Breast Cancer Screening: A Summary of the Evidence

Linda L. Humphrey, MD, MPH; Mark Helfand, MD, MS; Benjamin K.S. Chan, MS; and Steven H. Woolf, MD, MPH

Epidemiology

Breast cancer is the second leading cause of cancer death among North American women. Approximately 1 in 8.2 women will receive a diagnosis of breast cancer during her lifetime, and 1 in 30 will die of the disease.[1] Breast cancer incidence increases with age,[1] and although significant progress has been made in identifying risk factors and genetic markers, more than 50% of cases occur in women without known major predictors.[2–5]

This review was commissioned to assist the current U.S. Preventive Services Task Force (USPSTF) in updating its recommendations on breast cancer screening. We focus on information that was not available in 1996, when the previous USPSTF examined the issue.[6] Our goal was to critically appraise and synthesize evidence about the overall effectiveness of breast cancer screening, as well as its effectiveness among women younger than 50.

Methods

The analytic framework, literature search, and data extraction are described in detail in the Appendix. Briefly, we searched the Cochrane Controlled Trials Registry, MEDLINE, PREMEDLINE, and reference lists[6–8] for randomized, controlled trials of screening with death from breast cancer as an outcome. In all, we reviewed 154 publications from 8 eligible randomized trials of screening mammography and 2 trials of breast self-examination (BSE). We abstracted details about patient population, design, quality, data analysis, and published results at each reported length of follow-up. We also evaluated previous meta-analyses of these trials and of screening test characteristics and studies evaluating the harms associated with false-positive test results.

We used predefined criteria developed by the current USPSTF to assess the internal validity of the trials.[9] Two authors rated the internal validity of each study as "good," "fair," or "poor." Disagreements were resolved by further review and discussion. In the USPSTF system, a study that meets all the criteria for internal validity is rated as good quality.[9] The rating reflects a judgment that the results of the study are very likely to be correct. The fair-quality rating is used for studies that have important but not major flaws and implies that the findings are

From Oregon Health & Science University (Humphrey, Helfand, Chan) and Portland Veterans Affairs Medical Center (Humphrey, Helfand), Portland, Oregon; and Medical College of Virginia, Virginia Commonwealth University (Woolf), Fairfax, Virginia.

Address correspondence to: Linda L. Humphrey, MD, MPH, Oregon Health & Science University, Mailcode BICC, 3181 SW Sam Jackson Park Road, Portland, OR 97201-3098. E-mail: Humphrey@ohsu.edu.

Reprints are available from the AHRQ Web site (www.preventiveservices.ahrq.gov), through the National Guideline Clearinghouse (www.guideline.gov), or in print through the AHRQ Publications Clearinghouse (call 1-800-358-9295 or e-mail ahrqpubs@ahrq.gov).

The USPSTF recommendations based on this evidence review can be found in Screening for Breast Cancer: Recommendations and Rationale (which precedes this chapter), available on the AHRQ Web site and through the AHRQ Publications Clearinghouse.

This chapter first appeared as an article in *Ann Intern Med.* 2002;137(5 Part 1):347-360.

probably valid. A study that has a major flaw in design or execution—one that is serious enough to invalidate the results of the study—is rated as poor quality. We based our quality ratings on the entire set of publications from a trial rather than on individual articles.

The USPSTF criteria for internal validity are listed in Appendix Table 1. All of the mammography trials met the first 3 criteria: They clearly defined interventions, measured important outcomes, and used intention-to-treat analysis. Therefore, our quality ratings reflect differences among the studies on the remaining criteria: (1) initial assembly of comparable groups; (2) maintenance of comparable groups and minimization of differential loss to follow-up or overall loss to follow-up; and (3) use of outcome measurements that were equal, reliable, and valid. The Appendix describes our approach to applying these criteria in more detail.

We conducted new meta-analyses to incorporate new information about the quality of the trials and longer follow-up results. Breast cancer is known for its biological heterogeneity[10] as well as for late recurrences.[10] Thus, longer follow-up is relevant in evaluating mortality rates, particularly in younger women. In addition, for several of the trials, the most recent analyses correct flaws in earlier reports.

Six of the 8 mammography trials were designed to assess the effectiveness of mammography over a broad age range, rather than its comparative effectiveness in various age subgroups. One trial specifically examined women 40 to 49 years of age because the earliest trial seemed to show no benefit in this subgroup. The USPSTF posed these questions for the meta-analysis: (1) Does mammography reduce breast cancer mortality rates among women over a broad range of ages when compared with usual care? and (2) If so, does mammography reduce breast cancer mortality rates among women 40 to 49 years of age when compared with usual care?

We answered each question in 2 parts. First, using WinBUGS software (MRC Biostatistics Unit, Cambridge, United Kingdom), we constructed a 2-level Bayesian random-effects model to estimate the effect size from multiple data points for each study and to derive a pooled estimate of relative risk reduction and credible intervals (CrIs) for a given length of follow-up.[11] Second, we pooled the most recent results of each trial to calculate the absolute and relative risk reduction, using the results of the first analysis to estimate the mean length of follow-up.

To avoid bias that could result from excluding any data from valid studies, we included the results of all trials of fair quality or better in the base-case analysis. The disadvantage of this approach is that it combines results from 2 distinct types of studies.

The 6 population-based trials randomly assigned women to an invitation-to-screening group or to a control group that received "usual care" and was followed passively. In these trials, women who were invited to screening but chose not to be screened were included in the analysis of the "screened" group. Two trials from Canada, the Canadian National Breast Screening Study-1 (CNBSS-1) and the Canadian National Breast Screening Study-2 (CNBSS-2), differed from the other 6 trials. First, the Canadian trials used mass media to recruit a sample of volunteers, and all women randomly assigned to mammography had mammography at least once.[12,13] Second, in CNBSS-2, the control group was screened periodically with clinical breast examination (CBE). To estimate the relative risk reduction and the number needed to invite to screening to prevent one breast cancer death compared with usual care, we reanalyzed the data excluding the results of the Canadian studies.

Role of the Funding Source

This study was funded by the U.S. Agency for Healthcare Research and Quality. Agency staff and members of the USPSTF reviewed and made substantive recommendations about the analyses and final manuscript. Agency approval was required before the manuscript could be submitted for publication.

Results

Description of Trials

The 8 randomized trials of mammography identified in our review[12–23] varied in recruitment of participants, mammography protocol, control groups, and size (Table 1). Six trials examined the effectiveness of screening among women between 40 and 74 years of age; 1 trial enrolled women in their 40s, and 1 enrolled only women in their 50s. Four trials from Sweden tested mammography only,[14–17,23–26] and the other 4, from Canada, New York, and Edinburgh, Scotland, tested mammography and CBE.[12,13,18–22,27]

Study Quality

We found important methodologic limitations in all of the trials and rated all but 1 as fair, using USPSTF criteria. Table 1 lists the flaws of each trial and indicates how they influenced the overall ratings. The 2 reviewers rated the Swedish and Canadian trials as fair. Their initial ratings for the Edinburgh study and for the Health Insurance Plan of Greater New York (HIP) study differed. After extensive peer review, and detailed review of these trials' associated publications, the reviewers reached a consensus that the HIP study should be rated as fair and the Edinburgh study should be rated as poor.

The HIP trial (conducted from 1963 to 1966) was the first trial of breast cancer screening. It is difficult to critically appraise because publications that describe it differ in detail from more recent publications. We found several limitations of this trial, including inadequate description of allocation concealment and poor reporting of intervention and control group numbers. In addition, we found better ascertainment of clinical variables (including previous mastectomy) among the invitation-to-screening cohort than among the passively followed control group. However, we viewed this as an expected consequence of a study design in which a control group receives usual care and is not contacted. The screening and control groups differed from each other slightly in education, menopausal status, and previous breast lumps; however, the differences were not systematic and did not favor 1 group over the other. The strengths of the trial included intention-to-treat analysis, little contamination, and blind review of deaths. We did not find the faults severe enough to rate the study as poor quality and rated it as fair, which signifies that the results were probably valid at the time the study was conducted.

The Canadian trials met all of the USPSTF criteria for a rating of good quality, except for adequacy of allocation concealment. They differed from the other trials because all participants had a history and physical examination before randomization. This design permitted exclusion of patients who had a history of breast cancer and extensive examination of the baseline differences between groups.

The Swedish trials all had limitations that resulted in a rating of fair rather than good. The Stockholm and Malmo trials, which were individually randomized, did not report whether allocation was concealed. The Gothenburg trial and Swedish Two-County Trial, which were cluster randomized, had small differences in mean age between the invited and control groups. Such differences are expected to occur in a cluster-randomized trial, do not indicate failure of randomization or a problem in the trial execution, and can be adjusted for in statistical analyses.[28] Both the Gothenberg trial and the Swedish Two-County Trial provided insufficient data to determine whether randomization distributed other important confounders equally among the groups, but comparison of overall mortality rates in the invited and control groups do not suggest that a major imbalance occurred.[29]

As originally conducted, the Swedish trials had important flaws related to measurement of the primary outcome measure, death from breast cancer. In the Swedish Two-County, Gothenburg, and Stockholm trials, review of deaths was unblinded and criteria for the assignment of cause of death were unclear. Another concern about the Swedish trials as a group related to screening of the control groups. Originally, the Swedish trials used the "evaluation" method of analysis, in which mortality rates in the screened population were calculated only for cancer diagnosed between the time of

Table 1. Controlled trials of mammography and clinical breast examination

Trial	HIP[19]	CNBSS-1[13]	CNBSS-2[13,20]	Edinburgh[18]	Gothenburg[14,23]	Stockholm[17]	Malmo[25]	Swedish 2-County Trial[16]
Description								
Year study began setting/population	1963 New York health plan members	1980 15 centers in Canada, self-selected subjects	1980 15 centers in Canada, self-selected subjects	1978 All women aged 45-64 from 87 general practices in Edinburgh	1982 Entire female population, born between 1923-1944, of one Swedish town	1981 Residents of southeast greater Stockholm, Sweden	1976-1978 All women born between 1927-1945 living Malmo, Sweden	1977 From Ostergotland (E-County) and Kopparberg (W-County)
Age at enrollment *(years)*	40-64	40-49	50-59	45-64	39-59	40-64	45-70	40-74
Interventions								
Method of randomization	Age- and family size-stratified pairs of women randomized assigned individually by drawing from a list	Blocks (stratified by center and 5-year age group) after CBE		Cluster, based on general practitioner practices	Cluster, based on day of birth for 1923-1935 cohort (18%), by individual for 1936-1944 cohort (82%)	Individual, by day of month; ratio of screening to control group, 2:1	Individual, within birth year	Cluster, based on geographic units; blocks designed demographically homogeneous
Study Groups	Mammography + CBE vs usual care	Mammography + CBE vs usual care (all women prescreened and instructed in BSE)	Mammography + CBE vs CBE (all women prescreened and instructed in BSE)	Mammography + CBE vs usual care	Mammography vs usual care; controls offered screening after year 5, completed screening at approximately year 7	Mammography vs usual care; controls offered screening after year 5	Mammography vs usual care; controls offered screening after year 14	Mammography vs usual care; controls offered screening after year 7
Screening protocol:								
interval (months)	12	12	12	24	18	24-28	18-24	24-33
rounds (*n*)	4	4-5	4-5	4	5	2	9	3
views (*n*)	2	2	2	2 (1)	2 (1)	1	2 (1)	1
Subjects (*n*)								
Study group	30,239	25,214	19,711	28,628	20,724	40,318	21,088	77,080
Control group	30,256	25,216	19,694	26,015	28,809	19,943	21,195	55,985
Longest follow-up by 2002 (*years*)	18	13	13	14	12*	11.4*	11-13	20 15.5* 15.5*
Trial quality	**HIP[19]**	**CNBSS-1[13]**	**CNBSS-2[13,20]**	**Edinburgh[18]**	**Gothenburg[14,23]**	**Stockholm[17]**	**Malmo[25]**	**Swedish 2-County Trial[16]**
Assembly of comparable groups								
Allocation concealment and baseline groups	*Use of lists and pairs made subversion possible. More menopausal women and women with previous breast lumps in a sample of controls;*	*Use of lists and blocks made subversion possible.* In mammography arm, 17 had tumors with 4 nodes with initial screening vs 5 in control arm	*Use of lists and blocks made subversion possible*	Allocation concealment not described. *Significantly lower SES and higher all cause mortality in control group suggest inadequate randomization*	Allocation concealment not described	*Allocation concealment not described*	*Allocation concealment not described*	Allocation concealment not described; intervention women slightly older than controls
All cause mortality relative risk (screened vs control group)	0.98	1.02	1.06	*0.8 (statistically significant)*	0.98	NR	0.99	1

Continued on page 185

Table 1. Controlled trials of mammography and clinical breast examination *(continued)*

Trial	HIP[19]	CNBSS-1[13]	CNBSS-2[13,20]	Edinburgh[18]	Gothenburg[14,23]	Stockholm[17]	Malmo[25]	Swedish 2-County Trial[16]
Maintenance of comparable groups								
Screening attendance								
Round	1 2 3 4	1 2,4	1 2 5	1 7	1 2 - 5 control	1 2 control	1 2-5 control	1 2 3 control
%	67 54 50 46	100 85-89	100 90.4 86.5	61 44	85 75-78 66	81 81 77	74 70 ???	89 83 84 ???
Contamination (%)	Unknown, probably small	25	16	NR	20	NR	25	13
Post-randomization exclusions	Yes	No	No	Yes	One fewer death in screening group included in 1997 results	Yes	Yes	Yes
Variety of outcome assessment								
Deaths included in analysis (follow-up vs evaluation method)	*Breast cancer deaths diagnosed within 7 years of follow-up*	Follow-up method	Follow-up method	Follow-up method and evaluation method	Initially, all four Swedish trials used the evaluation method of analysis (breast cancer cases diagnosed after screening period were excluded from count of breast cancer deaths), but this was corrected in re-analyses of the data in 1993 and in 2002. *Control screening was delayed relative to the last screen in the mammography groups, resulting in bias because more cases of cancer were included in the control groups than in the intervention groups*.			
Method for verifying breast cancer deaths	Blinded review of the death certificate and medical records; *unclear how deaths were selected for review*	Blinded review of all deaths of women known to have breast cancer whose death certificate mentions liver, lung, colon cancer, or unknown primary, or whose medical record raised a question of breast cancer		All deaths with breast cancer deaths diagnosed within 14 years of follow-up; *not masked*	In the 1993 analysis, an independent panel used an explicit protocol to preform blinded assessment of cause of death.			
Analysis method								
Intention-to-treat analysis; completeness of reporting†	*Did not provide relative risk, confidence intervals, or P values in recent report; estimated the number of subjects*	Appropriate	Appropriate	–	In all the Swedish trials, sample sizes differed for different publications because different methods were used to estimate the size of the underlying population.			
External validity								
Comment	Poor mammography technique; only a third of cancer cases found by mammography alone	Many women with screening abnormalities (especially CBE) were "deemed not to require a diagnostic procedure," potentially reducing the sensitivity of screening		–	19% of controls and 13% of study women had mammography in the 2 years before the study	25% of all women entering the study had a mammogram before to entering the study	–	In the age group of 40-49 years, 3 women died after being invited to screening and 1 died before invitation but after randomization
GRADE	USPSTF Internal Validity	Fair	Fair or better	Fair or better	Poor	Fair	Fair	Fair

*Most recent results for age 40-49, if different.

