Efficacy of Breast Cancer Screening in the Community According to Risk Level

Joann G. Elmore, Lisa M. Reisch, Mary B. Barton, William E. Barlow, Sharon Rolnick, Emily L. Harris, Lisa J. Herrinton, Ann M. Geiger, R. Kevin Beverly, Gene Hart, Onchee Yu, Sarah M. Greene, Noel S. Weiss, Suzanne W. Fletcher

Background: The efficacy of breast cancer screening in the community may differ from that suggested by the results of randomized trials, and no data have been available on efficacy among women who have different levels of breast cancer risk. Methods: We conducted a matched case-control study among women enrolled in six health plans in Washington, Oregon, California, Massachusetts, and Minnesota. We examined the efficacy of screening by mammography and/or clinical breast examination among women in two age cohorts (40–49 years and 50–65 years) and in two breast cancer risk levels (average and increased risk). Women who died from breast cancer from January 1, 1983, through December 31, 1998, (N = 1351; case subjects) were matched to control subjects (N = 2501) on age and risk level. Increased risk was defined as a family history of breast cancer or a breast biopsy noted in the medical records before the index date (defined as date of first suspicion of breast abnormalities in case subjects, with the same date used for matched control subjects). Data on screening, risk status, and other variables were abstracted from medical records. Conditional logistic regression was used to examine the association between breast cancer mortality and receipt of screening. All statistical tests were two-sided. Results: There were small, non-statistically significant associations between breast cancer mortality and receipt of screening during the 3 years prior to the index date for both the younger women [odds ratio (OR) = 0.92; 95% confidence interval (CI) = 0.76 to 1.13] and the older women (OR = 0.87; 95% CI = 0.68 to 1.12). The association among women at increased risk (OR = 0.74; 95% CI = 0.50 to 1.03) was stronger than that among women at average risk (OR = 0.96; 95% CI = 0.80 to 1.14), but the difference was not statistically significant (P = .17). Conclusions: In this community-based study, screening history was not associated with breast cancer mortality. However, potential limitations of this study argue for a cautious interpretation of these findings. [J Natl Cancer Inst 2005;97:1035–43]

Although mammography and clinical breast examination are widely used to screen for breast cancer, which women benefit the most from screening and the magnitude of the benefit remain controversial (1). No studies, to our knowledge, have reported the efficacy of screening according to breast cancer risk factors other than age. Historically, women at increased risk for breast cancer, especially those in their 40s (1,2), have been encouraged to participate in screening.

A randomized controlled trial of mammography screening for women in their early 40s is underway in the United Kingdom (3) that may reveal whether screening is effective in this age group. However, it is unlikely that additional randomized trials of breast cancer screening will be initiated because screening with mammography and clinical breast examination is already the standard community practice in many countries, and because it would take 10–15 years for a new trial to provide answers on mortality reduction. Newer imaging modalities, which can be helpful diagnostic tools (e.g., ultrasound and magnetic resonance imaging), have not been approved for general population screening purposes and are unlikely to be approved in the near future (4). Thus, the mainstays of screening for some time to come will continue to be screening mammography and breast examination performed by a clinician.

Given the fact that screening by mammography and clinical breast examination is widely used in the population and that another randomized trial to determine the efficacy of mammography is unlikely to be conducted, we performed a multicenter, case-control study to assess the efficacy of community-based breast cancer screening among women in two different age groups (40–49 years and 50–65 years) and at two different levels of risk for breast cancer (average and increased risk). Our study was designed to compare the history of screening among women who died as a result of breast cancer (case subjects) with that among women who were at risk of developing breast cancer (control subjects). The observation of a smaller proportion of screened case subjects than screened

Affiliations of authors: School of Medicine, Harbormedical Center, University of Washington, Seattle, WA (JGE, LMR); Center for Health Studies, Group Health Cooperative, Seattle, WA (JGE, WEB, RKb, GH, OY, SMG); Department of Epidemiology, School of Public Health and Community Medicine, University of Washington, Seattle, WA (JGE, NWS); Center for Health Research, Kaiser Permanente, Portland, OR (ELH); HealthPartners Research Foundation, Minneapolis, MN (SR); Department of Ambulatory Care and Prevention, Harvard Pilgrim Health Care, Boston, MA (MMB, SWF); Research and Evaluation Department, Kaiser Permanente Southern California, Pasadena, CA (AMG); Division of Research, Kaiser Permanente Northern California, Oakland, CA (LJH); Biostatistics Department, School of Public Health and Community Medicine, University of Washington, Seattle, WA (WEB); Cancer Research and Biostatistics, University of Washington, Seattle, WA (WEB).

