the 7 population-based RCTs, only the HIP trial⁷ used 2 views on all examinations. For example, the Gothenburg trial¹⁸ used 2-view mammography at the first screen and either single view or 2 views at subsequent screens, depending on breast density. The Malmö trial¹¹ used 2 views in the first 2 screenings but only an MLO view alone on all subsequent screenings, except in patients with dense breasts. The Edinburgh trial²¹ used 2-view screening on the initial screening but only 1 view on all subsequent screenings. The Stockholm and Swedish Two-County trials^{8,16} used a single MLO view in all screenings.

Two fundamental reasons why RCTs underestimate the benefit from screening are as follows: (1) not all study group women accept the invitation to be screened (noncompliance); and (2) some

control group women obtained mammography screening outside the trial (contamination). Yet, to avoid selection bias, an RCT must compare the breast cancer death rate among all study group women, both screened and nonscreened, with that among all control group women, including those who are screened on their own initiative. Thus, both noncompliance of some study women and contamination of control group women reduce the calculated benefit from RCTs.

Among the RCTs, the noncompliance rate ranged from 10% to 39%.⁷² Studies performed on data from the individual trials have estimated that if all women in the study group had attended each screening round, there would have been an additional reduction in breast cancer deaths. Data from the Gothenburg, Malmö, and Swedish Two-County trials as well as the NBSSC indicate that the rate of control group contamination ranged from 13% to 25%.⁷²

Randomized trials might have also underestimated the potential benefit of screening, because screening intervals were too long.⁷³ Aside from the HIP trial, screening intervals in RCTs have been longer than the annual intervals now recommended. For example, women in the Swedish Two-County trial were screened every 24 to 33 months,⁸ and those in the Edinburgh trial every 24 months.²¹ Numerous studies suggest that greater benefit should result from annual screening, especially for women aged 40 to 49 years, in whom breast cancer growth rates seem to be faster.⁷⁴⁻⁷⁷ From a tumor growth rate model, Michaelson and colleagues⁷⁸ calculated that annual screening would result in a 51% reduction in the rate of distant metastatic disease compared with a 22% reduction at a screening interval of 2 years.

It has been estimated that the use of annual screening in the Swedish Two-County trial could have resulted in an additional 18% mortality reduction for women aged 40 to 49 years at entry, who were screened every 2 years, and an additional 12% mortality reduction for women aged 50 to 59 years at entry, who were screened every 33 months.⁷⁹ For women aged 39 to 49 years at entry into the Gothenburg trial, who were screened every 18 months, it has been estimated that annual screening could have resulted in an additional 20% mortality reduction.⁸⁰

Several investigators have used mathematical models of RCT data to calculate the benefit to an average woman who is screened every year and for whom results are not affected by noncompliance and contamination.^{46,74,80,81} For example, from an observed 45% reduction in breast cancer mortality among women aged 39 to 49 years offered screening every 18 months in the

Gothenburg trial, Feig calculated that the mortality reduction could have been as high as 65% with annual screening at the observed 80% compliance rate and as high as 75% at a 100% compliance rate.⁸⁰

The fact that the results of some trials were based on a short screening period and small numbers of screening rounds represents a sixth reason why such trials may underestimate the potential benefits of screening. Most trials do not achieve the highest steady state of benefit until after 8 to 11 years of follow-up.⁸² Short trial durations substantially underestimate the mortality reduction that would be achieved with continued annual screening from ages 40 to 80 years, inclusive.

Validity of RCT Results: the Gotzsche and Olsen Controversy

From results from RCTs conducted over the past quarter of a century and involving more than 500,000 women, consensus has been reached in the medical community in favor of screening mammography.

In the face of such near-unanimous agreement about the value of screening mammography, 2 studies made the seemingly incredible claim that none of the trials provided any convincing evidence that screening prevents breast cancer deaths.^{83,84} In these studies, Gotzsche and Olsen, asserted that only 2 of the 8 screening trials (the Malmö Mammographic Screening trial and the NBSSC) were valid and that neither of these trials found evidence of benefit. The studies, published in 2000 and 2001, were highly publicized because of the sensational nature of their claim, which questioned the widely held belief in the efficacy of early detection through mammography screening.

The only 2 points of which Gotzsche and Olsen and all other observers agree are as follows: (1) that the NBSSC failed to find benefit for screening in women aged 50 to 70 years with mammography and clinical examination versus clinical examination alone,^{24,25} and (2) the NBSSC found no benefit for mammographic screening of women aged 40 to 49 years.²⁶ However, at this point, advocates of screening part ways with Gotzsche and Olsen, because their explanations for the negative NBSSC results are vastly different. It is difficult to understand how Gotzsche and Olsen could view the NBSSC as the paradigm of a well-conducted study for several reasons. First, independent reviews found that the technical quality of mammography in the NBSSC was poor, even when measured by the standards of the 1980s,

when the trial was conducted.^{27–29} Second, performance of clinical breast examination before randomization of trial subjects may have allowed channeling of symptomatic women into the study group.^{29–36} The finding of an excess of advanced cancers in the study group aged 40 to 49 years was not shown in any of the other 7 trials and suggests that randomization at NBSSC was not performed blindly.

Third, NBSSC was not a population-based trial. Rather, participants were self-selected volunteers. Because self-selected women are more likely to be symptomatic, adequate randomization of such subjects is more problematic, especially when clinical examination has already been performed. Self-selected asymptomatic women may have higher survival rates than randomly selected asymptomatic women. Thus, benefit may be harder to show in a trial with such subjects than in a population-based trial. Contrary to the judgment of Gotzsche and Olsen, almost any of these problems to trial design and implementation could render the NBSSC incapable of providing meaningful results.

The statement by Gotzsche and Olsen that the Malmö trial showed no evidence of benefit is more difficult to understand. Gotzsche and Olsen considered only an early report of a small but insignificant 5% mortality reduction among women aged 45 to 70 years.¹² These investigators ignored later reports of breast cancer mortality reductions of 19% for women aged 45 to 70 years, 26% for women aged 55 to 70 years, and 36% for women aged 45 to 50 years at entry into the Malmö trial.¹³ Moreover, several months after publication of the second Olsen and Gotzsche study, Miettinen and colleagues reported using 3-year moving averages of RR estimates to estimate that the true mortality reduction from the Malmö trial was 55% for women aged 55 to 69 years and 60% for women aged 45 to 57 years at entry into screening.⁸²

Gotzsche and Olsen also claimed to have identified age differences of 1 to 5 months between study and control groups in the HIP and Edinburgh trials and all Swedish trials aside from the Malmö trial. The writers suggested that the observed reduction in breast cancer death rates was caused by these age differences rather than by the screening process itself. Gotzsche and Olsen were unaware that when the screening trials used cluster randomization rather than individual randomization, such small age differences were not unexpected, in some cases biased the studies against screening, and were in any case fully accounted for in analyses, leaving the conclusion of a significant mortality reduction unchanged.^{85–87}

Screening trials and therapeutic trials are different in nature and may be different in design. In therapeutic trials, all participants have disease. The main variables are treatment versus no treatment and dose regimen. Study and control groups are small. Individual randomization is required, and small age differences are significant to the study. In screening mammography trials, there is low disease prevalence, so large study and control groups are necessary. For this reason, individual randomization may not be necessary. Individual randomization may not be practical, and cluster randomization is usually necessary.⁸⁷

The age difference between the 2 groups that Gotzsche and Olsen purported to have discovered in the Swedish Two-County trial had been previously acknowledged by Tabar and colleagues⁸⁸ in 1989. After adjustment for age, mortality was only minimally different: 31% versus 30% for women aged 40 to 70 years in the Swedish Two-County trial and 45% instead of 46% for women aged 39 to 49 years in the Gothenburg trial.⁸⁶ Thus, there was no way that these small differences in age could have altered the overall conclusion that mammography screening results in a substantial reduction in deaths from breast cancer.

In another criticism, Gotzsche and Olsen suggested that assignment of the cause of death among women in the Swedish screening trials might have been inaccurate. Accurate assignment of cause of death is critical to proper assessment of trial results. Death in a woman with breast cancer may be either causally related or unrelated to her malignancy. Because screening trials compare deaths caused by breast cancer in women in study groups and control groups, attribution of the cause of death must be performed in a consistent and unbiased manner. However, the criticism by Gotzsche and Olsen was baseless. The methods had been previously described in detail by Nystrom and colleagues.⁸⁹ The process consisted of independent blind evaluation by 4 physicians and resulted in unanimous agreement in a remarkable 93% of cases.^{14,15,89}

Gotzsche and Olsen also observed that no statistically significant decrease in death rates from all causes combined had yet been shown in any of the Swedish trials. These investigators interpreted this observation to mean that any benefit from reduction in breast cancer deaths would be countered by increased deaths from other causes. This incorrect conclusion disregarded the fact that breast cancer accounts for only about 5% of total mortality. Thus, even the largest individual trial would be unlikely to show any statistically significant decrease in all-cause mortality. On this issue, Gotzsche and Olsen were again proved wrong.

After publication of the second Olsen and Gotzsche study, Nystrom and colleagues^{14,15} were able to find a 2% decrease in all-cause mortality among study group women in 5 Swedish trials combined. In addition, Tabar and colleagues⁹⁰ observed a significant 19% reduction in deaths from all causes among breast cancer cases in the group invited to screening in the Swedish Two-County trial. Thus, the Gotzsche and Olsen conjecture regarding all-cause mortality was incorrect.

To further their thesis that data from the Swedish Two-County trial were unreliable, Gotzsche and Olsen asserted that the reported study group size was different in the articles by Tabar and colleagues. In response to the criticism, Duffy and Tabar acknowledged that the study population size did differ among their published reports. Tabar and colleagues had previously noted that these differences were caused by progressive identification and exclusion of women diagnosed with breast cancer before the trial began. This is an acceptable and commendable practice. The irony of this unjustified criticism is that Tabar and colleagues were faulted for practicing good science.

In their studies, Gotzsche and Olsen also reiterated the conclusion of a study by Sjonell and Stahle,⁹¹ which claimed that widespread service screening in Sweden had not affected breast cancer mortality in the population. The basic mistake made by Sjonell and Stahle in this claim was that they had measured death rates too early after screening was started. Decreased mortality from screening should not be expected until 5 to 8 years after the start of screening. They had mistakenly begun to tally breast cancer deaths before the beginning of the service screening programs that they were attempting to assess. In addition, their calculations did not consider the increase in breast cancer incidence over time.