†All studies were analyzed using intention-to-treat methods.

Note: Italics indicate aspects of the design or conduct of trials that influenced the quality rating.

BSE indicates breast self-examination; CBE, clinical breast examination; CNBSS, Canadian National Breast Screening Study; HIP, Health Insurance Plan of Greater New York; NR, not reported; USPSTF, U.S. Preventive Services Task Force.

randomization and the last mammographic examination. When the evaluation method of analysis is used, control group screening can introduce bias unless it is performed concurrently with the final instance of mammography in the screened group.[30,31] This method is inferior to the "follow-up" method of analysis, in which all deaths that occur after randomization are included in the analysis. The follow-up method of analysis dilutes relative benefit over time, particularly in studies that offered screening to the control group and in areas where widespread screening is adopted.

We considered these flaws to be adequately corrected in subsequent analyses by the trialists. In a 1993 overview of the trials, an independent end point committee used an explicit protocol to perform blind assessment of cause of death.[32] Participants were linked to an external cancer registry and were excluded from the analysis if breast cancer had been diagnosed before the trial began. For the Swedish trials as a whole, death from every cause except breast cancer was similar in the compared groups.[33] In the Swedish Two-County Trial, the reduction in rates of advanced breast cancer,[34] which are not related to judgments about the causes of death, was similar to the reduction in breast cancer mortality rates.[35] The overview also reanalyzed the data by using the follow-up method of analysis and found very little difference between the recalculated and original relative risk values. A recent review[8] critical of the Swedish studies raised concern about bias in postrandomization exclusions, as evidenced by variation in the reported number of participants. This concern was effectively addressed in a recent update of these trials, which explained that this variation was due to the use of different methods for estimating the number of women in each birth cohort rather than to manipulation after randomization.[23] The update also reported more recent results of the Swedish trials by using both the follow-up and evaluation methods of analysis.

We rated the Edinburgh study as poor quality because of a serious imbalance between the control and screened groups. General practitioners' practices were randomized in clusters without matching for socioeconomic factors. As a result, socioeconomic status, a predictor of stage at diagnosis as well as death from breast cancer, was significantly lower in the control group than in the mammography group. All-cause mortality was dramatically higher in the control group than in the screened group (20.1 more deaths per 10,000 person-years [Confidence interval (CI), 13.3 to 26.9]).[29] This difference is close to 25 times larger than the difference in breast cancer deaths between the groups and confirms our assessment that the trial was severely flawed.

Sensitivity of Mammography

Since no gold standard can be applied to the entire screened population, the denominator used for estimating sensitivity is the total number of breast cancer cases diagnosed in a given interval. The results of recent, good-quality systematic reviews of the accuracy of mammography in the screening trials are summarized in Table 2.[36,37] The overall sensitivity for all rounds of screening was lowest in the HIP trial. Otherwise, 1 study was not clearly better or worse than another. For a 1-year screening interval, the sensitivity of first mammography ranged from 71% to 96%. Sensitivity was substantially lower for women in their 40s than for older women.

The data in Table 2 cannot be applied to individual patients because they are not adjusted for several factors that are known to affect sensitivity. These include patient factors (use of hormone replacement therapy, mammographic breast density), technical factors (the quality of mammography, the number of mammographic views), and provider factors (the experience of radiologists and their propensity to label the results of an examination abnormal, the choice of follow-up evaluation for abnormal mammograms).[36,38–42]

Specificity and Positive Predictive Value

In the randomized trials, the specificity of a single mammographic examination was 94% to 97%.[36,43–44] This indicates that 3% to 6% of women who did not have cancer underwent further diagnostic evaluation, typically a clinical examination, more mammographic views, or ultrasonography. The positive predictive value of 1-time mammography ranged from 2% to 22% for abnormal results

Table 2. Sensitivity of mammography*

Study	All rounds				First round only	
	Cases of cancer detected by screening	Total cases of cancer	%	Estimated sensitivity of mammography (no. of rounds)†	Sensitivity of screening at 1-year intervals	Sensitivity of screening at 2-year intervals
HIP (ages 40-64)	73	173	0.42	0.39 (4)		
Malmo (ages 45-69)	176	227	0.78	0.61 (2)	.92	
45-49					.73	
50-59					.71	
60-69					.85	
70-74					.81	
Swedish Two-County Trial (ages 40-74)					.95	.86
40-49	39	82	0.48		.81	
50-59	102	137	0.74		.96	
60-69	184	220	0.84		.95	
70-74	101	112	0.90		.98	
Stockholm (ages 40-64)					.86	.68
40-49	24	45	0.53	0.64		.53
50-59	71	95	0.75	0.89		.75
60-64	33	48	0.69			.69
CNBSS-1 (ages 40-49)	162	286	0.57	0.61 (4)	.77	0.56
CNBSS-2 (ages 50-59)	243	347	0.70	0.66 (4)	.88	.56

*Gothenburg is not listed because of insufficient data; the Edinburgh trial is excluded. Empty cells also indicate lack of sufficient data. All data are taken from reference 36, using the "detection" method, unless otherwise noted.

†Data taken from reference 37.

Note: CNBSS indicates Canadian National Breast Screening Study; HIP, Health Insurance Plan of Greater New York.

requiring further evaluation and from 12% to 78% for abnormal results requiring biopsy (Table 3).[36,45,46] Estimates from community settings suggest a graded, continuous increase in predictive value with age. For example, among 31,814 average-risk women screened in California from 1985 to 1992, the positive predictive value for further evaluation was 1% to 4% among those 40 to 49 years of age, 4% to 9% among those 50 to 59 years of age, 10% to 19% among those 60 to 69 years of age, and 18% to 20% among those 70 years of age and older.[47]

Effectiveness of Mammography in Reducing Breast Cancer Mortality

Table 4 summarizes the most recent results from trials that included at least some participants older than 50. The 4 Swedish trials that compared 2 to 6 rounds of mammography with usual care[23,26] reported 9% to 32% reductions in the risk for death from breast cancer. The results of the trials have changed little over time (Figure 1). The reduction was statistically significant in only 1 of these trials (the Swedish Two-County Trial) (relative risk [RR],

Table 3. Specificity and positive predictive value*

Study	Specificity(%)	Positive Predictive Value (%)	
	work-up method	work-up method	biopsy method
HIP[19]	NR	12	20
Malmo[25]	97.4	10-22	33-61
Swedish Two-County Trial[16]	95.6	12	50-75
Stockholm[17]	95.1	8-10	62-78
CNBSS-1[13]	93.5	2	12
CNBSS-2[13,20]		4-6	20
Gothenburg[14,23]		3-7 (complete mammography) 12-18 (CBE and FNA biopsy)	

*Adapted from references 36 and 45. Work-up method, a mammogram requiring further evaluation; biopsy method, a mammogram resulting in biopsy.

Note: CBE indicates clinical breast examination; CNBSS, Canadian National Breast Screening Study; FNA, fine-needle aspiration; HIP, Health Insurance Plan of Greater New York; NR, not reported.

0.68; CI, 0.59 to 0.80).[26] The number of times mammography was performed and the frequency of screening did not seem to explain the variation among the Swedish studies. A previous meta-analysis found little change when the individual trial results were adjusted for type of randomization and degree of adherence.[48]

Of the 4 studies that evaluated the combination of mammography and CBE (Table 4), 3 were of at least fair quality.[12,13,18,27,49] The HIP trial reported a relative risk reduction that began 5 years after randomization and remained below 1 after 16 or more years of follow-up (RR, 0.79). The CNBSS-2, which compared annual mammography and CBE with annual CBE among women 50 to 59 years of age, showed no benefit 13 years after the study began.[12,20] The CNBSS-1, which compared annual mammography and CBE with usual care in women 40 to 49 years of age, also showed no benefit.

In our meta-analysis of results from all age groups combined, we excluded the Edinburgh trial (which we rated as poor) and used the results from both Canadian trials. The summary relative risk was 0.84 (95% CrI, 0.77 to 0.91), equivalent to a number needed to screen of 1,224 (CrI, 665 to 2,564) an average of 14 years after study entry. To estimate the effectiveness of an invitation to screen compared with usual care, we also excluded the Canadian trials, which recruited volunteers. The relative risk reduction was 0.81 (CrI, 0.73 to 0.89), and the number needed to invite to screening was 1,008 (CrI, 531 to 2,128). The relative risks by year of observation (including trial plus follow-up time) are shown in Figure 1, which suggests a gradual decrease in benefit with longer observation time.

Effectiveness of Mammography among Women 40 to 49 Years of Age

Since 1963, 7 randomized, controlled trials have included women 40 to 49 years of age, approximately 200,000 participants. With the exception of 1 of the Canadian studies, none of the trials were planned to evaluate breast cancer screening in this age group and none had sufficient power. Two trials, the Stockholm trial and CNBSS-1, showed no benefit for this age group even with longer follow-up (Table 5). The other 5 trials suggest a benefit (risk reduction, 13% to 42%), and 1 (the Gothenburg trial) observed a statistically significant risk reduction since 1996. These findings reflect results after 11 to 19 years of observation; the

Table 4. Randomized controlled trials of mammography among women aged 39-74

Study (reference)	Ages	Median follow-up (years)	Number of breast cancer deaths/total number of women		Breast cancer death rate per 1,000 women		Relative risk for breast cancer death (95% confidence interval)	Absolute risk reduction per 1,000 women	Number needed to invite*
			Screened	Control	Screened	Control			
Studies of mammography alone									
Stockholm[23]	40-64	13.8	82/39,139	50/20,978	2.10	2.38	0.91 (0.65-1.27)	0.288	3,468
Gothenburg[23]	39-59	12.8	62/20,724	113/29,200	2.99	3.87	0.76 (0.56-1.04)	0.878	1,139
Malmo[23]	45-70	17.1	161/21,088	198/21,195	7.63	9.35	0.82 (0.67-1.00)	1.712	584
Swedish Two-County Trial[26]	40-74	17	319/77,080	333/55,985	4.14	5.95	0.68 (0.59-0.80)	1.809	553
Studies of mammography plus CBE									
CNBSS-1[22]	40-49	13	105/25,214	108/25,216	4.16	4.28	0.97 (0.74-1.27)	0.12	—
CNBSS-2[20]	50-59	13	107/19,711	105/19,694	5.43	5.33	1.02 (0.78-1.33)	-0.097	—
HIP[19]	40-64	16	232/30,239	281/30,256	5.46	6.89	0.79	1.438	883
Edinburgh[18]	45-64	13	156/22,926	167/21,342	6.80	7.82	0.79 (0.60-1.02)	1.020	980

*Number needed to invite for screening to prevent one death from breast cancer 13-20 years after randomization.

Note: CBE indicates clinical breast examination; CNBSS, Canadian National Breast Screening Study; HIP, Health Insurance Plan of Greater New York.

Table 5. Results of mammography trials among women younger than 50 years of age

Study (reference)	Ages	Median follow-up (years)	Number of breast cancer deaths/total number of women		Breast cancer death rate per 1,000 women		Relative risk for cancer death (95% confidence interval)	Absolute risk r eduction per 1,000 women	Number needed to invite*	Follow-up year controls screened
			Screened	Control	Screened	Control				
Studies of mammography alone										
Stockholm[23]	40-49	14.3	34/14,842	13/7,103	2.29	1.83	1.52 (0.8-2.88)	No reduction	—	5
Gothenburg[23]	39-49	12.7	22/11,724	46/14,217	1.88	3.24	0.58 (0.35-0.96)	1.36	736	7
Malmo[23]	45-50	13.3	53/13,568	66/12,279	3.91	5.38	0.73 (0.51-1.04)	1.47	681	4
Swedish Two-County Trial[16]	40-49	13	45/19,844	39/15,604	2.27	2.50	0.87 (0.54-1.41)	0.23	4316	7-8
Studies of mammography plus CBE										
CNBSS-1[22]	40-49	13	105/25,214	108/25,216	4.16	4.28	0.97 (0.74-1.27)	0.12	—	—
HIP[19,27]	40-49	14	64/13,740	82/13,740	4.66	5.97	0.78 (0.56-1.08)	1.31	763	—
Edinburgh[18]	45-49	13	49/11,749	53/10,267	4.17	5.16	0.75 (0.48-1.18)	0.99	1008	6-10

*Number needed to invite for screening to prevent one breast cancer death after 11-16 years.