Correspondence to: Joann G. Elmore, MD, MPH, Professor, Division of General Internal Medicine, University of Washington School of Medicine, Harbormedical Center, 325 Ninth Ave., Box 359780, Seattle, Washington 98104–2499 (e-mail: jelmore@u.washington.edu).

See “Notes” following “References.”

DOE: 10.1093/jnci/dji183
© The Author 2005. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oupjournals.org.
control subjects would support the hypothesis that screening is associated with a reduction in breast cancer mortality (5,6,7).

SUBJECTS AND METHODS

Study Setting

This study was conducted as part of the Cancer Research Network, a population-based research program that comprises the enrollees, research centers, and data systems of 10 U.S. health plans. Six of those health plans participated in this study: Group Health Cooperative (Seattle, WA), Harvard Pilgrim Health Care (Boston, MA), HealthPartners Research Foundation (Minneapolis, MN), and Kaiser Permanente in three regions: Northwest (Portland, OR), Northern California, and Southern California. Institutional approval was obtained at each site.

Definition of Index Date and Breast Cancer Risk Level

For case subjects, the index date was the date of the first mention in the medical record of the appearance of breast symptoms or a suspicion of cancer in the same breast and general location where the breast cancer was subsequently detected (i.e., the date of a breast self-examination during which a woman noticed a lump or the date of a screening clinical breast examination or mammogram during which an abnormality was first detected in an asymptomatic woman). Control subjects were assigned the same index date as their matched case subjects. The “index period” refers to the 3-year period before and including the index date.

Women who had any family history of breast cancer or whose medical records noted that they had had a breast biopsy at any time before the index date were classified as having an increased risk of breast cancer. Women who did not have a family history of breast cancer or a personal history of a breast biopsy were classified as having an average risk of breast cancer.

Definition of Case Subjects

Only women who died of breast cancer were selected as case subjects because we expected that breast cancer screening would lead to a reduction in subsequent breast cancer deaths. All case subjects had an initial breast cancer diagnosis from January 1, 1983, through December 31, 1993; were between 40 and 65 years of age at breast cancer diagnosis; had died from breast cancer or causes possibly related to breast cancer from January 1, 1983, through December 31, 1993; and were enrolled continuously in a health plan during the index period and were active health plan members at the time of the matched case subject’s breast cancer diagnosis. Eligible control subjects were matched to case subjects on health plan, age, and level of risk for breast cancer. We matched a priori on the latter factor because women with an increased risk of breast cancer might be more likely both to undergo screening and to become a case subject. We attempted to identify two matched control subjects for each case subject.

Determination of Eligibility

A total of 1987 potential case subjects and 31794 potential control subjects were identified by using automated systems linked to the health care plans. We included all potential case subjects who were 50–65 years of age from all but the two largest health plans (Kaiser Permanente of Southern California and Kaiser Permanente of Northern California). For those two health plans, we included a random sample of potential case subjects (25% and 33%, respectively) because we could obtain adequate statistical power without including all eligible women from these plans. Case subjects were excluded if they met any of the following criteria after medical record review: their breast cancer was not diagnosed between January 1, 1983, and December 31, 1993 (n = 69); their age at diagnosis was not 40 through 65 years (n = 5); death occurred after December 31, 1998 (n = 31); death was not due to breast cancer (n = 336); not continuously enrolled in the health plan during the index period (n = 99); medical chart information was not available (n = 43); medical chart was not reviewed because the funded study period had ended or for other reasons (n = 34); or no eligible control subjects were found (n = 19).