Although the report by Gotzsche and Olsen received considerable publicity in the US media, no medical organization or government changed its screening policy from their conclusions. After review of the Gotzsche and Olsen studies, 11 leading medical organizations (American Academy of Family Physicians, ACS, American Congress of Obstetricians and Gynecologists, American College of Physicians, American College of Preventive Medicine, American Medical Association, American Society of Internal Medicine, Cancer Research Foundation of America, National Medical Association, Oncology Nursing Society, and the Society of Gynecologic Oncologists) reaffirmed their support of screening in a full-page advertisement published in the *New York Times* on January 31, 2002.⁸⁷ Also,

the NCI and the United States Preventive Services Task Force (USPSTF) concluded that despite Gotzsche and Olsen's contentions, the results from RCTs of screening were still valid. In addition, the Swedish National Board of Health and Welfare, the Danish National Board of Health, the Health Council of the Netherlands, the European Institute of Oncology, and the World Health Organization dismissed Gotzsche and Olsen's arguments and concluded that the evidence for benefit of screening for breast cancer was convincing.⁹²⁻⁹⁶

SERVICE SCREENING STUDIES

After the success of the Swedish randomized trials, organized service screening mammography became routine in nearly all Swedish counties by the 1990s. Unlike randomized trials, which are conducted primarily as clinical research studies, service screening is performed mainly as a public health initiative. Nevertheless, results from service screening projects have provided strong confirmation that screening mammography is effective in reducing mortality from breast cancer.⁹⁷ Data from service screening can be analyzed using 3 different methods: incidence-based mortality (IBM) studies, case-control studies, and trend studies.⁹⁸

IBM Studies

IBM studies compare breast cancer mortality among women who were screened as well as the larger group of women who were offered screening and may or may not have agreed to be screened with breast cancer mortality expected in the absence of screening, which may be estimated in 1 of 3 ways:

1. Mortality in a cohort of women not yet invited to screening
2. Historical mortality data from the same region as well as from both historical and current data from a region
3. Historical data from the same region and if necessary adjusted for change in breast cancer mortality in time in nonparticipants

A study by Tabar and colleagues⁹⁹ measured the effect of mammography in a population in which service screening was offered to all women 40 years and older. These investigators compared breast cancer death rates in 2 Swedish counties over 3 periods: 1968 to 1977, when virtually no women were screened (prescreening era), 1978 to 1987, when half the population was offered screening in the RCT, and 1988 to 1996, after completion of the trial, when screening was

offered to all women, and 85% of the population was being screened. Compared with breast cancer death rates among women aged 40 to 69 years in the prescreening era, breast cancer death rates in 1998 to 1996 were 63% lower for screened women and 50% lower for the entire population (85% screened plus 15% nonscreened) (Table 4). During this time, reductions in death rates from breast cancer for screened women were similar to those for women screened during the trial (63% vs 57%, respectively). However, during the RCT trial period (1978–1987), only half of the population was offered screening; for that era, breast cancer death rate reduction in the entire population was only 21%. It seems probable that screening rather than advances in treatment was responsible for nearly all the benefit. The RRs of breast cancer death among nonscreened women aged 40 to 69 years were similar during the 3 consecutive periods (1.0, 1.7, and 1.19, respectively). Moreover, the breast cancer death rate for women aged 20 to 39 years, virtually none of whom was screened, showed no significant difference (1.0, 1.10, and 0.81, respectively) during these 3 consecutive periods. Possibly, women who agree to be screened have selection bias factors, which, apart from the screening process, improved their survival rates. Even assuming the maximum effect of selection bias, screening was shown to reduce breast cancer deaths by at least 50%.

A study by Duffy and colleagues¹⁰⁰ assessed the effect of service screening in 7 Swedish counties. Among women aged 40 to 69 years, according to breast cancer mortality trends, it was

Table 4
Reduction in population death rates from breast cancer in women diagnosed between ages 40 and 69 years in 2 Swedish counties^a

Screening Status	1979–1987 (Randomized Trial) (%)	1988–1996 (Service Screening) (%)
Screened	57	63
Invited to screening	43	48
Screened plus nonscreened	21	50

^a Time of diagnosis 1978–1987 or 1988–1996 compared with death rates from cancers diagnosed during 1969–1977 before screening began. All results were statistically significant at 95% confidence level.

Data from Tabar L, Vitak B, Chen HH, et al. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer* 2001;91:1724–31.

estimated that only 12% of the mortality reduction was a result of improved therapy and patient management apart from the screening process. A further study of the 20-year experience in 2 Swedish counties⁵⁰ found a significant 40% reduction in mortality with screening at ages 40 to 49 years.

A recent meta-analysis of 7 European IBM service screening studies⁹⁸ found statistically significant reductions in breast cancer mortality of 25% among women invited to screening and 38% among those screened (Table 5). The 7 nonoverlapping populations and ages screened were from 2 different areas of Finland (50–63),^{101,102} Denmark (60–69),¹⁰³ Italy (50–69),^{104,105} Norway (50–69),¹⁰⁶ Spain (45–69),¹⁰⁷ and Sweden (40–74).¹⁰⁸ Those women from Sweden were from counties other than the 7 counties reported by Duffy and colleagues.¹⁰⁰

Case-Control Studies

These types of studies involve a retrospective comparison of the presence or absence of screening among women who have died of breast cancer (case patients) and living members of the same population (control patients). Control patients have not died of breast cancer, but they may or may not have the disease. The rationale for case-control studies is that if screening reduces breast cancer mortality, then women who die from the disease should be less likely to have a history of screening than matched control patients randomly selected from the population. After the screening history of both case patients and control patients is retrospectively ascertained, reduction in breast cancer mortality as a result of screening can then be estimated.^{6,109}

A recent meta-analysis of results from 7 European case-control service screening studies found a statistically significant 31% breast cancer

Table 5
Summary of results from 7 IBM service screening studies

	RR	95% Confidence Interval	Mortality Reduction (%)
Invited vs not invited	0.75	0.69–0.81	25
Screened vs not screened	0.62	0.56–0.69	38

Data from Broeders M, Moss S, Nystrom L, et al, for the EUROSCREEN Working Group. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *J Med Screen* 2012; 19(Suppl 1):14–25.

mortality reduction among women invited for screening and 48% (after correction for possible self-selection) among those screened (Table 6).⁹⁸ The 7 nonoverlapping populations and ages were Iceland (40–69),¹¹⁰ Italy (50–69),¹¹¹ 3 different regions of Holland (50–74),^{112–116} England (50–64),¹¹⁷ and Wales (50–69).¹¹⁸

Trend Studies

Compared with IBM studies and case-control studies, both of which measure breast cancer deaths among different categories of contemporaneous individuals, trend studies compare breast cancer death rates in an entire population before and after the introduction of screening. There are many reasons why trend studies are less reliable than IBM and case-control studies to ascertain the benefit from screening. Results from trend studies are influenced by confounding factors such as changes in breast cancer incidence or treatment methods over time. Trend studies are unable to exclude deaths from breast cancers that were diagnosed before the introduction of screening or at an age younger than the screening age range.

Among 17 trend studies evaluated by the EUROSCREEN Working Group, 5 were descriptive and 12 were quantitative. Six of these 12 studies, had long enough follow-up for meaningful results. Among these 6 studies, 3 compared mortality before and after the introduction of screening. These studies screened women between ages 50 and 69 years in Florence, Italy,^{119,120} ages 45 and 65 years, in Navarre, Spain,¹⁰⁷ and ages 50 and 64 years in England.¹²¹ Estimated total reductions in breast cancer mortality were 30%, 36%, and 28%, respectively. The other studies included women ages 50 to 69 in Copenhagen, Denmark,¹²²

ages 55 to 74 in Holland,¹²³ and ages 45 to 65 in Navarre, Spain.^{107,124} The average mortality reductions per year (difference in annual mortality prescreening vs during screening/number of screening years) in these studies were 1%, 2.3% to 2.8%, and 9%, respectively.⁹⁸

Results from these many service screening studies indicate that the reductions in breast cancer mortality found in the RCTs can be obtained and exceeded in nonresearch, organized service screening settings. These programs effectively refute the claim by Gotzsche and Olsen that the benefits seen in the RCTs of screening were not real because of supposed flaws in randomization and ascertainment of cause of death.^{86,87}

Although there has not been any service screening study in the United States, screening mammography is commonly performed. Seventy percent of women age 40 years and older report having had a mammogram in the past 2 years, and 55% in the past year.^{125,126} As a consequence, the average woman with invasive breast cancer is 39% less likely to die from her disease than was her counterpart in the early 1980s when screening was less common.⁹⁷ Screening has also resulted in a substantial downstaging of breast cancer, enabling more conservative treatment and allowing every current surgical, medical, and radiation treatment to be more effective.

CAN MODERN TREATMENT SUBSTITUTE FOR EARLY DETECTION?

It is disheartening that new artificial controversies are periodically devised to raise unjustified doubts about the value of screening. Kalager and colleagues¹⁰⁶ reported on the short-term follow-up of Norwegian women aged 50 to 69 years who were offered screening every 2 years along with current (1996–2005) treatment of detected cancers. These women had a 28% mortality reduction compared with historical (1986–1995) breast cancer death rates. Women not offered screening but who received identical current treatment showed an 18% breast cancer mortality reduction compared with the same historical death rates. The investigators concluded that screening accounted for only a 10% mortality reduction (28%–18%) or about one-third (10/28) of the assumed benefit from screen-detected cancers treated with older methods. Thus, the investigators surmised that about two-thirds of the benefit seen in studies such as the Swedish Two-County trial, conducted in the 1980s, could be attained in the absence of screening by means of advances in chemotherapy, surgery, and radiotherapy.

Table 6
Summary of results from 7 European case-control service screening studies

	RR	95% Confidence Interval	Mortality Reduction (%)
Invited vs not invited	0.69	0.57–0.83	31
Screened vs not screened (corrected)	0.52	0.42–0.65	52

Data from Broeders M, Moss S, Nystrom L, et al, for the EUROSCREEN Working Group. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *J Med Screen* 2012; 19(Suppl 1):14–25.

As expected in our media-driven society, this study by Kalager and colleagues was quickly publicized in newspaper headlines and television news shows, a flagrant example of the power of sound bites based on a superficial analysis that failed to recognize the faulty methodology of the study.

The main flaw in Kalager and colleagues' study is the mean follow-up period of 2.2 years, which is insufficient to support their conclusions. This extremely short follow-up would be more than adequate for evaluating the treatment of an acute disease, such as the effect of antibiotic agents on pneumonia, but it is inadequate for studying the effect of early detection or treatment on a chronic disease such as breast cancer. No screening mammography trial has ever shown any significant separation of mortality curves between study and control groups in a period shorter than 4 to 5 years after the initiation of screening. Because the maximum length of follow-up for the Norwegian women was 8.9 years, Kalager and colleagues could have analyzed results for the subset of the population who had a mean follow-up of 4 to 5 years, but they did not report any such results.