Note: CBE indicates clinical breast examination; CNBSS, Canadian National Breast Screening Study; HIP, Health Insurance Plan of Greater New York.

Figure 1.

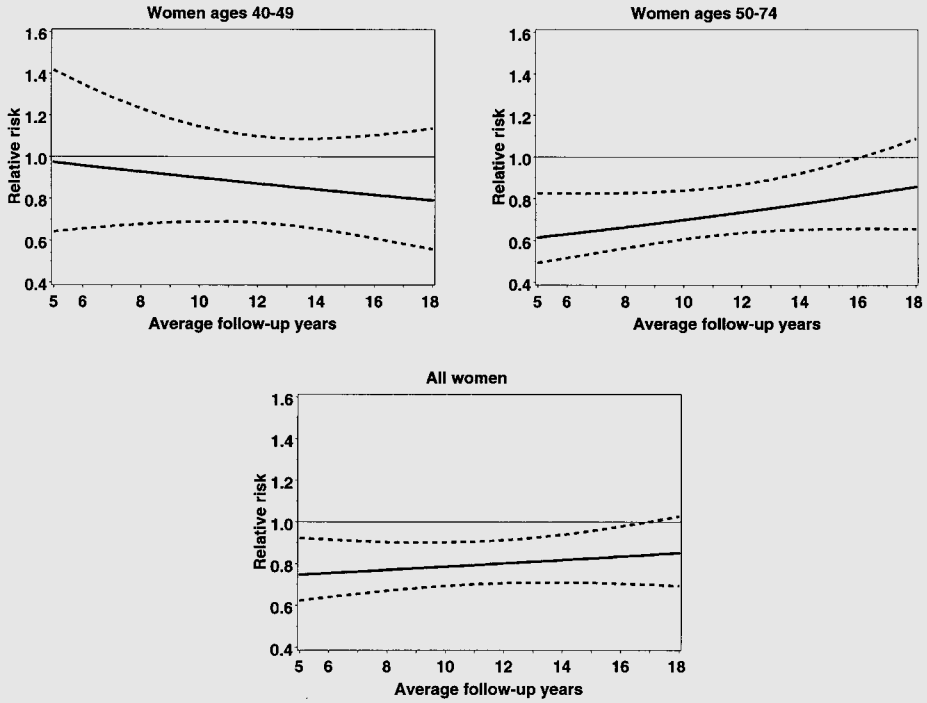


Figure 1. Relative risk compared with average years of follow-up for women 40 to 49 years of age, women 50 to 74 years of age, and all women. Estimated curves are from a herarchical meta-regression model. Dotted curves represent 95% credible intervals.

median period of active screening was 6 years (range, 4 to 15 years).

In our meta-analysis, excluding the Edinburgh trial, the summary relative risk was 0.85 (CrI, 0.73 to 0.99) after 14 years of observation, with a number needed to screen of 1,792 (CrI, 764 to 10,540) to prevent 1 death from breast cancer. Some might argue that the Canadian study should be excluded in calculating the number needed to invite to screening because its participants were prescreened volunteers who may have differed from the general population. When the Canadian study was excluded, the summary relative risk was 0.80 (CrI, 0.67 to 0.96) and the number needed to invite to screening was 1,385 (CrI, 659 to 6,060). Figure 1 shows an increasing screening benefit among this age group with a longer period of observation.

Among women 50 years of age or older, the summary relative risk was 0.78 (CrI, 0.70 to 0.87) after 14 years of observation, with a number needed to screen of 838 (CrI, 494 to 1,676) to prevent 1 death from breast cancer. As shown in Figure 1, the benefit has decreased with longer duration of follow-up.

We found 7 meta-analyses of the effectiveness of mammography in women 40 to 49 years of age (Table 6).[8,30,32,48,50–58] Our results, which reflect exclusion of 1 flawed trial, longer follow-up in 6 of the trials, and corrected results for the Swedish trials, were consistent with those of most previous meta-analyses. Two meta-analyses,[8,51] including 1 from the Cochrane Collaboration, produced results that differed substantially from ours. The Cochrane review reported a summary relative risk of 1.03 (CI, 0.77 to 1.38) but based this on only 2 trials.

Effectiveness of Mammography in Older Women

Direct evidence of effectiveness among older women is limited to 2 trials that included women

Table 6. Meta-analyses of randomized trials of screening mammography among women aged 40-49

Study (reference), Year	Assessed Quality?	Included Trials	Methods	Years of Follow-up	Relative Risk (95% Confidence Interval)	Number Needed to Screen
Larsson et al,[50] 1997; Nystrom et al,[32] 1993	No	5 Swedish trials	Weighted relative risks	12.8	0.77 (0.59-1.01)	
Cox,[51] 1997 Elwood,[52] 1993	No	All 8 trials	Fixed effects	10	0.93 (0.77-1.11)	
Glasziou and Irwig,[53,54] 1997	Yes. All studies were "good." Rated Malmo and CNBSS highest and Two-County trial and Gothenburg lowest	All 8 trials	Variance-weighted	13.13	0.85 (0.71-1.01)	
Hendrick et al,[55] 1997; Smart et al,[56] 1995	No	All 8 trials*	Fixed effects	12.7	0.82 (0.71-0.95)	1,540
Kerlikowske,[57,58] 1995,1997	No	All 8 trials	Fixed effects	≈ 12	0.84 (0.71-0.99)	2,500
Berry,[30] 1998	No	All 8 trials	Random effects†	12 -15	0.82 (0.49-1.17)	
Olsen and Gotzsche,[8] 2001	Yes. Excluded 6 trials rated "flawed" or "poor"	Canadian, Malmo	Fixed effects	13	1.03 (0.77-1.38)	
Current study, 2002	Yes. Rated Edinburgh "poor" and others fair or better	7 trials, excluding Edinburgh	Random effects	≈ 14	0.85 (0.73-0.99)	1,792

* Included an additional 17,000 subjects from the Malmo II trial.

† Hierarchical Bayes model; estimates are for the "next trial" analysis.

Note: For multiple publications, data from the most recent update are recorded in the table.

older than 65. Both of these trials reported relative risk reductions among women 65 to 74 years of age (RR, 0.68 [CI, 0.51 to 0.89]25 and 0.7959 among women 70 to 74 years of age). In the recent Swedish overview, the summary relative risk among women 65 to 74 years of age was 0.78 (CI, 0.62 to 0.99).[23,60]

Clinical Breast Examination

The test characteristics of CBE, based on data from trials designed specifically for breast cancer screening, were recently reviewed.[61] Sensitivity ranged from 40% to 69%, specificity from 88% to 99%, and positive predictive value from 4% to 50% when mammography and interval cancer were used as the criterion standard. One community study showed that over 10 years of biennial screening, 13.4% of women had false-positive results on CBE at least once; risk for such results was higher among women younger than 50.[62]

No trial has compared CBE alone with no screening. However, 2 randomized, controlled trials involving the use of mammography and CBE had mortality reductions of 29% and 14%.[18,27,63] A controlled, nonrandomized United Kingdom trial of

CBE and mammography showed a nonsignificant mortality reduction of 14% (RR, 0.86; CI, 0.73 to 1.01).[64]

What is the contribution of CBE to these reductions in mortality rate? Among studies showing a benefit of screening, mortality reductions in trials of CBE with mammography are similar to those in trials including mammography only. In the CNBSS-2, in which women 50 to 59 years of age were randomly assigned to annual CBE and mammography or to annual CBE,[65] the relative risk for death was 0.97 (CI, 0.62 to 1.52).[13] This suggests that mammography has little additive benefit in the setting of a careful, detailed CBE.

Breast Self-Examination

Because neither CBE nor mammography is 100% sensitive, BSE has been advised as an important screening method among women older than 20. However, its effectiveness in decreasing death from breast cancer has been controversial because evidence from clinical trials is limited. Observational studies evaluating BSE and breast cancer stage at diagnosis or death have had mixed results.[45,66]

In 2 randomized, controlled trials with 5 to 10 years of follow-up, both conducted outside the United States, breast cancer mortality rates were similar in women instructed in BSE and in noninstructed controls.[67–69] Both studies involved large numbers of women who were meticulously trained with proper technique and had numerous reinforcement sessions; mammography was not part of routine screening in the countries involved. In both trials, physician visits and biopsy for benign breast lesions increased among those educated in BSE. To date, no studies have evaluated other potential adverse outcomes of BSE, such as anxiety and subsequent screening behavior.

Adverse Effects

The most frequently discussed adverse effects of mammography are the anxiety, discomfort, and cost associated with positive test results, many of which are false positive, and the diagnostic procedures they generate. For a woman undergoing regular mammography, cumulative specificity may be more relevant than the specificity of a single examination. In 1 community setting involving 2,400 women 40 to 69 years of age, 6.5% of mammography results requiring further evaluation were false positives (specificity, 93.5%). When evaluated on an individual basis, however, approximately 23% of women had at least 1 false-positive result on mammography requiring further work-up during 10 years of biennial screening (average of 4 mammograms per woman), indicating a 10-year cumulative specificity of 76.2%. For every $100 spent on screening, $33 was spent on the evaluation of false-positive results.[62]

Anxiety over an abnormal mammogram is documented in some[70–74] but not all[71,75] studies. These studies generally suggest that anxiety dissipates after cancer is ruled out, but some studies suggest that some women worry persistently.[72,74–76] The anxiety associated with an abnormal mammogram does not seem to dissuade women from undergoing further screening[77] and may even be associated with improved adherence to recommended screening intervals.[70,78,79] Many women are willing to accept the risk for false-positive results. In 1 survey, 99% of women understood that false-positive examination results occur with screening, although they underestimated the likelihood. Of importance, 63% stated that they would accept 500 instances of false-positive examination results to save one life.[80]

Some view diagnosis and treatment of ductal carcinoma in situ (DCIS) as potential adverse consequences of mammography. There is incomplete evidence regarding the natural history of DCIS, the need for treatment, and treatment efficacy, and some women may receive treatment of DCIS that poses little threat to their health. In a 1992 study, 44% of women with DCIS were treated with mastectomy and 23% to 30% were treated with lumpectomy or radiation.[81,82] In 1 survey, only 6% of women were aware that mammography might detect nonprogressive breast cancer.[80]

Radiation exposure is also a potential risk associated with mammography.[83] Using risk

estimates provided by the Biological Effects of Ionizing Radiation report of the U.S. National Academy of Sciences, and assuming a 4 mGy mean glandular dose from each 2-views-per-breast bilateral mammography, Feig and Hendrick estimated that annual mammography of 100,000 women for 10 years beginning at 40 years of age would induce no more than 8 deaths from breast cancer.[84] Women with an inherited susceptibility to ionizing radiation damage have higher risk for radiogenic breast cancer,[10,85] although this has not been documented in association with mammography.

Discussion

Fair-quality, relatively consistent evidence suggests that mammography screening reduces breast cancer death among women 40 to 74 years of age. We found no evidence that inclusion of CBE conferred greater benefit than mammography alone. We also found no evidence supporting the role of BSE in reducing breast cancer mortality.

Over the 3 decades in which mammography trial data have been available, critical reviewers and the investigators themselves have discussed limitations and irregularities in data reporting. One highly publicized review by the Cochrane Collaboration criticized the trials in regard to randomization, postrandomization exclusions, and determination of deaths from breast cancer.[8] It found all but 2 of the trials, the Malmo trial and the Canadian trials, severely flawed or of poor quality and prompted some official bodies to question their support for screening mammography.

We identified many of the same design problems highlighted in the Cochrane review but reached different conclusions about their bearing on the validity of the findings. With the exception of the Edinburgh trial, we found inadequate evidence to conclude that the specific flaws identified introduced biases of sufficient magnitude or direction to invalidate the findings or to cause us to reject the inference that screening mammography reduces breast cancer mortality rates.

The effectiveness of screening in women 40 to 49 years of age is a longstanding controversy. In early years, it centered on the lack of evidence that observed risk reductions were statistically significant.[6,52,86] That argument has dissipated over time as more evidence has shown a significant separation in survival curves with longer follow-up. The delay in the separation of those curves, however, has prompted some to question whether the observed benefits are due to the detection of cancer after 50 years of age, suggesting little incremental benefit from initiating screening at 40 years of age and exposing women to the harms of screening for an extra decade.[87,88] We found little evidence to convincingly address this concern and some evidence that some benefit from screening women 40 to 49 years of age would be sacrificed if screening began at age 50 years.[27,89]

The use of 50 years of age as a threshold is somewhat arbitrary (except that it approximates the age of menopause). The risks for developing and dying of breast cancer are continuous variables that increase with age, and the greatest increase in incidence actually occurs before menopause.[90,91] We found that the relative risk reduction achieved with mammography screening does not differ substantially by age, although the time required to obtain the benefit is longer for younger women. On the other hand, younger women have more potential years of life to gain by screening. Thus, the variable most affected by age is absolute risk reduction, which increases as a continuum with age while the number needed to screen decreases. The age of 50 years has no special bearing on this pattern, and some question the scientific rationale for treating women 40 to 49 years of age as a special entity.[92]

What emerges as a more important concern, across all age groups, is whether the magnitude of benefit is sufficient to outweigh the harms. The risk for false-positive results and their consequences decreases with age. Thus, although mammography at any age poses a tradeoff of benefits and harms, the balance between increasing absolute risk reduction and decreasing harms grows more favorable over time. The age at which this tradeoff becomes acceptable is a subjective judgment that cannot be answered on scientific grounds, since early evidence suggests that women will tolerate a high risk for false-positive results. As noted earlier, 63% of women in one study stated that they would accept

500 instances of false-positive results to save one life.[80] On the basis of the results of our meta-analysis, we calculated that over 10 years of biennial screening among 40-year-old women invited to be screened, approximately 400 women would have false-positive results on mammography and 100 women would undergo biopsy or fine needle aspiration for each death from breast cancer prevented.