Control subjects were excluded if they met any of the following criteria: diagnosed with breast cancer before index date (n = 39); age not within the age limits for matching the case subject on the index date (n = 29); died on or before the date the matched case subject had died (n = 17); not continuously enrolled in the health plan during the index period (n = 214); risk level not the same as that of the case subject after medical record review (n = 2897); complete medical chart was not available (n = 20); or abstraction was not necessary because two control subjects had already been obtained for the matched case subject (n = 26077). Potential control subjects were sampled with replacement for different matched sets (i.e., a control subject could be matched to multiple case subjects or become a case subject at a later time point, although neither situation actually occurred in our final population) (9). For each case subject, we reviewed the medical records of a maximum of 10 potential control subjects. However, there were 201 case subjects for whom only one matched control subject was found after reviewing all 10 potential control subjects’ records. The final study population consisted of 1351 case subjects and 2501 control subjects.

Medical Record Review

Trained abstractors reviewed all available medical records for each subject, including outpatient records, mammography request
forms, mammography results, and hospitalization records. Site coordinators and abstractors were extensively trained in the standardized abstraction protocol and abstraction quality was monitored and maintained by monthly double review of one eligible and one ineligible chart abstraction performed in a blinded fashion and by bi-monthly conference calls, as previously described (10).

Information on breast cancer risk factors, demographics, comorbidity [classified as described in (11)], and other variables was abstracted for the 10-year period before, but not including, the index date. To obtain data on breast biopsy history (yes, no/unknown) and family history of breast cancer (yes, no/unknown) that were used to define the subject’s breast cancer risk level, abstractors reviewed information in all available medical charts up to the day before the index date. Risk factor data noted on the index date were not included because of the possibility that these factors were being noted more thoroughly in the medical charts of case subjects than in the medical charts of control subjects once a breast cancer was suspected for the case subject. Data on breast cancer characteristics and treatments were collected from NCI Surveillance, Epidemiology, and End Results (SEER) cancer registries or from approved tumor registries at four sites (Group Health Cooperative and the three Kaiser Permanente sites), and through medical chart reviews at the remaining two sites (Harvard Pilgrim Health Care and HealthPartners Research Foundation). Data were entered directly into computers; programmed data limits and subroutines performed initial checks of data for quality assessment simultaneously throughout the chart review process. Re-abstraction of 4% of the medical records (n =160) by abstractors who were blinded to case–control status produced high agreement between the initial unmasked abstraction and masked abstraction with respect to the classification of mammography examinations as screening or diagnostic (κ range = 0.76–0.91) (10).

Definitions of Breast Cancer Screening

Medical records were also reviewed to determine the subject’s breast cancer screening history (by mammography and/or clinical breast examination) during the index period. For the majority of women, a 3-year index period would be the expected time frame prior to diagnosis during which a tumor possibly could be identified by the screening test (i.e., the lead-time) (12–15).

Screening mammograms were defined as those that were performed when the woman was asymptomatic (i.e., she had no reported symptoms of breast cancer) and that were not being done as follow-up for an abnormality previously noted on an examination or mammogram within the previous year. Diagnostic mammograms were defined as mammograms obtained because of symptoms mentioned by the patient or to further evaluate an abnormal or suspicious finding noted during a previous clinical breast examination or mammogram. Clinical breast examinations were similarly defined as screening or diagnostic.

A committee of clinicians (JGE, SWF, MBB) who were blinded to case–control status reviewed the medical records in cases when the clinical scenarios suggested that coding of examinations as screening or diagnostic may have been incorrect (272 data queries on 191 women). Overall, coding remained unchanged in 91% of the examinations originally labeled as screening mammograms, 93% of those originally labeled as screening clinical breast examinations, 76% of those originally labeled as diagnostic mammograms, and 92% of those originally labeled as diagnostic clinical breast examinations.

