Breast Cancer Screening Outcomes in Women Ages 40–49: Clinical Experience With Service Screening Using Modern Mammography

Edward A. Sickles*

The several randomized controlled trials (RCTs) of breast cancer screening among women of ages 40 to 49 now collectively show a statistically significant reduction in breast cancer mortality. However, there have been numerous recent advances in mammography, such that it now is demonstrably better than when the RCTs were conducted. The use of surrogate measures of screening efficacy (tumor size, lymph node status, cancer stage), readily derived from modern service screening programs, demonstrates how the improved mammography of the 1990s should produce a greater degree of mortality reduction among women ages 40–49 than that already demonstrated in the RCTs. Indeed, these surrogate measures of mortality reduction are as favorable for women of ages 40–49 and 65+ as they are for women of ages 50–64, strongly suggesting that, since modern service screening is accepted as effectively reducing mortality among women of ages 50–64, it should also effectively reduce mortality among women in the 40–49 and 65+ age groups. [Monogr Natl Cancer Inst 1997;22:99–104]

The best evidence of mortality reduction from breast cancer by screening women ages 40 to 49 comes from the several randomized controlled trials (RCTs) that already have been conducted. Like the screening carried out in the RCTs, service screening is performed on entire populations of women, either by invitation or by self-selection. However, service screening does not provide data from randomly selected control groups of nonscreened women. Therefore, service screening programs do not generate outcomes data that are sufficiently rigorous to independently furnish convincing evidence on mortality reduction.

Nonetheless, there still is considerable value in the outcomes data from modern service screening programs, for several reasons: (a) the data from existing RCTs indicate the presence of a substantial and (as of January 1997) statistically significant mortality reduction, but controversy remains over the magnitude of this mortality reduction; (b) known deficiencies in design and execution of the RCTs may have diminished the efficacy of screening with mammography and thereby reduced the extent of observed mortality reduction (1,2); (c) since the conduct of the RCTs, there have been numerous advances in mammographic equipment, technical imaging factors, quality assurance procedures, education of personnel, and mammographic interpretation performance (1,3–6), such that the mammography of the 1990s is demonstrably better than that done when the RCTs were conducted—advances that also may have caused the RCTs to under-estimate the extent of currently achievable mortality reduc-

tion; and (d) the use of surrogate measures of screening efficacy, readily derived from modern service screening programs, demonstrates how the improved mammography of the 1990s can be expected to produce a greater degree of mortality reduction than that already demonstrated in the RCTs, thereby increasing the likelihood that modern mammography truly benefits screened women. Outcomes data from modern service screening programs also provide indicators of the frequencies with which abnormal screening interpretations, additional imaging evaluations, and screen-induced biopsies are performed in the real world, removed from the artificial conditions inherent in the design and conduct of the RCTs.

Many modern mammography service screening programs have published data in the peer review literature. These include: (a) the population-based screening program in Uppsala county, Sweden (7); (b) the province-wide Screening Mammography Program of British Columbia (SMPBC), Canada (8); (c) the University of California San Francisco (UCSF) screening program, which serves women throughout the San Francisco Bay Area (9); and (d) the X-Ray Associates of New Mexico screening program (10). These programs, which provide screening with mammography alone, were selected because they currently involve very large numbers of screening examinations (Uppsala, 127,515 examinations; SMPBC, 598,165 examinations; UCSF, 84,615 examinations; New Mexico, 104,371 examinations) and because they each collect extensive outcomes data including but not limited to linkage with comprehensive tumor registries in their respective geographic areas.

The outcomes data, displayed in tabular format throughout this article, are drawn from the UCSF program (11,12), utilizing updated results for all screening examinations performed through December 31, 1996. Because I have complete access to this source material, I can generate age-related data breakdowns that are not readily retrievable from any other service screening program or RCT.

Benefits of Screening With Modern Mammography

Since there has been controversy over the magnitude of the mortality reduction demonstrated by the RCTs among women

*Affiliation of author: Department of Radiology, School of Medicine, University of California, San Francisco.

Correspondence to: Edward A. Sickles, M.D., Department of Radiology, UCSF/Mount Zion Medical Center, 2330 Post Street, #180, San Francisco, CA 94115.

© Oxford University Press
ages 40–49 and since very few women aged 65+ have been studied in RCTs, surrogate measures of screening outcomes have been proposed, validated (using the same RCT data that indicate the presence of a mortality reduction), and widely used as alternative means to indicate the efficacy of screening (7–12, 22). Two very powerful surrogate measures (i.e., those highly likely to predict reduced mortality) are tumor size and axillary lymph node status. Cancer stage, which is derived primarily from these two indicators, is the penultimate surrogate measure for screening efficacy; in fact, this measure is so valuable clinically that it is widely used in formulating treatment plans for breast cancer patients. Another useful measure is tumor grade. However, this cannot be evaluated readily in most service screening programs in the United States, because American pathologists too frequently omit grading data in their breast cancer reports (40% of cases in the UCSF program) and because many different pathologists perform grading assessments in the remaining cases, potentially introducing substantial subjective variation (18,23). A final surrogate measure of mortality reduction involves interval cancers—those cancers that are identified in the interval between screening examinations. Interval cancers grow more rapidly and have a poorer prognosis than screen-detected cancers (7,15,24–27); thus, a low interval cancer rate is a strong indicator of effective screening performance. However, the most valuable measure of interval cancers is the rate at which they occur in proportion to the rate at which cancers surface clinically in the absence of screening. Unfortunately, this measure is difficult to determine in the service screening setting, since there is no readily accessible control population of nonscreened women to provide the needed comparative data.