Moreover, the mean follow-up period was shorter than 2.2 years. Screening in each Norwegian county was implemented gradually over a 2-year period. However, the investigators considered the beginning of the follow-up period as the date that countrywide screening began rather than the date that full or even partial screening was implemented. This gradual introduction of screening in a county also allowed late-stage cancers among screening group women not yet offered screening to be included in the study group merely because of their residence in the county. This problem could have been ameliorated by offering all women in the study group more screening rounds so that any pollution of the screening group by cancers detected either before the patient's screening or only shortly before reaching the clinical threshold would be minimal.

The Kalager and colleagues study did not include any follow-up data after 2005, even although it was published in late 2010. By that time, the Norwegian Cancer Registry could have provided follow-up data to 2008. Inexplicably, the investigators did not avail themselves of that opportunity for a larger, more robust follow-up sample.

Another weakness in the Norwegian study is that only 77% of women offered screening accepted the invitation to be screened, although presumably, 100% of patients with cancer accepted some form of advanced treatment. A related limitation of this study is that an unknown number of women in the control group obtained screening on their own, outside the program.

Although the investigators are unsure how often this screening happened, the comparatively low 12% breast cancer mortality in the control group suggests that such control group contamination may have been significant.

The study by Kalager and colleagues,¹⁰⁶ along with the accompanying editorial by Welch, mistakenly argues that substantial mortality reduction found in earlier mammography service screening studies conducted throughout Scandinavia is no longer valid, because mortality reduction was calculated from historical data from the prescreening era, with insufficient adjustment for advances in treatment.¹²⁷

Shortly after publication of their study, the hypothesis of Kalager and colleagues that the gains from screening mammography could largely be duplicated by advances in treatment was disproved by a subsequent study by Hellquist and colleagues,¹²⁸ which reported results from service screening in Sweden. Among the Swedish regions that implemented service screening, about half screened all women aged 40 to 59 years, whereas the remaining areas, because of financial constraints, could offer screening only to those aged 50 to 59 years. This practice allowed a natural experiment, because all regions had access to the same modern treatments through the Swedish national health care system, whereas only some regions screened women younger than 50 years. The investigators found that among women aged 40 to 49 years, death rates for breast cancer were reduced 26% for those offered screening and 29% for those screened. Thus, screening was found to confer a substantial benefit independent of treatment. In contrast to the inadequate 2.2-year follow-up period in the study by Kalager and colleagues from Norway, the Swedish study by Hellquist and colleagues had a follow-up period of more than 16 years.

Review of several other recent service screening studies finds other examples of contemporaneous breast cancer mortality for populations with identical access to modern treatment but different availability of screening. For example, in Finland, nationwide population-based breast cancer screening for women aged 50 to 59 years was introduced between 1987 and 1991 on a near-contemporaneous staggered basis. Women born in even years began screening in 1987 or 1988. Women born in odd years, who began screening in 1989 and 1991, served as controls. An effect of screening emerged in 3 to 4 years of follow-up and rapidly diluted as controls were screened. For this narrow window of time, Hakama and colleagues¹⁰² found that mortality from breast carcinoma was 24% lower among women who were

offered screening and 33% lower among women who were screened.

In another service screening study, Jonsson and colleagues^{129–131} found a 30% decrease in breast cancer mortality among women aged 40 to 74 years in 2 northern Swedish counties screened in the 1990s. Two adjacent counties where screening was not yet offered and that had otherwise identical breast cancer mortality served as controls.

THE OVERDIAGNOSIS CONTROVERSY

Overdiagnosis refers to the possibility that some cancers detected at screening might not result in death if they had been undetected, and hence, untreated. Presumably, such cancers grow slowly or not at all, metastasize infrequently, or are found in women who are very old or have substantial comorbid conditions, making the primary cause of death something other than breast cancer. The clinical significance of overdiagnosis is that such women may be subjected to unnecessary biopsy, lumpectomy, mastectomy, chemotherapy, and radiation therapy, as well as anxiety from awareness that they have cancer, along with excessive medical costs for them, their families, and society at large.

Overdiagnosis is different from a false-positive biopsy for a lesion that seems suspicious at mammography but is benign on pathologic examination. Overdiagnosis is essentially an epidemiologic rather than a pathologic concept, although such cases might disproportionately be in situ or minimally invasive cancers. The concept of overdiagnosis is related to length-biased sampling, the possibility that slower-growing tumors are more common among screen-detected cancers than among cancers that emerge clinically in an unscreened population. Length-biased sampling, lead time bias, and selection bias are the 3 major reasons why measurement of decreased mortality in RCTs rather than increased survival rates were necessary to provide the initial proof that the benefits from screening are real.

Was there Overdiagnosis in RCTs?

Randomized trials that compare breast cancer mortality for women in study groups offered screening and otherwise comparable control groups preclude any influence of these biases on mortality. Nevertheless, even documentation of mortality reduction, such as the statistically significant 35% mortality reduction among women aged 40 to 74 years at entry into the Swedish Two-County trial, does not prove that overdiagnosis did not occur.⁹ To assess possible overdiagnosis in that trial, Yen and colleagues¹³² found that the

cumulative incidence of breast cancer (invasive and in situ combined) in the study group and the control group were virtually identical on 29-year follow-up of 1 arm of the trial. Similarly, the total number of invasive cancers in each group was the same, as was the total number of in situ cancers in each group. The only indication of possible overdiagnosis was a nonsignificant excess of cancers among study group women aged 70 to 74 years, suggesting that if overdiagnosis occurred, it was mainly confined to the elderly. As desired for the goal of screening, there was a persistent excess of nonadvanced cancers (node negative or <20 mm in size) in the study group and a persistent excess of advanced cancers (node positive or >20 mm or both) in the control group at 29-year follow-up. This study suggests that the level of overdiagnosis from screening mammography is small and is largely confined to older women.

How Frequent was Overdiagnosis in Service Screening Studies?

The major challenge to estimating the frequency of overdiagnosis from direct observation of incidence rates on service screening studies is that screening advances the rate of detection from that encountered in the absence of screening. Thus, the cancer prevalence rate at first screen exceeds the incidence rate in a similar nonscreened population. Subsequent incidence rates on second, third, fourth screens, and so forth approach those of a nonscreened population. After the screening program ends, there is a catch-up point at which the cumulative number of cancers in the screened and nonscreened populations is equal. A second challenge is a possible difference in risk factors between screened and nonscreened women in some service screening studies. Failure to adjust calculations for lead time and risk factors, as well as an insufficient number of follow-up years, explain the widely different estimates for overdiagnosis (0%–54%) among 13 European service screening studies.¹³³ These disparities have been explained by Puliti and colleagues,¹³³ who reported that estimates for overdiagnosis among 10 studies lacking adjustment were mostly in the 30% to 54% range. Estimates from the 6 studies that made adjustment were lower, only 2.8% in the Netherlands,¹³⁴ 4.6% and 1.0% in Italy,^{105,135} 7% in Denmark,¹³⁶ and 10% and 3.3% in England and Wales.^{121,137}

What Length of Follow-Up Is Needed for an Accurate Estimate of Overdiagnosis?

Although the mean sojourn time (duration during which a nonpalpable cancer is potentially

detectable by screening) is approximately 2.85 years for women aged 40 to 74 years, the distribution of lead time (time between detection and palpability) is exponential and has a wide range, as shown in the Swedish Two-County trial.¹³⁸ From this range of distribution, Duffy and Parmar¹³⁹ have calculated that the number of excess cancers in a hypothetical screening population compared with a nonscreened population does not decrease less than 10% until after 25 years of follow-up, including 10 years after the upper age limit of screening. Applying these criteria for the length of follow-up needed to accurately assess the frequency of overdiagnosis explains why virtually all estimates for overdiagnosis have been too high and must be reduced. For example, Kalager and colleagues¹⁴⁰ estimated an overdiagnosis rate of 18% at 10 years of follow-up of the Norwegian services screening program. However, Duffy and Parmar¹³⁹ indicated that a 19% excess cancer rate would be expected from lead time alone.

Use of Trend Studies to Estimate Overdiagnosis

Analyses that crudely assess overdiagnosis from an increase in cancer incidence with the introduction of screening tend to estimate implausibly high rates of overdiagnosis, of the order of 30% to 50%. These estimates are flawed, because they do not fully take account of numerous complicating factors in incidence of breast cancer and its interface with screening. Reasons for increased incidence observed with the introduction of screening include:

1. Continuation of preexisting trends of increasing breast cancer incidence in the late twentieth century
2. Acceleration of these trends by lead time: if screening confers an average lead time of 2 years, for example, 2004 incidence is observed in 2002, 2005 incidence in 2003, and so on
3. Lead time similarly accelerates age effects: age 52 years incidence is observed at age 50 years, and so on
4. A major surge of additional cancers, mostly as a result of lead time, is diagnosed with the prevalence screen when the screening program is introduced
5. A similar surge of prevalence screen tumors occurs continuously as new patients enter the screening program
6. Overdiagnosis is a minor reason for the excess incidence observed at screening

Compared with direct observation of cumulative breast cancer detection rates among screened versus nonscreened women, temporal comparison of cancer incidence rates in the population represents the most unreliable method to gauge the frequency of overdiagnosis. For example, Bleyer and Welch¹⁴¹ estimated that as a result of screening mammography, 31% of all breast cancers in the United States are being overdiagnosed. These investigators found that breast cancer incidence during 2006 to 2008, when 60% of women age 40 years and older were screened at least once every 2 years, was 31% higher than predicted, assuming a 0.25% per year increase in breast cancer incidence from a 1976 to 1978 baseline, when little screening was being performed.¹⁴¹ A likely flaw in these investigators' method of calculation was that they ignored the fact that over a longer 40-year period (1940–1980), the increase in breast cancer incidence had been 4 times greater, that is, an increase of 1% per year instead of 0.25% per year.¹⁴² As indicated by Kopans,¹⁴³ breast cancer incidence during 2006 to 2008 (128 invasive cancers per 100,000 women) was less than expected (132 cases per 100,000 women per year) from a 1% per year increase. The claim that 31% of breast cancers in the United States are being overdiagnosed was uncritically publicized by the national media and medical journals. Reading these reports can dissuade women from being screened and even influence policies of health care organizations. This episode shows why estimation of overdiagnosis using inappropriate methods of assessment can be problematic.

In another widely publicized study, Esserman and colleagues¹⁴⁴ observed the following: (1) after the widespread introduction of screening in the United States, breast cancer incidence has never returned to prescreening levels, and (2) screening has not caused the relative incidence of regional cancer to decrease commensurate with the increased detection of early-stage cancer. Esserman and colleagues acknowledged that breast cancer mortality has decreased during the screening era in our country but expressed uncertainty about the relative contributions of screening versus treatment to this accomplishment. These investigators too are convinced that there is excessive diagnosis of slow-growing or biologically inert cancers. My critique of their viewpoint is as follows: (1) the increased incidence of breast cancer over the past 30 years is largely related to factors other than screening and (2) the proof of mortality reduction through screening has been established in randomized trials and service screening studies, and proof cannot be negated

by distracting observations regarding the higher than expected incidence of regional disease, which is influenced by factors such as overall breast cancer incidence, excessively long screening intervals, tumor growth rates, and inadequate screening compliance rates. It is likely that a disproportionate number of late-stage cancers cited by the investigators occurred in women who were not being screened.