Statistical Methods

Our primary analysis focused on estimating the association between death due to breast cancer and the occurrence of breast cancer screening during the index period. Only mammography and clinical breast examinations conducted for screening purposes were considered as the exposure of interest; diagnostic examinations were not included. We calculated the odds ratio (OR) with its 95% confidence interval (CI) by using conditional logistic regression retaining matching by health plan, calendar time, age, and breast cancer risk level (average or increased risk) (16,17). We also included in the model other possible confounders and/or predictors of breast cancer mortality (i.e., race, comorbidity, and age at first birth). Inclusion of other potential confounders in the model had no additional effect on the association between breast cancer mortality and screening. For each age group, we computed sample size estimates that assumed that the analysis was combined across the two risk groups. Although the absolute risk of breast cancer death could differ in the two risk groups, our sample size estimates assume that the odds ratio between screening and death due to breast cancer was the same for both strata. Under a complete one case subject:two control subjects matching protocol, we would have 80% power (at a two-sided alpha level of .05) to detect an odds ratio of 0.75 or lower for screened versus unscreened women.

We also performed analyses that excluded the 85 case subjects whose cause of death was possibly (but not definitely) breast cancer and their 152 matched control subjects. In addition, we performed analyses by using a 2-year, rather than a 3-year, screening interval. To describe screening behavior among the control subjects, we calculated screening in each calendar year separately for screening mammography and clinical breast examination by dividing the total number of screening events by the number of person-years contributed by the control subjects during that year. All statistical tests were two-sided.

RESULTS

Study Population

A total of 1351 case subjects and 2501 matched control subjects were identified. The numbers of case subjects and control subjects varied across the six health plans, and the number of case subjects per site ranged from 51 to 415. Case and control subjects had similar demographic and clinical characteristics (Table 1). Among the women who were at increased risk for breast cancer, 88% had only one risk factor (i.e., a family history of breast cancer or a previous breast biopsy) and the remaining 12% had both risk factors; these proportions were the same for case and control subjects (data not shown).