A limitation of our meta-analysis is that we combined studies that used different methods of analysis. In the most recent report from the Swedish trials,[23] Nystrom and colleagues did not report individual study-level data using the follow-up method. The pooled follow-up analysis reported by Nystrom and colleagues in 2002 suggests that the use of follow-up method would have resulted in a smaller estimate of relative risk reduction.

Women older than 70 have the highest incidence of breast cancer, and test performance in these women is likely to be similar to that in women 50 to 70 years of age. Therefore, theoretically, mammography should be at least as effective for women older than 65 as it is for younger women. Offsetting this potential benefit, however, is the greater comorbidity observed in elderly persons. The potential benefit of early detection is unlikely to be realized in women who have other diseases that diminish life expectancy, in those who would not tolerate evaluation or treatment, and in those with impaired quality of life (for example, dementia).[93] In addition, no data from randomized, controlled trials provide information about the morbidity associated with screening, follow-up, and treatment among women older than 74. Finally, a major concern in elderly women is the diagnosis and treatment of DCIS, since mortality rates from DCIS are low (1% to 2% at 10 years) and 99% of DCIS is treated surgically.[94]

The interval at which mammography was performed in the screening trials varied between 12 and 33 months, but annual mammography was no more effective than biennial mammography. Data from the Swedish Two-County Trial indicate that the period in which breast cancer can be detected before it presents clinically is shorter for women 40 to 49 years of age.[95–97] Annual screening may be more important in this age group than in older women, but we found no direct proof for this hypothesis in the controlled trials that have been completed so far.

We found no evidence that CBE or BSE reduces breast cancer mortality. Whether the BSE trials are generalizable to the United States, where the use of CBE and mammography and the incidence of breast cancer are higher, is uncertain. It is also uncertain whether BSE might be beneficial to women who are not in the age ranges at which mammography is recommended or do not avail themselves of mammography. In the setting of CBE and mammography, the probability of finding a significant decrease in mortality rates is likely to be small.

In summary, when judged as population-based trials of cancer screening, most mammography trials are of fair quality. Their flaws reflect tradeoffs in planning that make the trial results widely generalizable but decrease internal validity. In absolute terms, the mortality benefit of mammography screening is small enough that biases in the trials could erase or create it. However, we found that although these trials were flawed in design or execution, there is insufficient evidence to conclude that most were seriously biased and consequently invalid.

Future research should be directed toward developing new screening methods as well as methods of improving the sensitivity and specificity of mammography. Methods of reducing surgical biopsy rates and complications of treatment should also be studied, as should communication of the risks and benefits associated with screening to patients. Finally, efforts to identify breast cancer risk factors with high attributable risk, as well as appropriate prevention strategies, should continue. Even in the best screening settings, most deaths from breast cancer are not currently prevented.

Appendix

Analytic Framework

Because of the availability of population-based, randomized trials, mammography has the most direct type of evidence of any cancer screening program.[98] Nevertheless, mammography has been controversial since it was first proposed in the 1960s. To understand why, it is helpful to consider the assumptions underlying the steps in the causal chain from screening test to health outcomes. In the analytic framework (Appendix Figure 1), this evidence is shown by the overarching arc connecting screening with the outcomes, reduced morbidity and mortality. Mammography is aimed at early detection of invasive cancer, which is treated by major surgery (mastectomy or tumorectomy). This differs from screening for colorectal cancer and cervical cancer, which is aimed at detecting and removing precancerous lesions to prevent invasive cancer and to preserve the involved organ (colon or uterine cervix). This is 1 reason why, although it may be reasonable to endorse 1 cancer screening test (Papanicolaou smear) based on observational, indirect evidence, it may also be reasonable to require experimental evidence before endorsing another (mammography or prostate cancer screening).

It is important to note that the mammography trials do not necessarily provide the highest level of evidence about the efficacy of early treatment. While there is no doubt that screening results in earlier diagnosis of invasive breast cancer, the efficacy of earlier treatment of invasive cancer has not been established independently of the trials.[99] That is, there is no direct evidence from trials of surgical therapy (versus watchful waiting) that earlier treatment of invasive cancer reduces mortality. The mammography trials do not attempt to link specific treatments, such as radical mastectomy or adjuvant radiation, to improved outcomes.

The reliance on a theory of treatment rather than on evidence about the efficacy of treatment increases the burden of proof placed on the trials of mammography. It also distinguishes cancer screening from other screening services considered by the USPSTF, such as chlamydia, depression, or osteoporosis screening, for which randomized, placebo-controlled trials of treatment have been done.

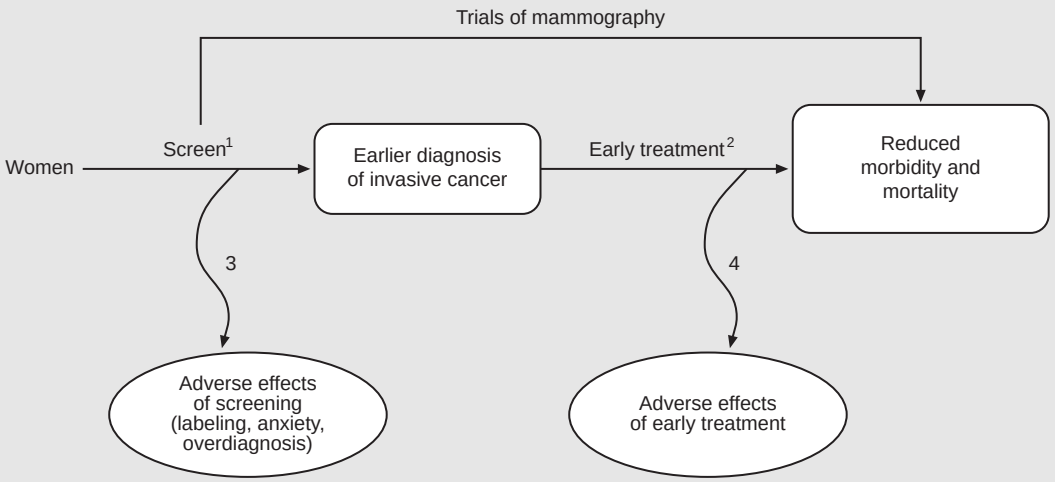
The threshold for sufficient evidence about efficacy also depends on the balance of benefits and harms. Because mammography technology, the timing and type of information provided to patients, and treatment approaches have changed over time, the adverse consequences of screening in current practice might be very different from those in the trials. Other sources of data must be used to estimate these consequences.

Identification and Selection of Articles

We identified controlled trials and meta-analyses by searching the Cochrane Controlled Trials Registry (all dates), as well as searching for recent publications in MEDLINE (January 1994 to December 2001). Other sources were a PREMEDLINE search (December 2001 through February 2002); the reference lists of previous reviews, commentaries, and meta-analyses[5,8,27,32,50,53,55,56,60,87,100–103]; the results of a broader search conducted for the systematic evidence review on which this article is based46; and suggestions from experts.

In the electronic searches, the terms *breast neoplasms* and *breast cancer* were combined with the terms *mammography* and *mass screening* and with terms for controlled or randomized trials to yield 954 citations. Titles and abstracts were reviewed to identify publications that were randomized, controlled trials of breast cancer screening and had a relevant clinical outcome (advanced breast cancer, breast cancer mortality, or all-cause mortality). In all, the searches identified 146 controlled trials, of which 132 were excluded at the title and abstract phase because they concerned promoting screening rather than the efficacy of mammography (Appendix Figure 2). Four of the remaining 12 trials were

Appendix Figure 1. Analytic framework



Numbers refer to actions and associations that were examined in the systematic review. Trials of mammography link screening to health outcomes, but do not address the intermediate steps (screening and early treatment) or harms (adverse effects of screening and early treatment). Arrows indicating screening and early treatment represent the intermediate steps in the casual chain linking screening with improved mortality and morbidity.

excluded. Two were randomized trials of screening with mammography that have not yet presented outcomes of mortality or advanced breast cancer.[104,105] The third was a controlled trial that reported a reduction in breast cancer mortality but was not randomized.[106,107] The fourth, the Malmo Prevention Study, was apparently a randomized trial of a variety of preventive interventions, including mammography.[108] It reported significantly fewer deaths from cancer among women younger than 40 at study entry but provided no information about the mammography protocol, referring readers to another randomized trial, the Malmo Mammographic Screening Program, for further information. We believe that the 2 trials were in fact separate and that the results of the Malmo Mammographic Screening Program probably do not include results for the 8,000 women who participated in the Malmo Prevention Study.

The remaining 8 randomized trials of mammography were conducted between 1963 and 1994. Four of these were Swedish studies: the Malmo, Kopparberg, Ostergotland, Stockholm, and Gothenburg studies. (Kopparberg and Ostergotland together are known as the Swedish Two-County Study.) The remaining studies were the Edinburgh study, the New York Health Insurance Plan (HIP) study, and the 2 Canadian National Breast Screening Studies (CNBSS-1 and CNBSS-2). Using the electronic searches and other sources, we retrieved the full text of 157 publications about these trials (these are listed in the bibliography accompanying the full systematic evidence review[46]). We also identified 10 previous systematic reviews of the trials. Seven of these concerned breast cancer mortality, and 3 addressed test performance.[36,37,45] The searches identified 3 nonrandomized, controlled trials[109–111] that are not included in the meta-analysis but are discussed in the larger report.[46] Two randomized trials of breast self-examination were identified and reviewed.

Two of the authors abstracted information about each randomized, controlled trial. We compiled an appendix consisting of detailed information about the patient population, design, potential flaws, missing information, and analysis conducted in each trial. For the primary end point of breast cancer mortality, we abstracted results for each reported

length of follow-up. Whenever possible, we abstracted data separately for participants by decade of age.

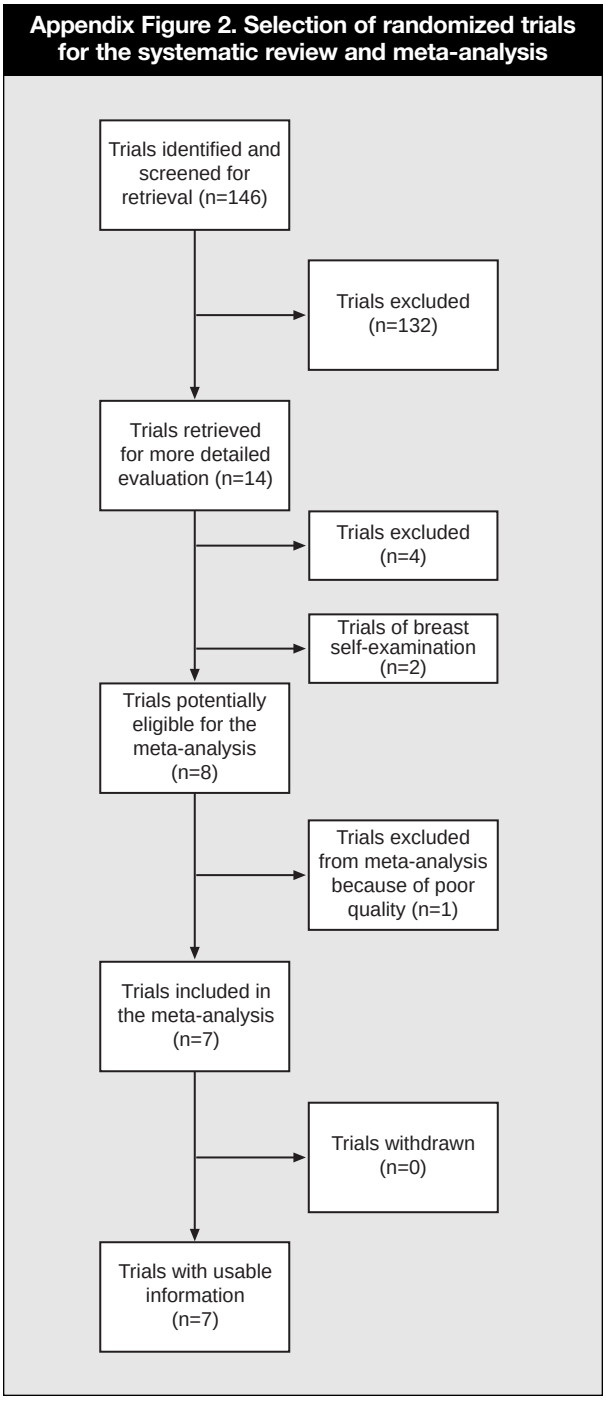
The randomized trials of screening provide little information about morbidity or the adverse effects of screening or treatment. A systematic review of adverse effects was beyond the scope of our review. In examining titles and abstracts, we obtained the full text of and reviewed recent articles reporting the frequency of false-positive results on screening mammography in the community and surveys of women's reactions to positive results on screening tests.

Assessment of Study Quality: General Approach

We used predefined criteria developed by the USPSTF to assess the internal validity of each study (Appendix Table 1).[9] Two authors rated each study as "good," "fair," or "poor," resolving disagreements by discussion among the authors after review of the data and of comments by 12 peer reviewers of earlier drafts of the report. We tried to apply the same standards to the mammography trials as we have applied to other prevention topics. We based our quality ratings on the entire set of publications from a trial rather than on individual articles.