IS DUCTAL CARCINOMA IN SITU A REAL CANCER?

The concept of overdiagnosis has just been discussed. However, detection of ductal carcinoma in situ (DCIS) is worth considering in more detail. Coincident with the increasing use of mammography, there has been a marked increase in the detection of DCIS. Before the era of mammographic screening, DCIS represented less than 5% of all malignancies of the breast.¹⁴⁵ However, DCIS now accounts for between 20% and 40% of all nonpalpable cancers detected at screening and about 20% of all newly diagnosed cancers (screen detected and non-screen detected) in the United States.¹⁴⁵ With appropriate treatment, the survival rate for patients with DCIS should be 99.5%. DCIS may be considered a frequent but nonobligate precursor of fatal breast cancer. All cases of invasive ductal carcinoma are believed to develop from DCIS, but not all cases of DCIS progress to invasive ductal carcinoma. Yet, critics of screening have referred to DCIS as a pseudo-cancer, false-positive, resulting in harm from screening by leading to unnecessary biopsies and excessive surgery.

Justification for the use of DCIS as an index of the benefit of screening depends on how often and how rapidly DCIS evolves into invasive ductal carcinoma. No direct method exists for determining the natural progression of DCIS. If patients with DCIS were never to undergo biopsy and the DCIS were left to develop into invasive carcinoma, there would be no way to establish that the initial lesion was DCIS. If DCIS is completely excised, its natural history has been stopped, but there is no proof that it would have evolved into invasive ductal carcinoma.

Results from autopsy studies of women with no clinical evidence of breast cancer show 6% to 14% prevalence of DCIS.¹⁴⁵ These rates have been used to suggest that most cases of DCIS may never become clinically apparent. However, there are reasons why this conclusion is not justified. First, most (45%–56%) of the autopsy-detected cases of DCIS could not be identified by radiography performed on the surgical

specimens. An even higher percentage would not have been seen at mammography. If autopsy rates vary around 10% but mammography-detected rates are considerably lower than 1%, it is clear that mammographically detectable DCIS in living women is either a different entity from autopsy-detectable DCIS in dead women, or a special subgroup thereof. The DCIS found at autopsy is not representative of the type of DCIS detected by screening mammography, which would be larger, calcified, and, therefore, a faster-growing lesion. Second, detection rates for invasive ductal carcinoma at prevalence screening (a woman's first screening mammography) are 2 to 3 times higher than the expected incidence, consistent with a 2-year to 3-year detection lead time. In the absence of screening, many cases of high-grade DCIS do not surface clinically as DCIS but rather as invasive carcinoma. Thus, it would not be surprising if the prevalence of mammographically visible DCIS at autopsy were even 10 to 20 times higher than the expected incidence of DCIS. Moreover, some cases classified as DCIS in the autopsy studies, which took place in the 1980s, would be reclassified as atypical hyperplasia according to current histology criteria.

Several follow-up studies of DCIS treated with biopsy alone also shed light on the invasive potential of DCIS. The lesions in these studies were categorized as benign at initial histologic review, and so wide excision was not performed. In 1 study, researchers found development of invasive ductal carcinoma at the biopsy site in 53% of cases within 9.7 years.¹⁴⁵ Another study showed development of invasive ductal carcinoma in 28% of cases by 10 years and 36% cases within 24 years.¹⁴⁵ Recurrence rates for DCIS in series such as these have suggested to some observers that DCIS is unlikely to progress to invasive disease.

However, there are 2 reasons why these studies should lead to just the opposite conclusion. First, these studies underestimate the invasive potential of DCIS, because they involved only cases of low-grade DCIS, that is, all histologic subtypes of DCIS except for comedocarcinoma, the most aggressive subtype. Comedocarcinoma typically accounts for 32% to 50% of all cases of DCIS detected at mammographic screening.^{146–148} Second, these studies include both cases in which the DCIS lesion was completely removed and cases in which some DCIS remained in the breast, because biopsy margins were not sufficiently wide. Invasive ductal carcinoma is expected only in this latter subgroup.

Based on a statistical model using the numbers of DCIS and invasive cancers detected at

5 different screening programs, Yen and colleagues¹⁴⁹ estimated that among cases of DCIS detected at prevalence (initial) screening, 63% were progressive and 37% were nonprogressive. At incidence (subsequent) screenings, 96% of detected cases of DCIS were progressive and only 9% were nonprogressive.¹⁴⁹

Among all cases of DCIS detected at the UK National Health Service Screening Programme,¹⁵⁰ 60% were high grade, 20% intermediate grade, and 20% low grade. It was estimated that 84% of high-grade DCIS would, if undetected, progress to invasive disease in 5 years, most intermediate-grade would progress to invasive disease in 10 years, and low-grade could become invasive in 15 years or longer.

Because screening detects cancers years earlier than they would normally appear clinically, the incidence of invasive cancers is lower than expected for several years after women cease participating in a screening program of limited duration. From observed and expected incidence rates for invasive cancer after cessation of screening in the United Kingdom, McCann and colleagues¹⁵¹ estimated that 75% of this subsequent decreased incidence and lower mortality was caused by screen-detected invasive cancers and 25% was from screen-detected DCIS. These investigators concluded that "Cancer for cancer, there is as much benefit from detection and treatment of DCIS as from detection and treatment of invasive cancer."

HOW FREQUENTLY SHOULD WOMEN BE SCREENED?

No randomized trial has been designed to compare the relative benefit from different screening intervals. Comparison of mortality reduction among randomized trials having different screening intervals has not been meaningful, because of their confounding factors, such as differences in trial design, screened populations, mammographic technique, and interpretation. Moreover, no single randomized trial has had arms that differed only in length of time between screening.

Much of the evidence for selection of screening intervals has been derived from calculation of tumor sojourn time (duration of preclinical disease).¹⁰⁹ Sojourn time begins when the tumor is potentially detectable by mammography and ends at the start of the clinical phase, which begins when the tumor is found clinically in the absence of screening.

Lead time (mean sojourn time) is the average length of time from the point at which the tumor

is detected by mammography to the point at which the clinical phase begins.¹⁰⁹ The lead time is always shorter than sojourn time but approaches the length of sojourn time as screening frequency is increased. Tabar and colleagues⁸ estimated that the lead time was 1.7 years for women aged 40 to 49 years and 2.6 to 3.8 years for women aged 50 to 74 years. Effective screening requires that the screening interval should be shorter than the lead time.⁷⁵⁻⁷⁷

Mathematical Models

Using a computer simulation model based on observed breast cancer growth rates, Michaelson and colleagues⁷⁸ estimated that triennial, biennial, annual, and biannual screening would reduce distant metastases by 19%, 27%, 52%, and 81%, respectively. These investigators found good correlation between screening frequency tumor size and survival. These findings suggest that annual screening could reduce breast cancer death rates by 50% and that screening every 6 months might reduce death rates by 82%.⁷⁸

Using a Markov chain model for progression of breast cancer, based on results from the Swedish Two-County trial, Duffy and colleagues¹⁵² and Chen and colleagues^{153,154} estimated that for women aged 40 to 49 years, screening every year, every other year, or every third year might result in reductions in mortality of 36%, 18%, and 4%, respectively. For women aged 50 to 59 years, the same screening intervals might result in reductions of 46%, 39%, and 34%, respectively.¹⁵²⁻¹⁵⁴ Thus, older women achieve greater mortality reduction than younger women screened at the same interval. However, older women receive less incremental increase in benefit for progressively shorter screening intervals.

Several investigators have estimated the increased benefits from annual screening versus the observed benefits from screening every 24 months in the Swedish Two-County trial and every 18 months in the Gothenburg, Sweden trial. Such calculations were made from the stage and expected death rates of interval cancers surfacing clinically between screens. In the Two-County trial, the observed mortality reduction was 24% for women aged 40 to 49 years and 39% for those aged 50 to 59 years. Using such data, Tabar and colleagues and Feig, in 2 separate reports, estimated mortality reductions of 35% and 46% for annual screening of these 2 respective age groups.^{46,79} In Gothenburg, Bjurstam and colleagues^{18,19} found a 45% mortality reduction observed from screening women aged 39 to 49 years every 18 months (80% compliance).

Using data from that trial, Feig estimated that annual screening of these women would have resulted in a mortality reduction between 65% (at 80% compliance) and 75% (at 100% compliance).⁸⁰

Clinical Observational Studies

Several clinical follow-up studies have shown that women aged 40 to 49 years who chose to be screened annually were more likely to be diagnosed with early-stage versus late-stage breast cancer. The Breast Cancer Surveillance Consortium (BCSC) is a large group of breast imaging facilities throughout the United States that are linked to regional tumor registries. Using data from the BCSC, White and colleagues¹⁵⁵ and Kerlikowske and colleagues¹⁵⁶ found an increase in late-stage disease among women aged 40 to 49 years screened with a 2-year versus those screened with a 1-year interval. The value of annual versus biennial screening in reducing the likelihood of late-stage breast cancer among women younger than 50 years was also shown in a study of women in Wisconsin by Ontilo and colleagues.¹⁵⁷

Hunt and colleagues¹⁵⁸ found that women aged 40 to 79 years screened annually at the University of California at San Francisco had invasive cancers that were smaller and lower stage than those among women screened biennially. About 60% of the screeners were aged 50 years or older.

At variance with these conclusions are findings from a 12.8-year follow-up of mortality of women aged 40 to 49 years selected for screening either annually (even year-of-birth cohorts) or triennially (odd year-of-birth cohorts) in Turku, Finland.¹⁵⁹ No differences in incidence based mortality according to screening frequency were found. However, because of lack of a control group with no screening, it cannot be determined whether this result was caused by a low efficacy of mammography in their study, or insufficient follow-up, or lack of incremental benefit for annual versus triennial screening.