Among the case subjects, 18 (1.3%) were initially diagnosed with ductal carcinoma in situ, 380 (28.1%) were diagnosed with local tumors, 749 (55.4%) had positive regional nodes, 179 (13.3%) had positive distant nodes or metastases, and 26 (1.9%) had unknown node status. At diagnosis, 25.7% of case subjects had a primary tumor smaller than 20 mm, 30.3% had a primary...
Table 1. Demographic and clinical characteristics of women who died of breast cancer (case subjects) and matched control subjects, by age*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>40 – 49 years Case subject (N = 786)</th>
<th>Control subject (N = 1483)</th>
<th>50 – 65 years Case subject (N = 565)</th>
<th>Control subject (N = 1018)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>518 (65.9)</td>
<td>908 (61.2)</td>
<td>410 (72.6)</td>
<td>772 (75.8)</td>
</tr>
<tr>
<td>African American</td>
<td>116 (14.7)</td>
<td>157 (10.6)</td>
<td>64 (11.3)</td>
<td>55 (5.4)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>35 (4.5)</td>
<td>85 (5.7)</td>
<td>13 (2.3)</td>
<td>26 (2.6)</td>
</tr>
<tr>
<td>Asian</td>
<td>47 (6.0)</td>
<td>117 (7.9)</td>
<td>27 (4.8)</td>
<td>38 (3.7)</td>
</tr>
<tr>
<td>Native American</td>
<td>0 (0)</td>
<td>2 (0.1)</td>
<td>0 (0)</td>
<td>3 (0.3)</td>
</tr>
<tr>
<td>Unknown</td>
<td>70 (8.9)</td>
<td>214 (14.4)</td>
<td>51 (9.0)</td>
<td>124 (12.2)</td>
</tr>
<tr>
<td>Age at first birth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nulliparous</td>
<td>101 (12.9)</td>
<td>173 (11.7)</td>
<td>55 (9.7)</td>
<td>79 (7.8)</td>
</tr>
<tr>
<td>50 – 65 years</td>
<td>357 (45.4)</td>
<td>580 (39.1)</td>
<td>263 (46.6)</td>
<td>499 (49.0)</td>
</tr>
<tr>
<td>≥30 years</td>
<td>84 (10.7)</td>
<td>128 (8.6)</td>
<td>42 (7.4)</td>
<td>50 (4.9)</td>
</tr>
<tr>
<td>Unknown</td>
<td>244 (31.0)</td>
<td>602 (40.6)</td>
<td>205 (36.3)</td>
<td>390 (38.3)</td>
</tr>
<tr>
<td>Medical comorbidity index†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>572 (72.8)</td>
<td>1113 (75.0)</td>
<td>260 (46.0)</td>
<td>560 (55.0)</td>
</tr>
<tr>
<td>1</td>
<td>135 (17.2)</td>
<td>255 (17.2)</td>
<td>190 (33.6)</td>
<td>311 (30.6)</td>
</tr>
<tr>
<td>≥2</td>
<td>79 (10.0)</td>
<td>115 (7.8)</td>
<td>115 (20.4)</td>
<td>147 (14.4)</td>
</tr>
<tr>
<td>Family history of breast cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative/Unknown</td>
<td>644 (81.9)</td>
<td>1230 (82.9)</td>
<td>461 (81.6)</td>
<td>793 (77.9)</td>
</tr>
<tr>
<td>First-degree relative</td>
<td>75 (9.5)</td>
<td>104 (7.0)</td>
<td>63 (11.1)</td>
<td>132 (13.0)</td>
</tr>
<tr>
<td>Non-first-degree relative</td>
<td>67 (8.5)</td>
<td>149 (10.1)</td>
<td>41 (7.3)</td>
<td>93 (9.1)</td>
</tr>
<tr>
<td>History of breast biopsy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No previous biopsy/Unknown</td>
<td>686 (87.3)</td>
<td>1317 (88.8)</td>
<td>444 (78.6)</td>
<td>855 (84.0)</td>
</tr>
<tr>
<td>≥1 biopsy</td>
<td>100 (12.7)</td>
<td>166 (11.2)</td>
<td>121 (21.4)</td>
<td>163 (16.0)</td>
</tr>
<tr>
<td>Breast cancer risk level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average risk</td>
<td>570 (72.5)</td>
<td>1107 (74.6)</td>
<td>370 (65.5)</td>
<td>671 (65.9)</td>
</tr>
<tr>
<td>Increased risk</td>
<td>216 (27.5)</td>
<td>376 (25.4)</td>
<td>195 (34.5)</td>
<td>347 (34.1)</td>
</tr>
</tbody>
</table>

*Percentages may not total 100% due to rounding.
†Categories defined according to the Charlson et al. (11) comorbidity index modified to include one point for hypertension. Scoring of comorbidity can be summarized as 0 = no disease; 1 = one disease; and ≥2 = at least two medical problems or one severe medical problem.

Efficacy of Screening

To assess efficacy, we considered three definitions of breast cancer screening in our models (Table 3). The first definition was screening by either or both modalities (clinical breast examination, screening mammography, or both). The second definition was screening by clinical breast examination only, ignoring mammography. The third definition was by screening mammography only, ignoring clinical breast examination.

Overall, compared with no screening, breast cancer screening was not associated with a reduction in breast cancer mortality in this population (OR = 0.91, 95% CI = 0.78 to 1.07; Table 3). However, most of the odds ratios associated with screening, either by clinical breast examination alone or by mammography and clinical breast examination together, were lower for women at increased risk of breast cancer than for women at average risk of breast cancer, even though almost all 95% confidence intervals included 1.0 (Table 3). For example, among women of both age groups combined, the odds ratios of breast cancer mortality associated with screening by clinical breast examination or mammography were 0.74 (95% CI = 0.53 to 1.03) for women in the increased risk group and 0.96 (95% CI = 0.80 to 1.14) for women in the average risk group. The difference between these two odds ratios was not statistically significant (P = .17). In separate analyses of screening by clinical breast examination alone or screening mammogram alone, neither screening modality was associated with any appreciable reduction in breast cancer mortality (Table 3). Similar differences in the efficacy of screening between increased and average risk women were observed in both age categories (Figure 2).
These results did not change when we excluded from the analysis the 85 case subjects whose cause of death was only possibly (but not definitely) due to breast cancer and their 152 matched control subjects. The results also did not change when we examined screening histories during a 2-year interval before the index date, rather than during a 3-year interval. Analyses that excluded case subjects who were diagnosed after 1988 and their matched control subjects (i.e., analyses that included a cohort in