The USPSTF criteria were designed to be adaptable to the circumstances of different clinical questions. Like other current systems to assess the quality of trials, the criteria are based as much as possible on empirical evidence of bias in relation to study characteristics. However, although the body of such evidence is growing, it does not permit a high degree of certainty about the importance of specific quality criteria in judging the mammography trials. This is because nearly all empirical evidence of the impact of bias on effect size examined drug treatment or other therapies, rather than screening.[112,113] Generalization of these findings to large, population-based trials of screening is not straightforward. In recognition of this fact, cancer screening literature from the 1970s emphasizes that design standards for conventional trials of treatment should not always be applied to cancer screening trials.[114]

Appendix Figure 2. Selection of randomized trials for the systematic review and meta-analysis



The quality of reporting of trials limits precision in critical appraisal.[115] This is a particular issue in the mammography screening trials, many of which were

Appendix Table 1. Criteria for grading the internal validity of individual studies

Randomized, controlled trials

- Clear definition of interventions
- All important outcomes considered
- Intention-to-treat analysis
- Initial assembly of comparable groups
  - adequate randomization, including first concealment and whether potential confounders were distributed equally among groups
  - Similar all-cause mortality among groups
- Maintenance of comparable groups (includes attrition, crossovers, adherence, contamination)
- Important differential loss to follow-up or overall high loss to follow-up
- Equal, reliable, and valid measurements (includes masking of outcome assessment)

Systematic Reviews

- Comprehensiveness of sources considered and search strategy used
- Standard appraisal of included studies
- Validity of conclusions
- Recency and relevance (especially important)

conducted in the 1960s and 1970s. Their methods were poorly described, which limits precision in critical appraisal. Although some reviewers have promoted extensive query of trial authors to fill in gaps in published articles, the reliability of such data, as well as the appropriate interpretation of query data that contradicts what has been published in multiauthored, peer-reviewed papers, is uncertain. Moreover, authors are often unable to provide clarifying information.[116]

Assessment of Study Quality: Application of Specific Criteria

All of the trials clearly defined interventions and co-interventions (CBE and BSE), all considered mortality outcomes, and all used intention-to-screen analysis. For this reason, the following received particular emphasis in judging the quality of the mammography trials: (1) initial assembly of comparable groups, (2) maintenance of comparable groups and minimization of differential or overall loss to follow-up, and (3) use of outcome measurements that were equal, reliable, and valid. As described below, we used a systematic approach to assess the flaws of the trials in each of these areas.

Initial Assembly of Comparable Groups

In the mammography trials, randomization was done individually or by clusters. Randomization of individuals is preferable because it is less likely to result in baseline differences among compared groups. In individually randomized trials, we classified allocation concealment as adequate, inadequate, or poorly described, according to the criteria used by Schulz and colleagues.[115] In a cluster-randomized trial, it is impossible to conceal the assignment of individual patients, and the importance of concealing the allocation of clusters is unclear. Accordingly, we placed more importance on concealment in individually randomized trials.

We rated the way in which each trial compared participants in the screened and control groups. To obtain the highest rating in this category, a trial had to obtain baseline data on possible covariates before randomization, and the distribution of these covariates had to be similar in screening and control groups. In a large, individually randomized trial, baseline differences in sociodemographic variables would suggest that randomization failed, especially if there were opportunities for subversion (that is, if allocation was not concealed).

This standard applies only if baseline data can be reliably collected in all patients in both groups. In several of the mammography screening trials, participants in the usual care group were followed passively, and there was no opportunity to collect baseline data from all of them. The decision not to contact each individual in the control group has logistic advantages and probably reduced contamination, but it limits comparison between the screened and control groups. Moreover, when clusters are used, some baseline differences in the compared groups are almost inevitable.

We evaluated whether the method of identifying clusters (for example, geographic areas, month or year of birth) was likely to result in bias and whether measures such as matching were used to reduce it. If bias in assigning clusters to intervention or control groups seemed likely, we considered this a major flaw that was enough to invalidate the findings and rated the study as poor. However, in contrast to individually randomized trials, we did not take small differences in the mean age of compared groups to be an indication that randomization failed to distribute more important confounders equally among the groups.

Several of the trials measured mortality rates from causes other than breast cancer to establish the comparability of the mammography and control groups. We recorded this information when it was available. Although comparable total mortality supports balanced randomization, it does not assure it. However, if there were dramatic differences in death from other causes, we considered it to be evidence that randomization failed.

Maintenance of Comparable Groups and Minimization of Differential or Overall Loss to Follow-up

Exclusions after randomization are considered to be a serious flaw in the execution of randomized trials, although empirical evidence of this bias is inconsistent.[112,113] Postrandomization exclusions were poorly described in several of the mammography trials and could have resulted in bias if the exclusions resulted in different levels of risk for death from breast cancer between the groups. In most of the mammography trials, however, exclusion of participants after randomization was an expected consequence of the protocol; some exclusion criteria, such as previous mastectomy, could not be applied to all participants before randomization because participants were not individually contacted. We examined the number of, reasons for, and methods for exclusion of participants after randomization. We based our rating on whether the methods used to ascertain patients were objective and consistent, not on the numbers of exclusions in the compared groups. Since ascertainment of clinical variables that might result in exclusion of a participant will be greater among intervention participants and is an expected consequence of the study design, we did not consider unequal numbers of excluded participants in the treatment and control groups after randomization to be definitive evidence of bias.

Use of Outcome Measurements That Were Equal, Reliable, and Valid (Including Masking of Outcome Assessment)

Over the duration of most of the trials, death from breast cancer (the primary end point) occurred in 2 to 9 per 1,000 participants. The relatively low numbers of events means that misclassification or biased exclusion of a few deaths could change the direction and statistical significance of the trial results. For this reason, selection of cases for review of cause of death on broad criteria, use of reliable sources of information to ascertain vital status (death certificates, medical records, autopsies, registries), and use of independent blinded review of the cause of death are important measures to prevent bias. We considered blinded review of deaths a requirement for a quality rating of fair or better.

Approach to Multiple Analyses

The mammography trials have been criticized for decades,[99,117–119] and the trialists have responded by conducting additional analyses intended to address these criticisms. In our assessment of quality, we took into account the results of these supplemental analyses. For example, the cluster-randomized trials have been criticized because they analyzed results using statistical methods appropriate only to individually randomized trials. However, an independent reanalysis using the correct statistical method found that the results were unchanged.[48] The Canadian trialists addressed criticisms that women who had palpable nodes might have been enrolled preferentially in the mammography group[120] by reanalyzing their data and showing that the exclusion of these participants did not affect the results.[22]

Data Synthesis

Four of the trials compared mammography alone with usual care, and 4 compared mammography plus CBE with usual care. Because of lack of certainty that CBE is effective, and in consultation with USPSTF members, we decided that these trials were qualitatively homogeneous. The homogeneity of the trials was also assessed by using the standard chi-square test. The *P* value was greater than 0.1, indicating the effect sizes estimated by the studies are homogeneous.

We conducted 2 meta-analyses to address 2 key questions posed by the USPSTF: (1) Does mammography reduce breast cancer mortality rates among women over a broad range of ages when compared with usual care? and (2) If so, does mammography reduce breast cancer mortality rates among women 40 to 49 years of age when compared with usual care? In the first analysis, we included all data from the 7 fair-quality trials, treating the 2 Canadian studies as 1 trial in participants 40 to 59 years of age. In the second analysis, we included the 6 fair-quality trials that reported results for women younger than 50.

We conducted each meta-analysis in 2 parts. First, using WinBUGS software, we constructed a 2-level Bayesian random-effects model to estimate the effect size from multiple data points for each study and to derive a pooled estimate of relative risk reduction and credible interval for a given length of follow-up.[11] The purpose of this analysis was to use repeated measures of the effect over time to estimate the relationship between length of follow-up and effect size. Appendix Table 2 shows the data we used in this analysis. Second, we pooled the most recent results of each trial to calculate the absolute and relative risk reduction, using the results of the first analysis to estimate the mean length of observation. Risks were modeled on the logit scale.

To model the relationship between length of follow-up and relative risk, a 2-level hierarchical model was used. The first level was the result of a trial at a given average or median follow-up time, x_{ij}, where *i* indexes the trial and *j* indexes the data point within a trial. The second level was the trial itself. The model allows for within-trial and between-trial variability. Specifically, the model was:

$\alpha^* \sim \text{Normal}(\cdot,\cdot)$

$\beta^* \sim \text{Normal}(\cdot,\cdot)$

$\alpha_i \sim \text{Normal}(\alpha^*, \sigma^2_\alpha)$

$\beta_i \sim \text{Normal}(\beta^*, \sigma^2_\beta)$

$\mu_{ij} = \alpha_i + \beta_i x_{ij} + \tau z_{ij}$

$\tau \sim \Gamma(\cdot,\cdot)$

$z_{ij} \sim \text{Normal}(0,1)$

$\log RR_{ij} \sim \text{Normal}(\mu_{ij}, s^2)$.

A global regression curve was estimated as log RR= $\alpha^* + \beta^* x$. The random effect was τz_{ij}. The model to estimate summary risk was:

$\#\ \text{deaths}_{\text{control},i} \sim \text{Binomial}(\pi_{\text{control},i}, n_{\text{control},i})$

$\#\ \text{deaths}_{\text{intervention},i} \sim \text{Binomial}(\pi_{\text{intervention},i}, n_{\text{intervention},i})$

$\text{logit}(\pi_{\text{control},i}) = \alpha + \tau z_i$

$\text{logit}(\pi_{\text{intervention},i}) = \alpha + \beta + \tau z_i$

$\alpha \sim \text{Normal}(\cdot,\cdot)$

$\beta^* \sim \text{Normal}(\cdot,\cdot)$

$\tau \sim \Gamma(\cdot,\cdot)$

Absolute risk difference was calculated as $\pi_{\text{control},i} - \pi_{\text{intervention},i}$. Relative risk was calculated as $\exp(\beta)$.

The models were estimated by using a Bayesian data analytic framework.[121] The data were analyzed by using WinBUGS,[11] which uses Gibbs sampling to simulate posterior probability distributions. Noninformative (proper) prior probability distributions were used: Normal(0, 10^6) and $\Gamma(0.001, 0.001)$. Five separate Markov chains with overdispersed initial values were used to generate draws from posterior distributions. Point estimates (mean) and 95% credible intervals (2.5 and 97.5 percentiles) were derived from the subsequent 5 ? 10,000 draws after reasonable convergence of the 5 chains was attained. The code to model the data in WinBUGS is available from the authors on request.