Clinical studies of screening frequency for women aged 50 years and older have varied in their conclusions. Using the BCSC database, White and colleagues¹⁵⁵ found that screening at 2-year versus 1-year intervals did not increase the likelihood of late-stage disease at diagnosis but did increase the likelihood of being diagnosed with invasive disease versus DCIS. Using the same database, the lack of effect of screening frequency on late-stage presentation among older women was confirmed by Kerlikowske and colleagues.¹⁵⁶

It has been observed that weaknesses in the design of both studies may nullify their conclusion that screening every 2 years is sufficient for older women. One criticism regards the definition of late-stage breast cancer: as either positive lymph nodes or metastases by White and colleagues,¹⁵⁵ or as stage 2b or higher or 2 cm or larger by Kerlikowske and colleagues.¹⁵⁶ These cut points may be too high to document benefit from annual screening. Cancers detected by screening mammography are usually stage 1b or earlier. Selection of lower cut points may have led to a different conclusion regarding the optimal screening frequency. It seems logical that reduction of stage 2a or stage 2b to stage 1 or stage 1 to stage 0 should reduce mortality.

Another limitation of these studies is that screening intervals were chosen by the patients or suggested by their physicians rather than made at random. Thus, intervals were subject to selection bias. It is unclear why some women were screened less frequently.

The finding that screening frequency does not influence stage at detection is implausible unless one believes that most breast cancers cease to grow after the patient is aged 50 years. As radiologists, we have observed some cancers, including those presenting as spiculated masses or calcifications, that do not grow over periods of 1 to 4 years.^{160,161} However, such cases are exceptional and worthy of reports in the radiologic literature.¹⁶²

Other clinical studies of older women found that those screened annually do have earlier disease than those screened less often. Field and colleagues¹⁶³ compared 1-year versus 2-year screening intervals among Michigan women aged 65 years and older. Those screened annually had smaller invasive tumors and a downstaging of their invasion disease and 3 times as many cases of DCIS. Ontilo and colleagues¹⁵⁷ found that women aged 50 years and older in Wisconsin screened at 2-year intervals had a 3 times greater likelihood of stage III or IV breast cancer than those screened at annual intervals. Late-stage breast cancers were also high among the group who had been screened at intervals of 3 years or longer.

USPSTF Controversy

In November, 2009, the USPSTF recommended against mammographic screening for women aged 40 to 49 years except for those at high risk; it recommended screening only every other year rather than annual screening for those aged 50 to 74 years; and it recommended no screening for

women 75 years and older.^{164,165} The recommendation to screen women in their 40s only if they were at high risk was controversial, because 80% of women with newly diagnosed breast cancers have no significant previous risk factors.^{166,167} USPSTF guidelines differ substantially from those of the ACS, American College of Radiology (ACR), Society of Breast Imaging, American College of Obstetricians and Gynecologists, and National Comprehensive Cancer Network, all of which recommend annual screening mammography beginning at age 40 years.^{73,168,169}

The USPSTF is a government-supported group of health experts who review published research and make recommendations about preventive health care issues such as screening for carcinoma of the breast, cervix, prostate, and colon. Many members are PhDs and of the few MD members, few are primarily involved in clinical care. There were no breast imagers or breast surgeons on the panel.

Public reaction to USPSTF was immediate and pronounced. Some women mistakenly believed that USPSTF supplanted older guidelines from the ACS, American College of Surgeons, and ACR. Most women were either outraged or confused. A *USA Today* poll found that 47% strongly disagreed and 29% disagreed, whereas only 5% strongly agreed and 17% agreed with the USPSTF advice.¹⁷⁰ The 2009 USPSTF recommendations were potentially more important than their previous ones issued in 2002, because 1 provision in the then proposed Affordable Healthcare Law forbid Medicare and private insurers from paying for any medical care that did not conform to USPSTF policy. After public outrage, Kathleen Sebelius, Secretary of Health and Human Services, was quick to deny that USPSTF recommendations were government policy.¹⁷¹ However, concern remains that USPSTF recommendations could encourage insurance companies to reduce their screening mammography coverage.

The USPSTF provided several reasons for its recommendations. First, by USPSTF calculations, RCTs found a breast cancer mortality reduction of only 15% for ages 40 to 49 years and 50 to 59 years compared with 32% for ages 60 to 69 years.^{164,165} The reason their calculations derived a lower percentage mortality reduction than those found in the RCTs and service screening studies cited earlier was that the USPSTF included results from the NBSSC, which found no benefit from mammography screening of women aged 40 to 60 years. There are several reasons why inclusion of NBSSC data is unjustified. Unlike all of the other RCTs, NBSSC was not population-based but

rather relied on volunteers. Clinical breast examination was performed by NBSSC staff members before patient allocation to either a study group or control group. As a likely consequence, the trial contained an excessive proportion of symptomatic women, many with advanced palpable cancers, who seem to have been preferentially channeled into the study group rather than the control group.^{28–36} In addition, many of the mammographic examinations were judged as technically inadequate by external expert reviewers, even by the standards of the 1980s when the NBSSC was performed.^{27–29} Technical problems included poor positioning, poor breast compression, faulty processing, and lack of mammographic grids.

Higher screening recall rates and false-positive biopsy rates relative to cancer detection rates among younger women were also used by USPSTF to justify the recommendation against routine screening for women in their 40s.¹⁷² The USPSTF cited current BCSC data from the United States that the numbers of women recalled from screening to undergo additional imaging in order to find 1 cancer at ages 40 to 49 years, 50 to 59 years, and 60 to 69 years were 47, 22, and 14, respectively.^{165,172,173} However, such recall rates are not a valid reason to not screen women in their 40s because more than 90% of all recalls do not result in biopsy and entail nothing more than supplementary mammographic views or ultrasonography.¹ Recall rates are always higher on initial screening than on subsequent screening, in which there are previous mammograms for comparison. Delaying the initiation of screening until age 50 years would only transfer the higher recall rates per detected cancer to that age group. The USPSTF also cited BCSC data that the biopsy positive predictive value (DCIS or invasive cancer per 100 biopsies) was 28% for women aged 40 to 49 years versus 44% for ages 50 to 59 years and 56% for ages 60 to 69 years.¹⁷² The USPSTF neglected to note that 99% of women surveyed considered even 500 or more false-positive mammography examinations to be an acceptable risk to save 1 life and that their anxiety from screening recall and biopsy was slight and transient.^{174,175} Then too, biopsy positive predictive value rates of 28% are well within the 20% to 40% range recommend by the US Agency for Healthcare Policy and Research, ACR, and other consensus panels.¹

Absolute and Relative Benefit

Numerous studies have shown that the relative mortality reduction through screening women

aged 40 to 49 years is 30% to 50%, similar to that from screening older women.^{9,37,50,51} However, absolute benefit (deaths averted per 1000 women screened) rather than relative benefit (percent mortality reduction) must be used to calculate the cost-effectiveness for screening different age groups, because breast cancer incidence varies according to age. For example, the probability of developing breast cancer in the next 10 years for women aged 40, 50, and 60 years is 1.44%, 2.39%, and 3.4%, respectively.¹⁷⁶ The odds of developing breast cancer in these respective decades are 1 in 69, 42, and 29. Thus, compared with older women, younger women obtain a lower absolute benefit from screening. For these reasons, the USPSTF used both absolute benefit and relative benefit to support their screening recommendations.

USPSTF made serious mistakes in their calculations of both relative and absolute benefit. As discussed, their estimated relative mortality reduction of only 15% for screening women aged 40 to 49 years was too low, because they included results from NBSSC. NBSSC was fatally flawed by poor mammographic technique, recruitment of symptomatic women, and the likely channeling of women with breast masses into the mammography arm. As a consequence, NBSSC was the only screening trial to have excess of late-stage breast cancers in the study group versus the control group. NBSSC was the only trial to find no benefit from screening any age group. These investigators' error in estimation of relative benefit led to another error in estimation of absolute benefit.

Using data from NBSSC and other trials, USPSTF calculated that the number of women needed to be invited to screening at ages 39 to 49 years, 50 to 59 years, and 60 to 69 years to prevent 1 death was 1904, 1339, and 377, respectively. However, excluding NBSSC, these results would have been 950, 670, and 377, respectively. These recalculated values would have been sufficient to recommend screening women in their 40s according to USPSTF's own criteria!¹⁷⁶

The second mistake made by USPSTF was selection of the number of women invited to be screened (NNI) rather than the number of women screened (NNS) for their measure of absolute benefit.

The reality is that not all women invited to screening agree to be screened. In some trials, as few as 32% of women have attended all screening rounds.⁷² Thus, NNS is always lower, often considerably lower, than NNI. When reporting USPSTF conclusions, the public media and medical journals often confused NNI with NNS.

This reporting error made screening seem less efficient. Using RCT data, Hendrick and Helvie¹⁷⁷ calculated that NNS at ages 40 to 49 years, 50 to 59 years, and 60 to 69 years were 746, 351, and 253, respectively. These numbers are lower than those for NNI used by USPSTF and thus are more favorable for screening. NNS can be specified in terms of either number of women needed to screen detect 1 cancer, save 1 life, or extend a woman's life by 1 year.

Because of the longer life expectancy of young women, NNS per life year gained (LYG) in these age groups was 28, 17, and 16, respectively. Use of LYG provides even more favorable comparisons of younger versus older women than use of number of deaths prevented.¹⁷⁶ Use of NNS to prevent 1 death or gain 1 life year strongly supports screening all women between ages 40 and 69 years.

Benefits and Costs

Increasing interest in reducing national health care expenditures has led to studies such as a recent one on the aggregate cost for screening mammography by O'Donoghue and colleagues.^{178,179} This study estimates costs for 3 different screening strategies: annual (ages 40–84 years), biennial (ages 50–69 years), and USPSTF (high risk ages 40–49 years, biennial ages 50–74 years). The annual cost for each of these plans was estimated at \$10.1 billion, \$2.6 billion, and \$3.5 billion, respectively. Studies such as this raise concerns that cost more than science is responsible for the screening controversies.

The deficiencies in the USPSTF recommendations have already been discussed in detail. Thus, I confine my critique to strategy 2, limiting screening to women aged 50 to 69 years every 2 years.

1. The investigators do not appreciate that annual screening of women aged 40 to 49 years can reduce breast cancer mortality by 40% to 50%. Moreover, 40% of years of life lost to breast cancer may result from cancer appearing during this decade.⁷
2. Most women aged 70 years and older need to be screened, because most still have substantial average life expectancies.
3. Numerous studies have concluded that annual screening is more effective than biennial screening (perhaps not double the benefit of biennial screening, but substantially more). Even USPSTF projected that for women ages 40–69 annual screening would increase life years gained by 37% more than biennial screening.¹⁷²

4. They did not acknowledge that the costs of screening is partly offset by reduced costs for more intensive treatment and long-term care.²
5. The investigators estimated the cost per LYG for annual versus biennial screening, but did not indicate their assumed level of benefit entered into their calculation. Their final value exceeded the acceptable upper limit. The more conventional measure of cost per LYG through annual screening has been calculated by other investigators and found to be lower than the commonly accepted upper limit of \$50,000 to \$100,000 per LYG.^{2,180}

SUMMARY

Numerous clinical studies have confirmed that screening women aged 40 years and older reduces breast cancer mortality by 30% to 50%. Several factors including faster breast cancer growth rates and lower breast cancer incidence among younger women, as well as shorter life expectancy and more comorbid conditions among older women, should also be considered in screening guidelines. Accordingly, annual screening beginning at age 40 years and continuing with no upper age limit, as long as a woman has a life expectancy of at least 5 years and no significant comorbid conditions, is recommended by the ACS, the ACR, and the Society of Breast Imaging.