Table 2. Breast cancer screening by mammography or clinical breast examination among case and control subjects, by risk level and age stratum

<table>
<thead>
<tr>
<th>Screening history</th>
<th>Women aged 40–49 years</th>
<th>Women aged 50–65 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average risk</td>
<td>Increased risk</td>
</tr>
<tr>
<td></td>
<td>Case subjects (N = 570)</td>
<td>Control subjects (N = 1107)</td>
</tr>
<tr>
<td></td>
<td>n  (%)</td>
<td>n  (%)</td>
</tr>
<tr>
<td>Clinical breast exam and mammogram</td>
<td>141  (24.7)</td>
<td>307  (27.7)</td>
</tr>
<tr>
<td>Clinical breast exam only</td>
<td>216  (37.9)</td>
<td>374  (33.8)</td>
</tr>
<tr>
<td>Mammogram only</td>
<td>21   (3.7)</td>
<td>31   (2.8)</td>
</tr>
<tr>
<td>No screening</td>
<td>192  (33.7)</td>
<td>395  (35.7)</td>
</tr>
</tbody>
</table>

Women aged 50–65 years

<table>
<thead>
<tr>
<th>Screening history</th>
<th>Women aged 40–49 years</th>
<th>Women aged 50–65 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average risk</td>
<td>Increased risk</td>
</tr>
<tr>
<td></td>
<td>Case subjects (N = 370)</td>
<td>Control subjects (N = 671)</td>
</tr>
<tr>
<td></td>
<td>n  (%)</td>
<td>n  (%)</td>
</tr>
<tr>
<td>Clinical breast exam and mammogram</td>
<td>132  (35.7)</td>
<td>226  (33.7)</td>
</tr>
<tr>
<td>Clinical breast exam only</td>
<td>116  (31.4)</td>
<td>218  (32.5)</td>
</tr>
<tr>
<td>Mammogram only</td>
<td>10   (2.7)</td>
<td>20   (3.0)</td>
</tr>
<tr>
<td>No screening</td>
<td>112  (30.3)</td>
<td>207  (30.8)</td>
</tr>
</tbody>
</table>

Fig. 1. Rates of screening clinical breast examination (CBE) and mammography examinations among control subjects, by age group and risk level for breast cancer. Triangle, women at average risk; circle, women at increased risk.
which all of the case subjects had a minimum follow-up time of 10 years and a maximum follow-up time of 15 years) did not produce a substantially different result from analyses including the entire cohort (n = 775 case subjects and n = 1444 control subjects; OR = 0.85; 95% CI = 0.70 to 1.04). We noted little or no difference in breast cancer screening histories between case subjects and control subjects, irrespective of the time interval during which the case subject had died (1983–90, 1991–94, or 1995–98). Because a large number of prevalent cases of breast cancer might be found among women who are age 40 years or age 50 years, the ages at which screening for breast cancer typically begins in clinical practice, and because the 3-year interval during which we examined screening history could have included an earlier decade of age, we also performed analyses in which we considered only those case subjects who were either 43–49 years old or 53–65 years old at diagnosis (and their matched control subjects). However, we observed no appreciable difference between the odds ratios obtained in these various analyses and those presented earlier.

Disentangling the possible effects of screening mammography from clinical breast examination proved difficult for several reasons. When mammography and clinical breast examination were performed on the same day, it was often unclear which examination was done first and whether the second examination was interpreted without knowledge of the results of the first examination. It was also possible that some screening mammograms were prompted by an abnormal clinical breast examination. Therefore, we performed an additional analysis in a smaller subset of women, from which those who had an indeterminate or suspicious screening clinical breast examination during the study period were excluded. However, in this analysis we found no consistent difference in the association between having screening mammography and breast cancer mortality between case and control subjects by age or by breast cancer risk level (data not shown).