Appendix Table 2. Data used in the analysis*

				Intervention Group				Control Group				95% CI		
Study	Ref	Age	Mean FU	Deaths	Subjects	Life-years	Rate/ 10,000	Deaths	Subjects	Life-years	Rate/ 10,000	RR	Lower	Upper
CNBSS	Miller, unpublished data	40-49	13.0	105	25,214	282,606	**3.7**	108	25,216	282,575	**3.8**	**0.97**	0.74	1.27
	Miller et al, 1997[21]†	40-49	10.5	82	25,214	**264,747**	**3.1**	72	25,216	**264,768**	**2.7**	**1.14**	0.83	1.56
	Miller et al, 1992[12]	40-49	8.5	38	25,214	**214,319**	**1.8**	28	25,216	**214,336**	**1.3**	**1.36**	0.84	2.21
		40-59	13.0	212	44,925	**584,025**	**3.6**	213	44,910	**583,830**	**3.6**	**1.00**	**0.82**	**1.20**
		40-59	8.5	76	44,925	**381,863**	**2.0**	67	44,910	**381,735**	**1.8**	**1.13**	**0.82**	**1.57**
	Miller et al, 2000[20]	50-59	13.0	107	19,711	216,133	**5.0**	105	19,694	216,042	**4.9**	**1.02**	0.78	1.33
	Miller et al, 1992[13]	50-59	8.3	38	19,711	**163,601**	**2.3**	39	19,694	**163,460**	**2.4**	**0.97**	0.62	1.52
HIP	Shapiro, 1997[27]†	40-49	18.0	49	13,740	**247,320**	**2.0**	65	13,740	**247,320**	**2.6**	**0.75**	**0.52**	**1.09**
	Habbema et al, 1986[122]	40-49	14.0	64	13,740	**192,360**	**3.3**	82	13,740	**192,360**	**4.3**	**0.78**	**0.56**	**1.08**
	Shapiro et al, 1988[19]	40-49	10.0	39	13,740	**137,400**	**2.8**	51	13,740	**137,400**	**3.7**	**0.76**	**0.50**	**1.16**
	Shapiro et al, 1988[19]	40-49	5.0	19	13,740	**68,700**	**2.8**	20	13,740	**68,700**	**2.9**	**0.95**	**0.51**	**1.78**
	Shapiro et al, 1988[19]	40-64	18.0	126	30,245	**544,410**	**2.3**	163	30,245	**544,410**	**3.0**	**0.77**	**0.61**	**0.98**
	Shapiro et al, 1985[123]	40-64	16.0	236	30,239	**483,824**	**4.9**	281	30,756	**492,096**	**5.7**	**0.85**	**0.72**	**1.02**
	Habbema et al, 1986[122]	40-64	14.0	165	30,245	**423,430**	**3.9**	212	30,245	**423,430**	**5.0**	**0.78**	**0.64**	**0.95**
	Shapiro et al, 1988[19]	40-64	10.0	95	30,245	**302,450**	**3.1**	133	30,245	**302,450**	**4.4**	**0.71**	**0.55**	**0.93**
	Shapiro et al, 1988[19]	40-64	5.0	39	30,245	**151,225**	**2.6**	63	30,245	**151,225**	**4.2**	**0.62**	**0.42**	**0.92**
	Shapiro et al, 1988[19]	50-64	18.0	77	16,505	**297,090**	**2.6**	98	16,505	**297,090**	**3.3**	**0.79**	**0.58**	**1.06**
	Habbema et al, 1986[122]	50-64	14.0	101	16,505	**231,070**	**4.4**	130	16,505	**231,070**	**5.6**	**0.78**	**0.60**	**1.01**
	Shapiro et al, 1988[19]	50-64	10.0	56	16,505	**165,050**	**3.4**	82	16,505	**165,050**	**5.0**	**0.68**	**0.49**	**0.96**
	Shapiro et al, 1988[19]	50-64	5.0	20	16,505	**82,525**	**2.4**	43	16,505	**82,525**	**5.2**	**0.47**	**0.27**	**0.79**
Gothenburg	Bjurstam et al, 1997[24]†	39-49	**11.8**	18	11,724	138,402	**1.3**	40	14,217	168,025	**2.4**	0.55	0.31	0.96
	Nystrom et al, 2002[23]	40-49	**12.7**	22	10,888	138,000	**1.6**	46	13,203	167,000	**2.8**	**0.58**	0.35	0.96
	Larsson et al, 1997[50]	40-49	**9.8**	16	10,821	106,000	**1.5**	33	13,101	129,000	**2.6**	**0.59**	0.33	1.06
	Nystrom et al, 2002[23]	40-59	**12.8**	62	21,000	268,000	**2.3**	113	29,200	373,000	**3.0**	**0.76**	0.56	1.04
	Nystrom et al, 1993[32]	40-59	**6.3**	27	20,724	129,000	**2.1**	47	28,809	181,000	**2.6**	0.86	0.54	1.37
	Nystrom et al, 2002[23]	50-59	**12.9**	40	10,112	130,000	**3.1**	67	15,997	206,000	**3.3**	0.94	0.62	1.43
Stockholm	Nystrom et al, 2002[23]	40-49	**14.3**	34	14,303	203,000	**1.7**	13	8,021	117,000	**1.1**	1.52	0.80	2.88
	Frisell and Lidbrink, 1997[124]†	40-49	**11.9**	24	14,842	173,866	**1.4**	12	7,103	87,826	**1.4**	1.08	0.54	2.17
	Larsson et al, 1997[50]	40-49	**11.5**	23	14,185	162,000	**1.4**	10	7,985	94,000	**1.1**	1.34	0.64	2.80
	Frisell et al, 1991[125]	40-49	**7.2**	16	14,375	99,155	**1.6**	8	7,103	54,446	**1.5**	1.09	0.40	3.00
	Frisell et al, 1997[17]	40-64	**11.8**	66	40,318	473,153	**1.4**	45	19,943	239,460	**1.9**	**0.74**	0.50	1.10
	Frisell et al, 1991[125]	40-64	**7.1**	39	39,164	270,247	**1.4**	30	19,943	147,373	**2.0**	**0.71**	0.40	1.20
	Nystrom et al, 2002[23]	40-65	**13.8**	82	39,139	535,000	**1.5**	50	20,978	296,000	**1.7**	**0.91**	0.65	1.27
	Nystrom et al, 1993[32]	40-65	**7.6**	53	38,525	287,000	**1.8**	40	20,651	164,000	**2.4**	**0.80**	0.53	1.22
	Nystrom et al, 2002[23]	50-59	**13.7**	25	15,946	217,000	**1.2**	24	8,421	118,000	**2.0**	0.56	0.32	0.97
	Frisell et al, 1997[17]	50-64	**11.8**	42	25,476	299,287	**1.4**	33	12,840	151,634	**2.2**	0.62	0.38	1.00
	Frisell et al, 1991[125]	50-64	**7.0**	23	24,789	171,092	**1.3**	22	12,840	92,927	**2.4**	**0.57**	0.30	1.10

Continued on page 203

Appendix Table 2. Data used in the analysis* *(continued)*

				Intervention Group				Control Group				95% CI		
Study	**Ref**	**Age**	**Mean FU**	**Deaths**	**Subjects**	**Life-years**	**Rate/ 10,000**	**Deaths**	**Subjects**	**Life-years**	**Rate/ 10,000**	**RR**	**Lower**	**Upper**
Malmö I + II	Nystrom et al, 2002[23]	43-49	**13.3**	53	13,568	184,000	**2.9**	66	12,279	160,000	**4.1**	**0.73**	**0.51**	**1.04**
	Andersson and Janzon, 1997[15]†	43-49	**12.0**	57	13,528	165,596	**3.4**	78	12,242	144,036	**5.4**	**0.64**	0.45	0.89
	Nystrom et al, 2002[23]	43-70	**15.3**	190	30,669	473,000	**4.0**	231	29,407	448,000	**5.2**	**0.79**	**0.65**	**0.96**
	Nystrom et al, 2002[23]	45-49	**18.0**	24	3,987	71,000	**3.4**	33	4,067	74,000	**4.5**	**0.74**	**0.44**	**1.25**
	Larsson et al, 1997[50]	45-49	**15.4**	15	3,945	61,000	**2.5**	23	4,017	62,000	**3.7**	0.67	0.35	1.27
	Nystrom et al, 2002[23]	45-54	**18.2**	71	8,673	158,000	**4.5**	78	8,311	151,000	**5.2**	**0.87**	0.63	1.20
	Andersson et al, 1988[25]	45-54	**9.0**	28	7,981	71,775	**3.9**	22	8,082	72,635	**3.0**	**1.29**	0.74	2.25
	Andersson et al, 1988[25]	45-69	**8.8**	63	21,088	186,297	**3.4**	66	21,195	187,016	**3.5**	**0.96**	0.68	1.35
	Nystrom et al, 2002[23]	45-70	**17.1**	161	21,088	360,000	**4.5**	198	21,195	362,000	**5.5**	**0.82**	0.67	1.00
	Nystrom et al, 1993[32]	45-70	**11.5**	87	20,695	239,000	**3.6**	108	20,783	240,000	**4.5**	**0.81**	0.62	1.07
	Nystrom et al, 2002[23]	50-70	**16.9**	137	17,101	289,000	**4.7**	165	17,128	288,000	**5.7**	**0.83**	**0.66**	**1.04**
	Nystrom et al, 2002[23]	55-64	**17.2**	63	8,194	141,000	**4.5**	83	8,679	149,000	**5.6**	**0.80**	0.57	1.12
	Andersson et al, 1988[25]	55-69	**8.7**	35	13,107	114,522	**3.1**	44	13,113	114,381	**3.8**	**0.79**	0.51	1.24
	Nystrom et al, 2002[23]	55-70	**16.3**	90	12,415	202,000	**4.5**	120	12,884	211,000	**5.7**	**0.78**	**0.59**	**1.02**
2-county, Kopparberg	Tabár et al, 2000[26]	40-49	**17.3**	NR	NR	NR	NR	NR	NR	NR	NR	0.76	0.42	1.40
	Tabár et al, 1995[16]	40-49	13.0	22	9,582	**124,566**	**1.8**	16	5,031	**65,403**	**2.4**	0.73	0.37	1.41
	Tabár et al, 1989[28]	40-49	7.9	13	9,582	**75,698**	**1.7**	9	5,031	**39,745**	**2.3**	**0.76**	**0.32**	**1.77**
	Tabár et al, 1985[35]	40-49	6.0	8	9,625	**57,750**	**1.4**	3	5,053	**30,318**	**1.0**	**1.40**	**0.37**	**5.28**
	Tabár et al, 2000[26]	40-74	**17.3**	152	NR	672,482	**2.3**	121	NR	326,091	**3.7**	**0.61**	NR	NR
	Tabár et al, 1995[16]	40-74	13.0	126	38,589	**501,657**	**2.5**	104	18,582	**241,566**	**4.3**	0.60	0.46	0.79
	Tabár et al, 1989[28]	40-74	7.9	77	38,589	**304,853**	**2.5**	58	18,582	**146,798**	**4.0**	**0.64**	**0.46**	**0.90**
	Tabár et al, 1985[35]	40-74	6.0	51	39,051	**234,306**	**2.2**	39	18,846	**113,076**	**3.4**	**0.63**	**0.42**	**0.96**
	Tabár et al, 2000[26]	50-59	**17.3**	NR	NR	NR	NR	NR	NR	NR	NR	0.46	0.30	0.71
	Tabár et al, 1995[16]	50-59	13.0	34	11,728	**152,464**	**2.2**	34	5,557	**72,241**	**4.7**	0.48	0.29	0.77
	Tabár et al, 1989[28]	50-59	7.9	20	9,582	**75,698**	**2.6**	20	5,031	**39,745**	**5.0**	**0.53**	**0.28**	**0.98**
	Tabár et al, 1995[16]	50-74	13.0	104	29,007	**377,091**	**2.8**	88	13,551	**176,163**	**5.0**	0.58	0.43	0.78
	Tabár et al, 1989[28]	50-74	7.9	64	29,007	**229,155**	**2.8**	49	13,551	**107,053**	**4.6**	**0.61**	**0.42**	**0.89**
	Tabár et al, 1985[35]	50-74	6.0	43	29,426	**176,556**	**2.4**	36	13,793	**82,758**	**4.4**	**0.56**	**0.36**	**0.87**
2-county, Östergötland	Tabár et al, 2000[26]	40-49	**17.3**	NR	NR	NR	NR	NR	NR	NR	NR	1.06	0.65	1.76
	Nystrom et al, 2002[23]	40-49	**16.8**	31	10,285	172,000	**1.8**	30	10,459	176,000	**1.7**	1.05	0.64	1.71
	Tabár et al, 1995[16]	40-49	13.0	23	10,262	**133,406**	**1.7**	23	10,573	**137,449**	**1.7**	1.02	0.52	1.99
	Tabár et al, 1989[28]	40-49	7.9	15	10,262	**81,070**	**1.9**	15	10,573	**83,527**	**1.8**	**1.03**	**0.50**	**2.11**
	Tabár et al, 1985[35]	40-49	6.0	8	10,312	**61,872**	**1.3**	7	10,625	**63,750**	**1.1**	**1.18**	**0.43**	**3.25**
	Tabár et al, 2000[26]	40-74	**17.3**	167	NR	660,242	**2.5**	213	NR	**643,696**	**3.3**	**0.76**	NR	NR
	Nystrom et al, 2002[23]	40-74	**15.2**	177	38,942	589,000	**3.0**	190	37,675	**572,000**	**3.3**	**0.90**	0.73	1.11
	Tabár et al, 1995[16]	40-74	13.0	135	38,491	**500,383**	**2.7**	173	37,403	**486,239**	**3.6**	0.78	0.60	1.01