Annual screening is more effective than screening every 2 years for women aged 40 to 49 years and probably for those aged 50 years and older as well. However, the benefit from screening every year is less than double that from screening every other year.

Past controversies regarding the effectiveness of screening women in their 40s have largely been resolved. Differences in recommendations for screening women in this age group are now mostly related to the magnitude of benefit, the lower absolute benefit, the lower cost-effectiveness, and the higher false-positive biopsy rates compared with those in older women.

The previous controversy initiated by Gotzsche and Olsen regarding the validity of screening trial results for women of all ages has been essentially resolved as well. Numerous medical organizations in the United States and Europe concur that the data showing mortality reduction are scientifically sound.

The claim of Kalagar and colleagues that modern treatment can largely substitute for early detection has no scientific support, whereas there is strong evidence that mammography

remains indispensable for the substantial decline in breast cancer death rates over the past 30 years.

Most cases of screen-detected DCIS are capable of transformation into invasive disease. There is no reliable way for the radiologist to distinguish these cases from the few cases that represent low-grade DCIS. This is a continuing challenge for the pathologist and surgeon to ensure appropriate treatment.

Overdiagnosis refers to the possibility that some screen-detected cancers may not eventuate in death if undetected by mammography, yet receive unnecessary treatment by surgery, chemotherapy, and radiation therapy. One highly publicized study by Bleyer and Welch claimed that 31% of all breast cancers are misdiagnosed. More reliable calculations by Puliti and colleagues found that the frequency of overdiagnosis is extremely low, between 0% and 5% of all screen-detected cancers.

REFERENCES

1. Feig SA. Adverse effects of screening mammography. *Radiol Clin North Am* 2004;42(5):807–21.
2. Feig SA. Cost-effectiveness of mammography, MRI, and ultrasonography for breast cancer screening. *Radiol Clin North Am* 2010;48(5):879–91.
3. Seidman H, Gelb SK, Silverberg E, et al. Survival experience in the breast cancer detection demonstration project. *CA Cancer J Clin* 1987;37:258–90.
4. Smart CR, Bryne C, Smith RA, et al. Twenty year follow-up of the breast cancers diagnosed during the Breast Cancer Detection Demonstration Project. *CA Cancer J Clin* 1997;47:134–49.
5. Ries L, Eisner MP, Kosary CL, et al. SEER Cancer Statistics Review, 1973-1998. Bethesda (MD): National Cancer Institute; 2001.
6. Feig SA. Methods to identify benefit from mammographic screening. *Radiology* 1996;201:309–16.
7. Shapiro S, Venet W, Strax P, et al. Periodic screening for breast cancer, the health insurance plan project and its sequelae 1963-1976. Baltimore (MD): Johns Hopkins University Press; 1988.
8. Tabar L, Vitak B, Chen JJ, et al. The Swedish Two-County Trial twenty years later. *Radiol Clin North Am* 2000;38:625–52.
9. Tabar L, Vitak B, Chen TH, et al. Swedish Two-County trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology* 2012;260:658–63.
10. Tabar L, Fagerberg CJ, Gad A, et al. Reduction in mortality from breast cancer after mass screening with mammography. Randomized trial from the Breast Cancer Screening Working Group in the

- Swedish National Board of Health and Welfare. *Lancet* 1985;1:829–32.
11. Andersson I, Aspegren K, Janzon L, et al. Mammographic screening and mortality from breast cancer: the Malmo Mammographic Screening Trial. *BMJ* 1988;297:943–8.
 12. Andersson I, Nystrom L. Mammography screening. *J Natl Cancer Inst* 1995;87:1263–4.
 13. Andersson I, Janzon L. Reduced breast cancer mortality in women under 50: updated results from the Malmo Mammographic Screening Program. *J Natl Cancer Inst Monographs* 1997;22:63–8.
 14. Nystrom L, Andersson I, Bjurstam N, et al. Long-term effects of mammography screening: updated overview of the Swedish randomized trials. *Lancet* 2002;359:909–19.
 15. Nystrom L. Screening mammography re-evaluated [letter to the editor]. *Lancet* 2002;355:748–9.
 16. Frisell J, Eklund G, Helstrom L, et al. Randomized study of mammography screening: preliminary report on mortality in the Stockholm trial. *Breast Cancer Res Treat* 1991;18:49–56.
 17. Frisell J, Lidbrink E, Helstrom L, et al. Follow-up after 11 years: update of mortality results in the Stockholm mammographic screening trial. *Breast Cancer Res Treat* 1997;45:263–70.
 18. Bjurstam N, Bjorneld L, Duffy SW, et al. The Gothenburg Breast Screening Trial: first results on mortality, incidence, and mode of detection for women ages 39–49 years at randomization. *Cancer* 1997;80:2091–9.
 19. Bjurstam N, Bjorneld L, Duffy SW, et al. The Gothenburg Breast Screening Trial. *Cancer* 1998;83:188–90 [authors' reply].
 20. Bjurstam J, Bjorneld L, Warwick J, et al. The Gothenburg Breast Screening Trial. *Cancer* 2003;97:2387–96.
 21. Roberts MM, Alexander FE, Anderson TJ, et al. Edinburgh trial of screening for breast cancer: mortality at seven years. *Lancet* 1990;335:241–6.
 22. Alexander FE, Anderson TJ, Brown HK, et al. 14 years of follow-up from Edinburgh randomized trial of breast cancer screening. *Lancet* 1999;353:1903–8.
 23. Moss SM, Cuckle H, Evans A, et al. Effect of mammographic screening from age 40–49 years on breast cancer mortality at 10 years follow-up: a randomized controlled trial. *Lancet* 2006;368:2053–80.
 24. Miller AB, Baines CJ, To T, et al. Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50–59 years. *Can Med Assoc J* 1992;147:1477–88 [Erratum appears in *Can Med Assoc J* 1993;148:718].
 25. Miller AB, To T, Baines CJ, et al. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50–59 years. *J Natl Cancer Inst* 2000;92:1490–9.
 26. Miller AB, To T, Baines CJ, et al. The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up: a randomized screening trial of mammography in women age 40 to 49 years. *Ann Intern Med* 2002;137:305–12.
 27. Baines CJ, Miller AB, Kopans DB, et al. Canadian National Breast Screening Study: assessment of technical quality by external review. *AJR Am J Roentgenol* 1990;155:743–7.
 28. Kopans DB. The Canadian Screening Program: a different perspective. *AJR Am J Roentgenol* 1990;155:748–9.
 29. Kopans DB, Feig SA. The Canadian National Breast Screening Study: a critical review. *AJR Am J Roentgenol* 1993;161:755–60.
 30. Warren-Burhenne LJ, Burhenne HJ. The Canadian National Breast Screening Study: a Canadian critique. *AJR Am J Roentgenol* 1993;161:761–3.
 31. Bailar JC, MacMahon B. Randomization in the Canadian National Breast Screening Study: a review of evidence for subversion. *Can Med Assoc J* 1997;156:193–9.
 32. Boyd NE, Jong RA, Yaffe MJ, et al. A critical appraisal of the Canadian National Breast Screening Study. *Radiology* 1993;189:661–3.
 33. Boyd NF. The review of randomization in the Canadian National Breast Screening Study: is the debate over? *Can Med Assoc J* 1997;156:207–9.
 34. Sun J, Chapman J, Gordon R. Survival from primary breast cancer after routine clinical use of mammography. *Breast J* 2002;8:199–208.
 35. Mettlin CJ, Smart CR. The Canadian National Breast Screening Study: an appraisal and implications for early detection policy. *Cancer* 1993;72:1461–5.
 36. Tarone RE. The excess of patients with advanced breast cancer in young women screened with mammography in the Canadian National Breast Screening Study. *Cancer* 1995;75:997–1003.
 37. Smith RA, Duffy SW, Gabe R, et al. The randomized trials of breast cancer screening: what have we learned? *Radiol Clin North Am* 2004;42:793–806.
 38. Fletcher SW, Black W, Harris R, et al. Report of the International Workshop on Screening for Breast Cancer. *J Natl Cancer Inst* 1993;85:1644–56.
 39. Smith RA. Breast cancer screening among women younger than age 50: a current assessment of the issues. *CA Cancer J Clin* 2000;50:312–36.
 40. Hurley SF, Kaldor JM. The benefits and risks of mammographic screening for breast cancer. *Epidemiol Rev* 1992;14:101–30.
 41. 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* 1998;80:1125–32.