**DISCUSSION**

We found that breast cancer screening, as practiced in U.S. community settings in the 1980s and early 1990s, was associated with, at most, a modest reduction in breast cancer mortality. This was true both for women in their 40s and for women 50 to 65 years of age. Our results also suggest that screening may be more efficacious among women who have an increased risk for breast cancer; the odds ratio of breast cancer mortality for women 40–65 years old at increased risk was 0.74 (95% CI = 0.53 to 1.03), compared with an odds ratio of 0.96 (95% CI = 0.80 to 1.14) for women at average risk. However, the difference between these odds ratios was not statistically significant.

![Fig. 2. Adjusted odds ratios and 95% confidence intervals for screening by clinical breast examination or mammography, by age group and risk level for breast cancer. Models were adjusted for race (white, African American, other, unknown), medical comorbidity index (0, 1, ≥2), and age at first birth (<30 years, ≥30 years, unknown).](https://academic.oup.com/jnci/article-abstract/97/14/1035/2521304)
Overall, our findings suggest that breast cancer screening in the community was minimally effective in preventing death from breast cancer. At first glance, this conclusion seems questionable, especially for women in their 50s and 60s, given that results of randomized trials (1) have shown that breast cancer screening is associated with reduced breast cancer mortality among women in this age group. However, controversy exists about the methods of some of those trials (18–23) and concerns have been raised of possible harms of screening (24–28). For all studies, and especially those with unexpected results such as those reported here, it is critical to examine potential explanations for the findings.

We considered four possible explanations for the lack of a reduction in breast cancer mortality. One is chance. That is, our results are statistically compatible with a true 24% reduction in mortality from breast cancer associated with screening among women in their 40s, and with a true 32% decrease among women in their 50s and 60s. It is plausible that a 24–32% decrease in mortality reflects the actual benefit of screening when the dominant modality is clinical breast examination, a relatively insensitive method of detection (26–28).

Second, our case–control study could have been biased through differential misclassification of status of exposure to screening. Specifically, it is possible that women who sought screening, especially mammography, may have had symptoms that they did not report or that were not recorded in their medical records. The retrospective design of our study also lends itself to potential nondifferential misclassification, which could result from errors in chart abstraction or missing data on key variables. However, in a masked sub-study in the same cohort (10), we found that extensive training of abstractors resulted in excellent agreement among abstractors on assessment of exposure to screening examinations. In addition, results of re-abstraction of ambiguous screening histories by clinicians who were masked to case–control status lead us to believe that nondifferential misclassification was likely kept to a minimum.

A third possible reason for the lack of evidence of an association between screening and mortality is that patients were followed up for a relatively short time from diagnosis to death. Women in our study had 1 to 15 years of follow-up before they died of breast cancer, with the average time from diagnosis to death 4.2 years. Our study period, therefore, may not have provided enough follow-up time after screening to see an association between screening and mortality. However, limiting the analysis to case subjects who had 10–15 years of possible follow-up did not change the results. We purposely did not expand the follow-up time period beyond 15 years because we would have had to include case subjects who were diagnosed in the 1970s. Since that time, there have been secular increases in screening rates in the United States (29) and improvements in treatment since that time (30,31).

Fourth, we could identify risk factors for breast cancer only by reviewing subjects’ clinical records. Case subjects might still have been at an inherently greater likelihood than their matched control subjects of developing breast cancer and of seeking screening. However, we believe that this was not so, because our data showed that the proportions of women with one versus two risk factors were similar for increased-risk case subjects and control subjects. Our data were collected before testing for BRCA1 and BRCA2 mutations was available; however, such testing is applicable to only a small percentage of the screened population and thus the lack of this information is unlikely to have influenced the results.