Continued on page 204

Appendix Table 2. Data used in the analysis* *(continued)*

				Intervention Group				Control Group				95% CI		
Study	**Ref**	**Age**	**Mean FU**	**Deaths**	**Subjects**	**Life-years**	**Rate/ 10,000**	**Deaths**	**Subjects**	**Life-years**	**Rate/ 10,000**	**RR**	**Lower**	**Upper**
	Tabár et al, 1989[28]	40-74	7.9	83	38,491	**304,079**	**2.7**	109	37,403	**295,484**	**3.7**	**0.74**	**0.56**	**0.98**
	Tabár et al, 1985[35]	40-74	6.0	36	39,034	**234,204**	**1.5**	47	37,936	**227,616**	**2.1**	**0.74**	**0.48**	**1.15**
	Tabár et al, 2000[26]	50-59	**17.3**	NR	NR	NR	**NR**	NR	NR	NR	**NR**	0.76	0.53	1.10
	Nystrom et al, 2002[23]	50-59	**16.1**	53	12,011	194,000	**2.7**	54	11,495	185,000	**2.9**	**0.94**	0.66	1.35
	Tabár et al, 1995[16]	50-59	13.0	44	11,757	**152,841**	**2.9**	51	11,248	**146,224**	**3.5**	0.85	0.52	1.38
	Tabár et al, 1989[28]	50-59	7.9	25	11,757	**92,880**	**2.7**	34	11,248	**88,859**	**3.8**	**0.70**	**0.42**	**1.18**
	Nystrom et al, 2002[23]	50-74	**14.9**	146	28,657	**417,000**	**3.5**	160	25,920	**396,000**	**4.0**	**0.83**	**0.66**	**1.03**
	Tabár et al, 1995[16]	50-74	13.0	112	28,229	**366,977**	**3.1**	150	26,830	**348,790**	**4.3**	0.73	0.56	0.97
	Tabár et al, 1989[28]	50-74	7.9	68	28,229	**223,009**	**3.0**	94	26,830	**211,957**	**4.4**	**0.69**	**0.50**	**0.94**
	Tabár et al, 1985[35]	50-74	6.0	28	28,722	**172,332**	**1.6**	40	27,311	**163,866**	**2.4**	**0.67**	**0.41**	**1.08**
Kopparberg + Östergötland	Tabár et al, 1995[16]	40-49	13.0	45	19,844	**257,972**	**1.7**	39	15,604	**202,852**	**1.9**	0.87	0.54	1.41
	Tabár et al, 1989[28]	40-49	7.9	28	19,844	**156,768**	**1.8**	24	15,604	**123,272**	**1.9**	0.92	0.52	1.60
	Tabár et al, 1989[28]	40-49	7.9	28	19,844	**156,768**	**1.8**	24	15,604	**123,272**	**1.9**	**0.92**	**0.53**	**1.58**
	Tabár et al, 1985[35]	40-49	6.0	16	19,937	**119,622**	**1.3**	10	15,678	**94,068**	**1.1**	**1.26**	0.56	2.84
	Tabár et al, 2000[26]	40-74	**17.3**	319	77,080	1,332,724	**2.4**	334	55,985	969,787	**3.4**	0.68	0.59	0.80
	Tabár et al, 1995[16]	40-74	**12.5**	269	77,080	965,405	**2.8**	277	55,985	701,207	**4.0**	0.69	0.57	0.84
	Tabár et al, 1989[28]	40-74	7.9	160	77,080	**608,932**	**2.6**	167	55,985	**442,282**	**3.8**	**0.70**	**0.56**	**0.86**
	Tabár et al, 1985[35]	40-74	6.0	87	78,085	**468,510**	**1.9**	86	56,782	**340,692**	**2.5**	0.69	0.51	0.92
	Tabár et al, 1995[16]	50-59	13.0	78	23,485	**305,305**	**2.6**	85	16,805	**218,465**	**3.9**	0.66	0.46	0.93
	Tabár et al, 1989[28]	50-59	7.9	45	23,485	**185,532**	**2.4**	54	16,805	**132,760**	**4.1**	**0.60**	**0.40**	**0.89**
	Tabár et al, 1995[16]	50-74	13.0	224	57,236	**744,068**	**3.0**	238	55,985	**727,805**	**3.3**	0.66	0.54	0.81
	Tabár et al, 1989[28]	50-74	7.9	132	57,236	**452,164**	**2.9**	143	40,381	**319,010**	**4.5**	**0.65**	**0.51**	**0.83**
	Tabár et al, 1985[35]	50-74	6.0	71	58,148	**348,888**	**2.0**	76	41,104	**246,624**	**3.1**	0.61	0.44	0.84
Edinburgh	Alexander et al, 1999[18]	45-49	**12.2**	47	11,479	139,868	**3.4**	53	10,267	126,413	**4.2**	0.75	0.48	1.18
	Alexander, 1997[126]†	45-49	**12.2**	46	NR	139,871	**3.3**	52	NR	126,417	**4.1**	0.88	0.55	1.41
	Alexander et al, 1994[127]	45-49	**8.5**	25	11,505	97,206	**2.6**	31	10,269	88,766	**3.5**	0.78	0.46	1.31
	Roberts et al, 1990[128]	45-49	**6.9**	13	5,913	40,851	**3.2**	13	5,810	40,009	**3.2**	**0.98**	NR	NR
	Alexander et al, 1999[18]	45-64	**13.0**	156	22,926	301,155	**5.2**	167	21,342	276,363	**6.0**	0.79	0.60	1.02
	Alexander et al, 1994[127]	45-64	**9.5**	96	22,944	219,215	**4.4**	106	21,344	201,821	**5.3**	0.82	0.61	1.11
	Roberts et al, 1990[128]	45-64	**6.8**	68	23,226	157,946	**4.3**	76	21,904	147,854	**5.1**	0.83	0.58	1.18
	Alexander et al, 1999[18]	50-64	**12.9**	129	17,149	222,393	**5.8**	134	15,748	200,637	**6.7**	**0.87**	NR	NR
	Alexander et al, 1994[127]	50-64	**9.4**	79	17,149	162,465	**4.9**	85	15,748	147,233	**5.8**	0.85	0.62	1.15
	Roberts et al, 1990[128]	50-64	**6.7**	55	17,313	117,095	**4.7**	63	16,094	107,845	**5.8**	**0.80**	0.54	1.17

*Numbers in bold are calculated from data in the spreadsheet; all other numbers were taken from publications.

†Used in reference 30.

Note: CI indicates confidence interval; CNBSS, Canadian National Breast Screening Study; FU, follow-up; HIP, Health Insurance Plan of greater New York; NR, not reported; RR, relative risk.

Peer Review and Revisions

Our review began early in 2000. A first draft was presented to the USPSTF in December 2000. Throughout 2001, the manuscript underwent extensive critical review by a broad range of experts. Subsequent versions were reviewed by the USPSTF in September 2001 and in January 2002.

Note: This manuscript is based on a longer systematic evidence review that was reviewed by outside experts and representatives of professional societies.

Acknowledgments: The authors thank Stephanie Detlefsen, MD, for her contribution to this evidence review and David Atkins, MD, MPH, from the Agency for Healthcare Research and Quality and members of the U.S. Preventive Services Task Force for their comments on earlier versions of the review. We also thank Kathryn Pyle Krages, AMLS, MA, Susan Carson, MPH, Patty Davies, MS, Susan Wingenfeld, and Jim Wallace for their help with preparation of the manuscript and the full systematic evidence review.

Grant Support: This study was conducted by the Oregon Health & Science University Evidence-based Practice Center under contract to the Agency for Healthcare Research and Quality (contract no. 290-97-0018, task order no. 2), Rockville, Maryland.

References

1. American Cancer Society. Cancer facts and figures, 2001. Available at: http://www.cancer.org/downloads/STT/F&F2001.pdf.
2. Gail M, Brinton, LA, Byar, DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst.* 1989;81:1879-1886.
3. Colditz GA, Willett WC, Hunter DJ, et al. Family history, age, and risk of breast cancer. Prospective data from the Nurses' Health Study [published erratum appears in *JAMA.* 1993;270(13):1548]. JAMA. 1993;270(3):338-343.
4. Seidman H, Stellman SD, Mushinski MH. A different perspective on breast cancer risk factors: some implications of the nonattributable risk. *Cancer.* 1982;32:301-312.
5. Strax P. Mass screening of asymptomatic women. In: Ariel IM, Cleary J, eds. *Breast Cancer: Diagnosis and Treatment.* New York: McGraw-Hill; 1987:145-151.
6. U.S. Preventive Services Task Force. *Guide to Clinical Preventive Services.* 2nd ed. Washington, DC: Office of Disease Prevention and Health Promotion; 1996.
7. Sirovich BE, Sox HC Jr. Breast cancer screening. *Surg Clin North Am.* 1999;79(5):961-990.
8. Olsen O, Gotzsche PC. Cochrane review on screening for breast cancer with mammography. *Lancet.* 2001;358(9290):1340-1342.
9. Harris RP, Helfand M, Woolf SH, et al. Methods of the third U.S. Preventive Services Task Force. *Am J Prev Med.* 2001;20(suppl 3):21-35.
10. Harrison TR. Breast cancer. In: Fauci AS, ed. *Principles of Internal Medicine.* 14th ed. New York: McGraw Hill; 1998:564-567.
11. *WinBUGS Version 1.2 User Manual.* Cambridge, England: MRC Biostatistics Unit; 1999.
12. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years [published erratum appears in *CMAJ.* 1993;148(5):718]. *CMAJ.* 1992;147:1459-1476.
13. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years [published erratum appears in *CMAJ.* 1993;148(5):718]. *CMAJ.* 1992;147:1477-1488.
14. Bjurstam N, Bjorneld L, Duffy SW, et al. The Gothenburg Breast Cancer Screening Trial: preliminary results on breast cancer mortality for women aged 39-49 [letter]. *J Natl Cancer Inst* Monogr. 1997;22:53-55.
15. Andersson I, Janzon L. Reduced breast cancer mortality in women under age 50: updated results from the Malmo Mammographic Screening Program. *J Natl Cancer Inst* Monogr. 1997;22:63-67.
16. Tabar L, Fagerberg G, Chen HH, et al. Efficacy of breast cancer screening by age: new results from the Swedish Two-County Trial. *Cancer.* 1995;75:2507-2517.
17. Frisell J, Lidbrink E, Hellstrom L, Rutqvist LE. Followup after 11 years: update of mortality results in the Stockholm mammographic screening trial. *Breast Cancer Res Treat.* 1997;45:263-270.

18. Alexander FE, Anderson TJ, Brown HK, et al. 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *Lancet.* 1999;353(9168):1903-1908.

19. Shapiro S, Venet W, Strax P, Venet L. Current results of the breast cancer screening randomized trial: the Health Insurance Plan (HIP) of greater New York study. In: Day NE, Miller AB, eds. *Screening for Breast Cancer.* Toronto: Hans Huber; 1988:3-15.

20. Miller AB, To T, Baines CJ, Wall C. The 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(18):1490-1499.

21. Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study: update on breast cancer mortality. *J Natl Cancer Inst Monogr.* 1997;22:37-41.

22. Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study - I, breast cancer mortality after 11-16 years of follow-up in women age 40-49. *Ann Intern Med.* 2002;137:305-312.

23. Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet.* 2002;359:909-19.

24. 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(11):2091-2099.

25. Andersson I, Aspegren K, Janzon L, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. *BMJ*. 1988;297:943-948.

26. Tabar L, Vitak B, Chen HH, et al. The Swedish Two-County Trial twenty years later: updated mortality results and new insights from long-term follow-up. *Radiol Clin North Am.* 2000;38(4):625-651.

27. Shapiro S. Periodic screening for breast cancer: the HIP Randomized Controlled Trial. Health Insurance Plan. *J Natl Cancer Inst Monogr.* 1997;22:27-30.

28. Tabar L, Fagerberg G, Duffy SW, Day NE. The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. *J Epidemiol Community Health.* 1989;43:107-114.

29. Black WC, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst .* 2002;94(3):167-173.

30. Berry DA. Benefits and risks of screening mammography for women in their forties: a statistical appraisal. *J Natl Cancer Inst.* 1998;90(19):1431-1439.

31. Sjonell G, Stahle L. Halsokontroller med mammografi minskar inte dodlighet i brostcancer. [Mammographic screening does not reduce breast cancer mortality]. *Lakartidningen.* 1999;96:904-905; 908-913.

32. Nystrom L, Rutqvist LE, Wall S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials [published erratum appears in Lancet. 1993;342(8883):1372]. *Lancet.* 1993;341(8851):973-978.

33. Nystrom L, Larsson LG, Wall S, et al. An overview of the Swedish randomised mammography trials: total mortality pattern and the representivity of the study cohorts. *J Med Screen.* 1996;3:85-87.

34. Tabar L, Gad A, Holmberg L, Ljungquist U. Significant reduction in advanced breast cancer.: results of the first seven years of mammography screening in Kopparberg, Sweden. *Diagn Imaging Clin Med.* 1985;54:158-164.

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

36. Mushlin AI, Kouides RW, Shapiro DE. Estimating the accuracy of screening mammography: a meta-analysis. *Am J Prev Med.* 1998;14(2):143-153.

37. Shen Y, Zelen M. Screening sensitivity and sojourn time from breast cancer early detection clinical trials: mammograms and physical examinations. *J Clin Oncol.* 2001;19(15):3490-3499.

38. Laya MB, Larson EB, Taplin SH, White E. Effect of estrogen replacement therapy on the specificity and sensitivity of screening mammography. *J Natl Cancer Inst.* 1996;88(10):643-649.

39. Greendale GA, Reboussin BA, Sie A, et al. Effects of estrogen and estrogen-progestin on mammographic parenchymal density. Postmenopausal Estrogen/Progestin Interventions (PEPI) Investigators. *Ann Intern Med.* 1999;130:262-269.

40. Marugg RC, van der Mooren MJ, Hendriks JH, Rolland R, Ruijs SH. Mammographic changes in postmenopausal women on hormonal replacement therapy. *Eur Radiol.* 1997;7(5):749-755.

41. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. *JAMA.* 1996;276(1):33-38.

42. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Likelihood ratios for modern screening mammography: risk of breast cancer based on age and mammographic interpretation. *JAMA.* 1996;276:39-43.

43. Eddy D. Screening for breast cancer. *Ann Intern Med.* 1989;111:389-399.

44. Lidbrink E, Elfving J, Frisell J, Jonsson E. Neglected aspects of false positive findings of mammography in breast cancer screening: analysis of false positive cases from the Stockholm trial. *BMJ.* 1996;312(7026):273-276.

45. Fletcher S, Black W, Harris R, Rimer BK, Shapiro S. Report of the International Workshop on Screening for Breast Cancer. *J Natl Cancer Inst.* 1993;85:1644-1656.

46. Humphrey L, Helfand M. *Screening for Breast Cancer.* Systematic Evidence Review No. 15 (Prepared by the Oregon Health & Science University Evidence-based Practice Center under Contract No. 290-97-0018). Rockville, MD: Agency for Healthcare Research and Quality. September 2002. (Available on the AHRQ Web site at: www.ahrq.gov/clinic/serfiles.htm).

47. Kerlikowske K, Grady D, Barclay J, Sickles EA, Eaton A, Ernster V. Positive predictive value of screening mammography by age and family history of breast cancer. *JAMA.* 1993;270(20):2444-2450.

48. Glasziou PP, Woodward AJ, Mahon CM. Mammographic screening trials for women aged under 50: a quality assessment and meta-analysis. *Med J Aust.* 1995;162(12):625-629.

49. Miller AB, Baines CJ, To T, Wall C. Screening mammography re-evaluated [letter]. *Lancet.* 2000;355(9205):747.

50. Larsson LG, Andersson I, Bjurstam N, et al. Updated overview of the Swedish Randomized Trials on Breast Cancer Screening with Mammography: age group 40-49 at randomization. *J Natl Cancer Inst Monogr.* 1997:57-61.

51. Cox B. Variation in the effectiveness of breast screening by year of follow-up. *J Natl Cancer Inst Monogr.* 1997(22):69-72.

52. Elwood JM, Cox B, Richardson AK. The effectiveness of breast cancer screening by mammography in younger women. *Online J Curr Clin Trials* [serial online]. 1993;doc 32.

53. Glasziou P, Irwig L. The quality and interpretation of mammographic screening trials for women ages 40-49. *J Natl Cancer Inst Monogr.* 1997(22):73-77.