42. Elwood JM, Cox B, Richardson AK. The effectiveness of breast cancer screening by mammography in younger women. *Online J Curr Clin Trials* 1993. February 25, Doc No 32.
43. Glaszion PP, Woodward AJ, Mahon CM. Mammographic screening trials for women under age 50: a quality assessment and meta-analysis. *Med J Aust* 1995;162:625–9.
44. Kerlikowske K, Grady D, Rubin SM, et al. Efficacy of screening mammography: a meta-analysis. *JAMA* 1995;273:149–54.
45. Smart CR, Hendrick RE, Rutledge JH, et al. Benefit of mammography screening in women ages 40 to 49 years: current evidence from randomized controlled trials. *Cancer* 1995;75:1619–26 [Erratum appears in *Cancer* 75:2788].
46. Falun Meeting Committee and Collaborators. Breast-cancer screening with mammography in women aged 40–49 years. Swedish Cancer Society and the Swedish National Board of Health and Welfare. *Int J Cancer* 1996;68:693–9.
47. Hendrick RE, Smith RA, Rutledge JH, et al. Benefit of screening mammography in women aged 40–49: a new meta-analysis of randomized controlled trials. *Natl Cancer Inst Monogr* 1997;33:87–92.
48. Wald NJ, Murphy P, Major P, et al. UKCCCR multicentre randomized controlled trial of one and two view mammography in breast cancer screening. *BMJ* 1995;311:1189–93.
49. Evans AJ, Kutt E, Record C, et al. Radiologic findings of screen-detected cancers in a multicentre randomized, controlled trial of mammographic screening in women from age 40 to 48 years. *Clin Radiol* 2006;61:784–8.
50. Tabar L, Yen MF, Vitak B, et al. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. *Lancet* 2003;361:1405–10.
51. Coldman A, Phillips N, Warren L, et al. Breast cancer mortality after screening mammography in British Columbia women. *Int J Cancer* 2007;120:1076–80.
52. US Bureau of the Census. Statistical abstract of the United States. 212th edition. Washington, DC: US Government Printing Office; 2012.
53. Feig SA. Mammographic screening of elderly women. *JAMA* 1996;276:446.
54. Mandelblatt JS, Wheat ME, Monane M, et al. Breast cancer screening for elderly women with and without comorbid conditions. *Ann Intern Med* 1992;116:722–30.
55. Yancik R, Reis LG, Yates JW. Breast cancer in women: a population based study of contrasts in stage, survival, and surgery. *Cancer* 1989;163:976–81.
56. Faulk RM, Sickles EA, Solitto RA, et al. Clinical efficacy of mammographic screening in the elderly. *Radiology* 1995;194:193–7.
57. Wilson TE, Helvie MA, August DA. Breast cancer in the elderly patient: early detection with mammography. *Radiology* 1994;190:203–7.
58. Constanza ME. Issues in breast cancer screening in older women. *Cancer* 1994;74:2009–15.
59. Walter LC, Covinsky KE. Cancer screening in elderly patients: a frame-work for individual decision making. *JAMA* 2001;285:2750–6.
60. Conway BJ, Suleiman OH, Reuter FG, et al. National survey of mammographic facilities in 1985, 1988, and 1992. *Radiology* 1994;191:323–30.
61. Haus AG. Dedicated mammography x-ray equipment, screen-film processing-systems, and viewing conditions for mammography. *Semin Breast Dis* 1999;2:30–54.
62. Feig SA. Screening mammography: effect of image quality on clinical outcome. *AJR Am J Roentgenol* 2002;178:803–7.
63. Young K, Wallis MG, Ramsdale ML. Mammographic film density and detection of small breast cancers. *Clin Radiol* 1994;49:461–5.
64. Glynn CE, Farria DM, Monsees BS, et al. Effect of transition to digital mammography on clinical outcomes. *Radiology* 2011;200(3):664–70.
65. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast cancer screening. *N Engl J Med* 2005;363:1773–83.
66. Andersson I, Hildell J, Muhlow A, et al. Number of projections in mammography: influence on detection of breast disease. *AJR Am J Roentgenol* 1978;130:349–51.
67. Anttinen I, Pamiilo M, Roiha M, et al. Baseline screening mammography with one versus two views. *Eur J Radiol* 1989;9:241–3.
68. Bassett LW, Bunnell DH, Jahanshahi R, et al. Breast cancer detection: one versus two views. *Radiology* 1987;165:95–7.
69. Muir BB, Kirkpatrick AE, Roberts MM, et al. Oblique-view mammography adequacy for screening. *Radiology* 1984;151:39–41.
70. Sickles EA, Weber WN, Galvin HB, et al. Baseline screening mammography: one vs two views per breast. *AJR Am J Roentgenol* 1986;147:1149–53.
71. Thurfjell G, Taube A, Tabar L. One-versus two-view mammography screening: a prospective population based study. *Acta Radiol* 1994;35:340–4.
72. Humphrey LL, Helfant M, Chan BK, et al. Breast cancer screening: a summary of the evidence for the US Preventive Services Task Force. *Ann Intern Med* 2002;137:347–60.
73. Smith RA, Saslow D, Sawyer KA, et al. American Cancer Society Guidelines for Breast Cancer Screening: update 2003. *CA Cancer J Clin* 2003; 53:141–69.
74. Feig SA. Determination of mammographic screening intervals with surrogate measures for

- women aged 40-49 years. *Radiology* 1994;193:311-4.
75. Moskowitz M. Breast cancer: age specific growth rates and screening strategies. *Radiology* 1986;161:37-41.
 76. Pelikan S, Moskowitz M. Effects of lead-time, length bias, and false-negative reassurance on screening for breast cancer. *Cancer* 1993;71:1998-2005.
 77. Tabar L, Fagerberg G, Day NE, et al. What is the optimum interval between screening examination? An analysis based on the latest results of the Swedish Two-County Breast Cancer Screening trial. *Br J Cancer* 1987;55:47-51.
 78. Michaelson JS, Halpern E, Kopans DB. Breast cancer computer simulation method for estimation of optimal intervals for screening. *Radiology* 1999;212:551-60.
 79. Feig SA. Estimation of currently attainable benefit from mammographic screening of women aged 40-49 years. *Cancer* 1995;75:2412-9.
 80. Feig SA. Increased benefit from shorter screening mammography intervals for women ages 40-49 years. *Cancer* 1997;80:2035-9.
 81. Tabar L, Fagerberg G, Chen HH. Efficacy of breast cancer screening by age: new results from the Swedish Two-County Trial. *Cancer* 1995;75:2507-17.
 82. Miettinen OS, Henschke CI, Pasmantier MW, et al. Mammographic screening: no reliable supporting evidence? *Lancet* 2002;359:404-6.
 83. Gotzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet* 2000;355:129-34.
 84. Olsen O, Gotzsche PC. Cochrane review on screening for breast cancer with mammography. *Lancet* 2001;358:1340-2.
 85. de Koning HJ. Assessment of nationwide cancer-screening programmes. *Lancet* 2000;355:80-1.
 86. Duffy SW. Interpretation of the breast screening trials: a commentary on the recent paper by Gotzsche and Olsen. *Breast* 2001;10:209-12.
 87. Feig SA. How reliable is the evidence for screening mammography? *Recent Results Cancer Res* 2003;163:129-39.
 88. 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* 1989;43:107-14.
 89. Nystrom L, Larsson LG, Rutqvist LE, et al. Determination of cause of death among breast cancer cases in the Swedish randomized mammography screening trials. A comparison between official statistics and validation by an endpoint committee. *Acta Oncol* 1995;34:145-52.
 90. Tabar L, Duffy SW, Yen MF, et al. All cause mortality among breast cancer patients in a screening trial: support for breast cancer mortality as an end point. *J Med Screen* 2002;9:159-62.
 91. Sjonell G, Stahle L. Mammography screening does not reduce breast cancer mortality. *Lakartidningen* 1999;96:904-13 [in Swedish].
 92. Swedish Board of Health and Welfare. Vilka Effekter Har Mammografic screening?. Referat av ett expertmote anordnat av Socialstyrelsen och ancerfonden I; Stockholm den 15 Februari 2002.
 93. Rosen M, Rehnqvist N. No need to reconsider breast screening programme on basis of results from defective study [letter to the editor]. *BMJ* 1999;318:809-10.
 94. Health Council of the Netherlands. The benefit of population screening for breast cancer with mammography. The Hague (The Netherlands): Health Council of the Netherlands; 2002.
 95. Veronisi U, Forrest P, Wood W. Statement from the chair: global Summit on Mammographic Screening. Milan (Italy): European Institute of Oncology; 2002.
 96. International Agency for Research on Cancer. Mammography screening can reduce deaths from breast cancer. Lyon (France): IARC Press; 2002.
 97. Feig SA. Effect of service screening mammography on population mortality from breast carcinoma. *Cancer* 2002;95:451-7.
 98. Broeders M, Moss S, Nystrom L, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *J Med Screen* 2012;19(Suppl 1):14-25.
 99. Tabar L, Vitak B, Chen HH, et al. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer* 2001;91:1724-31.
 100. Duffy SW, Tabar L, Chen HH, et al. The impact of organized mammography service screening on breast cancer mortality in seven Swedish counties: a collaborative evaluation. *Cancer* 2002;95:458-69.
 101. Sarkeala T, Heinavaara S, Anttila A. Breast cancer mortality with varying invitational policies in organized mammography. *Br J Cancer* 2008;98:641-5.
 102. Hakama M, Pukkala E, Heikkila M, et al. Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. *BMJ* 1997;314:864-7.
 103. Olsen AH, Njor SH, Vejborg I, et al. Breast cancer mortality in Copenhagen after the introduction of mammography screening: a cohort study. *BMJ* 2005;330:220.
 104. Paci E, Giorgi D, Bianchi S, et al. Assessment of the early impact of the population-based breast cancer screening programme in Florence (Italy) using mortality and surrogate measures. *Eur J Cancer* 2002;38:568-73.

105. Puliti D, Zappa M, Miccinesi G, et al. An estimate of overdiagnosis 15 years after the start of mammographic screening in Florence. *Eur J Cancer* 2009;45:3166–71.
106. Kalager M, Zelen M, Langmark F, et al. Effect of screening mammography on breast-cancer mortality in Norway. *N Engl J Med* 2010;363:1203–10.
107. Ascunce EN, Moreno-Iribas C, Barcos Urtiaga A, et al. Changes in breast cancer mortality in Navarre (Spain) after introduction of a screening programme. *J Med Screen* 2007;14:14–20.
108. Swedish Organised Service Screening Evaluation Group. Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data. *Cancer Epidemiol Biomarkers Prev* 2006;15:45–51.
109. Morrison AS. Screening for chronic disease. 2nd edition. New York: Oxford University Press; 1992.
110. Gabe R, Tryggvadottir L, Sigfusson BF, et al. A case-control study to estimate the impact of the Icelandic population-based mammography screening program on breast cancer death. *Acta Radiol* 2007;48:948–55.
111. Puliti D, Miccinesi G, Collina N, et al. Effectiveness of service screening: a case-control study to assess breast cancer mortality reduction. *Br J Cancer* 2008;99:423–7.
112. Otto SJ, Fracheboud J, Looman CW, et al. Initiation of population-based mammography screening in Dutch municipalities and effect on breast-cancer mortality: a systematic review. *Lancet* 2003;361:1411–7.
113. Otto SJ, Fracheboud J, Verbeek AL, et al. Mammography screening and risk of breast cancer death: a population-based case-control study. *Cancer Epidemiol Biomarkers Prev* 2012;21:66–73.
114. van Schoor G, Moss SM, Otten JD, et al. Increasingly strong reduction in breast cancer mortality due to screening. *Br J Cancer* 2011;104:910–4.
115. Paap E, Holland R, den Heeten GJ, et al. A remarkable reduction of breast cancer deaths in screened vs unscreened women: a case referent study. *Cancer Causes Control* 2010;21:1569–73.
116. Broeders MJ, Verbeek AL, Straatman H, et al. Repeated mammographic screening reduces breast cancer mortality along the continuum of age. *J Med Screen* 2002;9:163–7.
117. Allgood PC, Warwick J, Warren RM, et al. A case-control study of the impact of the East Anglian breast screening programme on breast cancer mortality. *Br J Cancer* 2008;98:206–9.
118. Fielder HM, Warwick J, Brook D, et al. A case-control study to estimate the impact on breast cancer death of the breast screening programme in Wales. *J Med Screen* 2004;11:194–8.
119. Gorini G, Zappa M, Miccinesi G, et al. Breast cancer mortality trends in two areas of the province of Florence, Italy, where screening programmes started in the 1970s and 1990s. *Br J Cancer* 2004;90:1–4.
120. Barchielli A, Paci E. Trends in breast cancer mortality, incidence, and survival, and mammographic screening in Tuscany, Italy. *Cancer Causes Control* 2001;12:249–55.
121. Duffy SW, Tabar L, Olsen AH, et al. Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England. *J Med Screen* 2010;17:25–30.
122. Jergensen KJ, Zahl PH, Gotzsche PC. Breast cancer mortality in organized mammography screening in Denmark: comparative study. *BMJ* 2010;340:c1241.
123. Otten JD, Broeders MJ, Fracheboud J, et al. Impressive time-related influence of the Dutch screening programme on breast cancer incidence and mortality, 1975–2006. *Int J Cancer* 2008;123:1929–34.
124. Ugarte MD, Goicoa T, Etxeberria J, et al. Age-specific spatio-temporal patterns of female breast cancer mortality in Spain (1975–2005). *Ann Epidemiol* 2010;20:906–16.
125. American Cancer Society. Cancer prevention and early detection facts and figures, 2011–2012. Atlanta (GA): American Cancer Society; 2013.
126. Available at: <http://seer.cancer.gov/statfacts/html/breast.html>. Accessed March 10, 2010.
127. Welch HG. Screening mammography—a long run for a short slide? *N Engl J Med* 2010;363:1276–8.
128. Hellquist BN, Duffy SW, Abdsaleh S, et al. Effectiveness of population-based service screening with mammography for women ages 40–49 years: evaluation of the Swedish Mammography Screening in Young Women (SCRY) cohort. *Cancer* 2011;117:714–22.
129. Jonsson H, Bordas P, Wallin H, et al. Service screening with mammography in Northern Sweden: effects on breast cancer mortality—an update. *J Med Screen* 2007;14:87–93.
130. Jonsson H, Nystrom L, Tornberg S, et al. Service screening with mammography of women aged 50–69 years in Sweden: effects on mortality from breast cancer. *J Med Screen* 2001;8:152–60.
131. Jonsson H, Nystrom L, Tornberg S, et al. Service screening with mammography. Long-term effects on breast cancer mortality in the county of Gavleborg, Sweden. *Breast* 2003;12:183–93.
132. Yen AM, Duffy SW, Chen TH, et al. Long-term incidence of breast cancer by trial arm in one county of the Swedish Two County Trial of Mammography screening. *Cancer* 2012;118:5728–32.
133. Puliti D, Duffy SW, Miccinesi G, et al. Overdiagnosis in mammographic screening for breast cancer in