We examined our data for numerous additional potential sources of bias. To remove potential bias caused by mammograms ordered in response to an abnormal clinical breast examination, we analyzed data from the subgroup of women who had no abnormal clinical breast examinations. We did not find any stronger association between receipt of a mammogram and reduced mortality in this analysis. We also reduced the time interval during which screening history was ascertained from 3 years to 2 years because it has been suggested that a 3-year index period is too long to show an effect of screening (12). This modified-interval analysis also failed to find any stronger association between screening and mortality. Screening techniques and quality may have improved over time; however, there was no substantial difference in breast cancer screening histories between case subjects and control subjects, irrespective of whether the histories were during the 1980s or the 1990s. In addition, the results were not affected by restricting the analysis to women aged 53–65 years (i.e., by excluding women aged 50–52 years who had screens during their 40s that could fall within the 3-year screening study period). Finally, an analysis that excluded case subjects (and their matched control subjects) whose cause of death was not definitely identified as breast cancer gave results similar to the analyses that had included those women.

Our findings may, therefore, reflect a possible reduction in the accuracy of screening as it moves from highly controlled randomized trials to real-life clinical practice. For screening mammography, questions have been raised about its quality in the community (32), the importance of the radiologists’ experience (33), and the importance of the health care and auditing systems (34–39). Although fewer data are available on the accuracy of screening clinical breast examination, the sensitivity of this modality reported in community practice is lower than that from clinical trials (26–28). However, our data were gathered from highly regarded U.S. health plans, two of which have reported better survival among health plan members with breast cancer than national averages (40).

Improvements in breast cancer treatment during the past 2 decades (30,31,41,42,43) could also have affected the efficacy of screening. If newer treatments cured some women who, in the 1970s, would have died, these improved treatments could make it difficult to demonstrate a difference in breast cancer mortality between screened and unscreened women.

This study provides a timely examination of a clinically important topic and has several strengths. Specifically, it examined breast cancer screening in five U.S. states as it is actually practiced in the community, included a large sample of women 40–49 years of age, and included a more ethnically diverse group of women (approximately 30% were not white) than any randomized controlled trial of screening to date. Our case–control study is also the largest completed to date on this topic and is representative of the broad U.S. population, having been drawn from an underlying population of 8.68 million covered lives (approximately 3% of the U.S. population). This study was large enough to provide the first data on separate results of screening efficacy by breast cancer risk level.

The case–control study design has been suggested as an efficient alternative to a randomized trial for the evaluation of
screening tests (3,6). In particular, results of one study (7) support the use of case-control studies in evaluating the efficacy of screening mammography when randomized controlled trials are not feasible. In a randomized trial of screening, women would be randomly assigned to receive regular screening (the intervention group) or to not receive screening (the control group). Follow-up data would then be obtained after 10–15 years to compare the subsequent breast cancer mortality rates between the two groups. A randomized trial would include women with screen-detected cancers as well as interval cancers, which would include the fast-growing lethal tumors that screening might not detect early enough for effective treatment. In a randomized trial, the consequence of these interval cancers would be to decrease the protective effect of screening because some of the breast cancer deaths (i.e., those from interval cancers) would not have been preventable. Similarly, in a case-control study, interval cancers would not be detected by screening and thus would decrease the protective effect of screening. The design of our case-control study was as similar as possible to that of a randomized trial.

In conclusion, we observed no appreciable association between breast cancer mortality and screening history, regardless of whether screening took place during a woman’s 40s, 50s, or early 60s. Although our results suggest that screening might be efficacious only among women who are at increased risk for breast cancer, the differences in the estimated efficacy of screening according to women’s risk levels were well within the limits of chance. In addition, the non-randomized study design and potential limitations of the available data argue for a cautious interpretation of these results. Nevertheless, it is possible that the efficacy of screening for breast cancer may be lower in community settings than in randomized clinical trials or as the treatment of breast cancer improves.

REFERENCES


**NOTES**

This project was supported by a grant from the National Cancer Institute (CA79689, PI Dr. Edward H. Wagner). We thank the project coordinators and medical record abstractors at each site for their hard work, Sherry Falls, Cary Williams, and Joe Egger for coordination of the project and Roy Pardee for development of our automated database and data management.

Manuscript received October 21, 2004; revised May 9, 2005; accepted May 20, 2005.