54. Glasziou PP. Meta-analysis adjusting for compliance: the example of screening for breast cancer. *J Clin Epidemiol.* 1992;45(11):1251-1256.

55. Hendrick RE, Smith RA, Rutledge JH III, Smart CR. Benefit of screening mammography in women aged 40-49: a new meta-analysis of randomized controlled trials. *J Natl Cancer Inst Monogr.* 1997(22):87-92.

56. Smart CR, Hendrick RE, Rutledge JH III, Smith RA. Benefit of mammography screening in women ages 40 to 49 years: current evidence from randomized controlled trials [published erratum appears in Cancer. 1995;75(11)]. *Cancer.* 1995;75(7):1619-1626.

57. Kerlikowske K, Grady D, Ernster V. Benefit of mammography screening in women ages 40-49 years: current evidence from randomized controlled trials [letter]. *Cancer.* 1995;76(9):1679-1681.

58. Kerlikowske K. Efficacy of screening mammography among women aged 40 to 49 years and 50 to 69 years: comparison of relative and absolute benefit. *J Natl Cancer Inst Monogr.* 1997(22):79-86.

59. Tabar L, Fagerberg G, Chen HH, Duffy SW, Gad A. Screening for breast cancer in women aged under 50: mode of detection, incidence, fatality, and histology. *J Med Screen.* 1995;2:94-98.

60. Larsson LG, Nystrom L, Wall S, et al. The Swedish randomised mammography screening trials: analysis of their effect on the breast cancer related excess mortality. *J Med Screen.* 1996;3(3):129-132.

61. Barton MB, Harris R, Fletcher SW. Does this patient have breast cancer? The screening clinical breast examination: should it be done? How? *JAMA.* 1999;282(13):1270-1280.

62. Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast

examinations. *N Engl J Med.* 1998;338(16):1089-1096.

63. Shapiro S. Evidence on screening for breast cancer from a randomized trial. *Cancer* (Philadelphia). 1977;39:2772-2782.

64. Anonymous. 16-year mortality from breast cancer in the UK Trial of Early Detection of Breast Cancer. *Lancet.* 1999;353(9168):1909-1914.

65. Baines CJ. The Canadian National Breast Screening Study: responses to controversy. *Womens Health Issues.* 1992;2(4):206-211.

66. Richert-Boe KE, Humphrey LL. Screening for cancers of the cervix and breast. *Arch Intern Med.* 1992;152(12):2405-2411.

67. Semiglazov VF, Moiseyenko VM, Bavli JL, et al. The role of breast self-examination in early breast cancer detection (results of the 5-years USSR/WHO randomized study in Leningrad). *Eur J Epidemiol.* 1992;8(4):498-502.

68. Semiglazov VF, Moiseenko VM, Manikhas AG, et al. [Interim results of a prospective randomized study of self-examination for early detection of breast cancer (Russia/St.Petersburg/WHO)]. The role of breast self-examination in early breast cancer detection (results of the 5-years USSR/WHO randomized study in Leningrad). *Vopr Onkol.* 1999;45(3):265-271.

69. Thomas DB, Gao DL, Self SG, et al. Randomized trial of breast self-examination in Shanghai: methodology and preliminary results. *J Natl Cancer Inst*. 1997;89(5):355-365.

70. Pisano ED, Earp J, Schell M, Vokaty K, Denham A. Screening behavior of women after a false-positive mammogram. *Radiology.* 1998;208(1):245-249.

71. Ekeberg O, Skjauff H, Karesen R. Screening for breast cancer is associated with a low degree of psychological distress. *The Breast.* 2001;10(1):20-24.

72. Lampic C, Thurfjell E, Bergh J, Sjoden PO. Short- and long-term anxiety and depression in women recalled after breast cancer screening. *Eur J Cancer.* 2001;37(4):463-469.

73. Meystre-Agustoni G, Pacccaud F, Jeannin A, Dubois-Arber F. Anxiety in a cohort of Swiss women participating in a mammographic screening programme. *J Med Screen.* 2001;8(4):213-219.

74. Lerman C, Track B, Rimer BK, Boyce A, Jepson C, Engstrom PF. Psychological and behavioral implications of abnormal mammograms. *Ann Int Med.* 1991;114:657-661.

75. Rimer BK, Bluman LG. The psychosocial consequences of mammography. *J Natl Cancer Inst Monogr.* 1997(22):131-138.

76. Olsson P, Armelius K, Nordahl G, Lenner P, Westman G. Women with false positive screening mammograms: how do they cope? *J Med Screen.* 1999;6:89-93.

77. O'Sullivan I, Sutton S, Dixon S, Perry N. False positive results do not have a negative effect on reattendance for subsequent breast screening. *J Med Screen.* 2001;8:145-148.

78. Lipkus IM, Kuchibhatla M, McBride C, et al. Relationships among Breast Cancer Percieved Absolute Risk, Comparative Risk, and Worries. *Cancer Epidemiol Biomarkers Prev.* 2000;9:973-975.

79. Burman ML, Taplin SH, Herta DF, Elmore JG. Effect of false-positive mammograms on interval breast cancer screening in a health maintenance organization. *Ann Intern Med.* 1999;131(1):1-6.

80. Schwartz LM, Woloshin S, Sox HC, Fischhoff B, Welch HG. U.S. women's attitudes to false positive mammography results and detection of ductal carcinoma in situ: cross sectional survey. *BMJ.* 2000;320(7250):1635-1640.

81. Harstall C. *Mammography Screening: Mortality Rate Reduction and Screening Interval.* Edmonton: Alberta Heritage Foundation for Medical Research; 2000.

82. Ernster VL, Barclay J. Increases in ductal carcinoma in situ (DCIS) of the breast in relation to mammography: a dilemma. *J Natl Cancer Inst Monogr.* 1997(22):151-156.

83. Mattsson A, Leitz W, Rutqvist LE. Radiation risk and mammographic screening of women from 40 to 49 years of age: effect on breast cancer rates and years of life. *Br J Cancer.* 2000;82(1):220-226.

84. Feig SA, Hendrick RE. Radiation risk from screening mammography of women aged 40-49 years. *J Natl Cancer Inst Monogr.* 1997(22):119-124.

85. Swift M, Morrell D, Massey RB, Chase CL. Incidence of cancer in 161 families affected by ataxia-telangiectasia. *N Engl J Med.* 1991;325(26):1831-1836.

86. U.S. Preventive Services Task Force. *Guide to Clinical Preventive Services.* Baltimore, MD: Williams and Wilkins; 1989.

87. Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography: a meta-analysis. *JAMA.* 1995;273(2):149-154.

88. Fletcher SW. Why question screening mammography for women in their forties? *Radiol Clin North Am.* 1995;33(6):1259-1271.

89. Tabar L, Duffy SW, Chen HH. Re: Quantitative interpretation of age-specific mortality reductions from the Swedish Breast Cancer-Screening Trials. *J Natl Cancer Inst.* 1996;88(1):52-55.

90. McPherson K, Steel CM, Dixon JM. ABC of breast diseases: breast cancer-epidemiology, risk factors, and genetics. *BMJ*. 2000;321(7261):624-628.

91. Ries LAG, Eisner MP, Kosary CL, et al. SEER *Cancer Statistics Review,* 1973-1997. Bethesda, MD: National Cancer Institute; 2000. NIH Pub. No. 00-2789.

92. Kopans DB. An overview of the breast cancer screening controversy. *J Natl Cancer Inst Monogr.* 1997(22):1-3.

93. Satariano WA, Ragland DR. The effect of comorbidity on 3-year survival of women with primary breast cancer. *Ann Intern Med.* 1994;120(2):104-110.

94. Kerlikowske K, Salzmann P, Phillips KA, Cauley JA, Cummings SR. Continuing screening mammography in women aged 70 to 79 years: impact on life expectancy and cost-effectiveness. *JAMA.* 1999;282(22):2156-2163.

95. Tabar L, Faberberg G, Day NE, Holmberg L. What is the optimum interval between mammographic screening examinations? An analysis based on the latest results of the Swedish two- county breast cancer screening trial. *Br J Cancer.* 1987;55:547-551.

96. Duffy SW, Day NE, Tabar L, Chen HH, Smith TC. Markov models of breast tumor progression: some age-specific results. *J Natl Cancer Inst Monogr.* 1997;22:93-97.

97. Duffy SW, Chen HH, Tabar L, Fagerberg G, Paci E. Sojourn time, sensitivity and positive predictive value of mammography screening for breast cancer in women aged 40-49. *Int J Epidemiol.* 1996;25:1139-1145.

98. Kramer BS, Brawley OW. Cancer screening. *Hematol Oncol Clin North Am.* 2000;14(4):831-848.

99. Skrabanek P. False premises and false promises of breast cancer screening. *Lancet.* 1985(August 10):316-320.

100. Ringash J. Canadian Task Force on Preventive Health Care. Preventive health care, 2001 update: screening mammography among women aged 40-49 years at average risk of breast cancer. *CMAJ.* 2001;164(4):469-476.

101. Tabar L, Vitak B, Chen HHT, et al. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer.* 2001;91(9):1724-1731.

102. Gotzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet.* 2000;355(9198):129-134.

103. Rajkumar SV, Hartmann LC. Screening mammography in women aged 40-49 years. *Medicine.* 1999;78(6):410-416.

104. Moss S. A trial to study the effect on breast cancer mortality of annual mammographic screening in women starting at age 40. Trial Steering Group. *J Med Screen.* 1999;6:144-148.

105. Ng EH, Ng FC, Tan PH, et al. Results of intermediate measures from a population-based, randomized trial of mammographic screening prevalence and detection of breast carcinoma among Asian women: the Singapore Breast Screening Project [published erratum appears in Cancer. 1998;83(1):191]. *Cancer.* 1998;82:1521-1528.

106. Hakama MPEHM, Kallio M. Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. *BMJ.* 1997;314:864-867.

107. Hakama M, Pukkala E, Söderman B, et al. Implementation of screening as a public health policy: issues in design and evaluation. *J Med Screen.* 1999;6:209-216.

108. Berglund G, Nilsson P, Eriksson KF, et al. Long-term outcome of the Malmo preventive project: mortality and cardiovascular morbidity. *J Intern Med.* 2000;247(1):19-29.

109. Verbeek AL, Hendriks JH, Holland R, et al. Reduction of breast cancer mortality through mass screening with modern mammography: first results of the Nijmegen project, 1975-1981. *Lancet.* 1984;1:1222-1224.

110. Chamberlain J, Coleman D, Ellman R, et al. Verification of the cause of death in the trial of early detection of breast cancer: UK Trial of Early Detection of Breast Cancer Group. Trial Co-ordinating Centre. *Br J Cancer.* 1991;64(6):1151-1156.

111. Collette HJ, de Waard F, Rombach JJ, et al. Further evidence of benefits of a (non-randomised) breast

cancer screening programme: the DOM project. *J Epidemiol Community Health.* 1992;46(4):382-386.

112. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? [see comments.]. *Lancet.* 1998;352(9128):609-613.

113. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA.* 1995;273(5):408-412.

114. Prorok PC, Hankey BF, Bundy BN. Concepts and problems in the evaluation of screening programs. *J Chron Dis.* 1981;34:159-171.

115. Schulz KF, Grimes DA, Altman DG, et al. Blinding and exclusions after allocation in randomized controlled trials: survey of published parallel group trials in obstetrics and gynecology. *BMJ.* 1996;312:742-744.

116. Johansen HK, Gotzsche PC. Problems in the design and reporting of trials of antifungal agents encountered during meta-analysis. *JAMA.* 1999;282(18):1752-1759.

117. Bailar JC. Mammography: a contrary view. *Ann Intern Med.* 1976;84:77-84.

118. Skrabanek P. Mass mammography: the time for reappraisal. *Int J Technol Assess Health Care.* 1989;5(3):423-430.

119. Schmidt JG. The epidemiology of mass breast cancer screening—a plea for a valid measure of benefit. *J Clin Epidemiol.* 1990;43:215-225.

120. Tarone R. 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.

121. Sutton AJ, Abams KR, Jones DR, et al. *Methods for Meta-analysis in Medical Research* (Wiley Series in Probability and Statistics). Chichester: John Wiley and Sons, Ltd.; 2000.

122. Habbema JD, van Oortmarssen GJ, van Putten DJ, Lubbe JT, van der Maas PJ. Age-specific reduction in breast cancer mortality by screening an analysis of the results of the Health Insurance Plan of Greater New York study. *J Natl Cancer Inst.* 1986;77:317-320.

123. Shapiro S, Venet W, Straz P, Venet L, Roeser R. Selection, follow-up, and analysis in the Health Insurance Plan Study: a randomized trial with breast cancer screening. *Natl Cancer Inst Monogr.* 1985;67:65-74.

124. Frisell J, Lidbrink E. The Stockholm Mammographic Screening Trial: risks and benefits in age group 40-49 years. *J Natl Cancer Inst Monogr.* 1997;22:49-51.

125. Frisell J, Eklund G, Hellstrom L, Lidbrink E, Rutqvist LE, Somell A. Randomized study of mammography screening: preliminary report on mortality in the Stockholm trial. *Breast Cancer Res Treat.* 1991;18:49-56.

126. Alexander FE. The Edinburgh Randomized Trial of Breast Cancer Screening. *J Natl Cancer Inst Monogr.* 1997;22:31-35.

127. Alexander FE, Anderson TJ, Brown HK. The Edinburgh randomised trial of breast cancer screening: results after 10 years of follow-up. *Br J Cancer.* 1994;70:542-548.

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

AHRQ Pub. No. 03-507B
August 2002