- Europe: a literature review. *J Med Screen* 2012; 191:42–56.
134. de Gelder R, Heijnsdijk EA, van Ravesteyn NT, et al. Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev* 2011;33:111–21.
 135. Paci E, Miccinesi G, Puliti D, et al. Estimate of overdiagnosis of breast cancer due to mammography after adjustment for lead time. A service screening study in Italy. *Breast Cancer Res* 2006;8:R68.
 136. Olsen AH, Agbaje OF, Myles JP, et al. Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. *Breast J* 2006; 12:338–42.
 137. Waller M, Moss S, Watson J, et al. The effect of mammographic screening and hormone replacement therapy use on breast cancer incidence in England and Wales. *Cancer Epidemiol Biomarkers Prev* 2007;16:2257–61.
 138. Tabar L, Duffy S, Vitak B, et al. The natural history of breast cancer. What have we learned from screening? *Cancer* 1999;86:449–62.
 139. Duffy SW, Parmar D. Overdiagnosis in breast cancer screening: the importance of length of observation period and lead time. *Breast Cancer Res* 2013; 15:R41.
 140. Kalager M, Adami H, Bretthauer M, et al. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. *Ann Intern Med* 2010;156: 491–9.
 141. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast cancer incidence. *N Engl J Med* 2012;367(2):1998–2005.
 142. Garfinkel L, Boring CC, Heath CW Jr. Changing trends: an overview of breast cancer incidence and mortality. *Cancer* 1994;79(Suppl):222–7.
 143. Kopans DB. The New England Journal of Medicine article suggesting overdiagnosis from mammography screening is scientifically incorrect and should be withdrawn. *J Am Coll Radiol* 2013; 10(5):317–9.
 144. Esserman L, Shieh Y, Thompson I. Rethinking screening for breast cancer and prostate cancer. *JAMA* 2009;302(15):1685–92.
 145. Feig SA. Ductal carcinoma in situ: implications for screening mammography. *Radiol Clin North Am* 2000;38:653–68.
 146. Lagios MD, Margolin FR, Westdahl PR, et al. Mammographically detected ductal carcinoma in situ. Frequency of local recurrence following tylectomy and prognostic effect of nuclear grade on recurrence. *Cancer* 1989;63:618–24.
 147. Schwartz GF, Finkel GC, Garcia JC, et al. Subclinical ductal carcinoma in situ of the breast. Treatment by local excision and surveillance along. *Cancer* 1992;70:2468–74.
 148. Silverstein M, Waisman JR, Gamigami P, et al. Intraductal carcinoma of the breast (208 cases): clinical factors influencing treatment choice. *Cancer* 1990; 66:102–8.
 149. Yen MF, Tabar L, Smith RA, et al. Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. *Eur J Cancer* 2003;39:1746–54.
 150. UK National Health Service. NHS Breast Screening Programme and Association of Breast Surgery at British Association of Surgical Oncology: an audit of screen detected breast cancers for the year of screening April 2007 to March 2008. UK West Midlands Cancer Intelligence Unit;2009.
 151. McCann J, Treasure P, Duffy SW. Modeling the impact of detecting and treating ductal carcinoma in situ in a breast screening programme. *J Med Screen* 2004;11:117–25.
 152. Duffy SW, Chen HH, Tabar L, et al. Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. *Stat Med* 1995;14:1521–43.
 153. Chen HH, Duffy SW, Tabar L, et al. Markov chain models for progression of breast cancer. Part I: tumor attributes and the preclinical screen-detectable phase. *J Epidemiol Biostat* 1997;2:9–25.
 154. Chen HH, Duffy SW, Tabar L, et al. Markov chain models for progression of breast cancer. Part II: prediction of outcomes for different screening regimes. *J Epidemiol Biostat* 1997;2:25–35.
 155. White E, Miglioretti DL, Yankaska BC, et al. Biennial versus annual mammography and the risk of late-stage breast cancer. *J Natl Cancer Inst* 2004; 96(24):1832–9.
 156. Kerlikowske K, Zhu W, Hubbard RA, et al. Outcome of screening mammography by frequency, breast density, and postmenopausal hormone therapy. *JAMA Intern Med* 2013;173(9):807–15.
 157. Ontilo AA, Engel JM, Liang H, et al. Mammographic utilization: patient characteristics and breast cancer stage at diagnosis. *AJR Am J Roentgenol* 2013;201:1057–63.
 158. Hunt KA, Rosen EL, Sickles EA. Outcome analysis for women undergoing annual versus biennial screening mammography: a review of 23,211 examinations. *AJR Am J Roentgenol* 1999;173:285–9.
 159. Parvinen I, Chiu S, Pylkkanen L, et al. Effects of annual vs triennial mammography interval on breast cancer incidence and mortality in ages 40–49 in Finland. *Br J Cancer* 2011;105:1388–91.
 160. Meyer JE, Kopans DB. Stability of a mammographic mass: a false sense of security. *AJR Am J Roentgenol* 1981;137:595–8.
 161. Lev-Toaff AS, Feig SA, Saitas VL, et al. Stability of malignant breast microcalcifications. *Radiology* 1994;192:153–6.

162. Feig SA. Carcinoma without interval change. In: Bassett LW, Feig SA, Hendrick RE, et al, editors. Breast disease (third series, test and syllabus). Reston (VA): American College of Radiology; 1999. p. 167–79.
163. Field LR, Wilson TE, Strawderman M, et al. Mammographic screening in women more than 64 years old: a comparison of 1- and 2-year intervals. *AJR Am J Roentgenol* 1998;170:961–5.
164. US Preventive Services Task Force. Screening for breast cancer: US Preventive Services Task Force recommendation statement. *Ann Intern Med* 2009;151:716–26.
165. Nelson HD, Tyne K, Naik A, et al, US Preventive Services Task Force. Screening for breast cancer: an update for the US Preventive Services Task Force. *Ann Intern Med* 2009;151:727–37.
166. Seidman H, Stellman SD, Mushinski MH. A different perspective on breast cancer risk factors: some implications of nonattributable risk. *Cancer* 1982;32:301–13.
167. Smith RA. Risk-based screening for breast cancer: is there a practical strategy? *Semin Breast Dis* 1992;2:280–91.
168. Lee CH, Dershaw DD, Kopans D, et al. Breast cancer screening with imaging: recommendation from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. *J Am Coll Radiol* 2010;7:18–27.
169. American College of Obstetricians and Gynecologists Breast Cancer Screening (ACOG). ACOG practice bulletin no. 122. Breast Cancer Screening. *Obstet Gynecol* 2011;118:372–82.
170. Szabo L. Women are insistent on mammograms, poll shows. McLean (VA): USA Today; 2009. p. A1.
171. Sack K, Kolata G. Breast cancer screening policy won't change, US officials say. *New York Times*; 2009. A1, A3.
172. Mandelblatt JS, Cronin KA, Bailey S, et al, Breast Cancer Working Group of the Cancer Intervention and Surveillance Modeling Network. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Ann Intern Med* 2009;151:738–47.
173. Kerlikowske K. Evidence-based breast cancer prevention: the importance of individual risk. *Ann Intern Med* 2009;151:750–2.
174. Schwartz LM, Woloshin S, Fowler FJ Jr, et al. Enthusiasm for cancer screening in the United States. *JAMA* 2004;291:71–8.
175. Schwartz LM, Woloshin S, Sox HC, et al. US women's attitudes to false-positive mammography results and detection of ductal carcinoma in situ: cross sectional survey. *BMJ* 2000;320:1635–40.
176. Feig SA. Number needed to screen: appropriate use of this new basis for screening mammography guidelines. *AJR Am J Roentgenol* 2012;198(5):1214–7.
177. Hendrick RE, Helvie MA. Mammography screening: a new estimate of number needed to screen to prevent one breast cancer death. *AJR Am J Roentgenol* 2012;198(3):723–8.
178. O'Donoghue C, Eklund M, Ozanne EM, et al. Aggregate cost of mammography screening in the United States: comparison of current practice and advocated guidelines. *Ann Intern Med* 2014;160:145–53.
179. Elmore JG. The cost of breast cancer screening in the United States: a picture is worth...a billion dollars? *Ann Intern Med* 2014;160:203–4.
180. Rosenquist CJ, Lindfors KK. Screening mammography beginning at age 40 years: a reappraisal of cost-effectiveness. *Cancer* 1998;82:2235–40.