Surrogate measures of clinical efficacy are especially useful when employed in comparative studies—for example, in assessing the efficacy of different screening protocols (17). In this article, surrogate measures of mortality reduction will be used to compare the efficacy of screening women aged 40–49 and women aged 65+ with women of ages 50–64, the age range for which screening is widely accepted as being efficacious.

There is considerable evidence on the tumor size, lymph node status, and stage of cancers detected during modern service screening mammography. Indeed, these surrogate measures of mortality reduction appear to be as favorable for women ages 40–49 and 65+ as they are for women of ages 50–64 (see Table 1, which provides an update from the UCSF screening program, involving 72,145 examinations). Similar results also have been reported from the SMPBC, Uppsala, and New Mexico service screening programs (7–10,20). Thus, the surrogate-measure data strongly suggest that, since modern service screening is accepted as effectively reducing mortality among women in the 50–64 age group, it should also effectively reduce mortality among women of ages 40–49 and 65+.

There also is substantial evidence that the optimal screening interval for women aged 40–49 is one year, rather than the two-year interval used in most of the RCTs. Analysis of results from the Kopparberg portion of the Swedish two-county RCT, the Uppsala service screening program, and the Cincinnati Breast Cancer Detection Demonstration Project (another, albeit older service screening program) indicates that the lead time from screening women in their forties is substantially less than that from screening older women (7,20,21,27). Finally, as shown in Table 2, data from the UCSF service screening program demonstrate a substantial decline in sensitivity for screening women age 40–49 when the screening interval is increased from one year to two years, twice as large a decline as is observed for older women (28). This suggests that substantially more (poor-prognosis) interval cancers will occur if younger women are screened every two years rather than annually. Furthermore, the sensitivity for screening women age 40–49 at a one-year interval is equivalent to that of screening women age 50 and older at a two-year interval (28), an interval for the older cohort of women that already has been shown to produce statistically significant mortality reduction in the Swedish two-county RCT (22). Thus, these various lines of evidence all support the concept that if screening is recommended for women at ages 40–49, it should be done at an annual rather than a biennial interval.

It has been suggested that the slightly lower sensitivity for screening women aged 40–49 compared to older women may be a result of younger women’s breasts being more radiographically dense. This argument is supported by the observation that the proportion of women with dense breasts is slightly higher at age 40–49 than it is in older women (29,30). However, the data on screening sensitivity from the UCSF program show that breast density did not influence the sensitivity of mammography in women aged 40–49; sensitivity was 90% for women with primarily dense breasts, compared with 88% for women with primarily fatty breasts (28). Similar findings also have been observed in the Swedish two-county RCT (22). A much more likely explanation for the slightly lower screening sensitivity in women age 40–49 is that rapid tumor growth rates among younger women result in more (poor-prognosis) interval cancers between screening examinations, as implied in the preceding discussion of optimal screening interval. This theory is further supported by the UCSF data, which show that screening sensitivity decreases with increasing tumor size, especially for women age 40–49 (28), suggesting that cancers in younger women grow more rapidly. Again, screening women age 40–49 at an annual rather than a biennial interval should result in sensitivity equivalent to that of screening older women at a two-year interval.

<table>
<thead>
<tr>
<th>Table 1. Surrogate measures of breast cancer screening efficacy as a function of age at screening*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median size (invasive cancers)</td>
</tr>
<tr>
<td>Nodal metastasis (invasive cancers)</td>
</tr>
<tr>
<td>Stage 2 or higher cancers</td>
</tr>
</tbody>
</table>

*Based on data from the UCSF service screening program, involving 425 cancers and 72,145 screening examinations.

<table>
<thead>
<tr>
<th>Table 2. Sensitivity of initial screening mammography as a function of age at screening*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Follow-up interval</td>
</tr>
<tr>
<td>1 year</td>
</tr>
<tr>
<td>2 years</td>
</tr>
</tbody>
</table>

*Based on data from the UCSF service screening program, derived from (28).
Although invasive breast cancer seems to grow more rapidly in women age 40–49 than in older women, does it really disseminate more frequently when very small in size (≤10 mm—i.e., detected primarily by screening)? Data from the Nijmegen population-based mammography screening program (1975–1990) for women with invasive cancers 10 mm or less show that 40% of women younger than age 50 had positive axillary lymph nodes, whereas only 20% of women age 50 and older had positive nodes (31). Dutch investigators have attributed these findings to presumed age-related differences in tumor biology, suggesting that, in younger women, cancers disseminate early in their evolution, whereas cancers in older women produce nodal metastasis more slowly. However, parallel evidence from the more modern UCSF service screening program is strikingly different. Among women with small (≤10 mm) invasive cancers, none aged 40–49 had positive axillary lymph nodes, and 6% of women aged 50–64 had positive nodes (18). Almost identical results have been reported for the New Mexico service screening program (10) and for the Swedish two-county RCT (15). Therefore, it appears likely that the Nijmegen observations do not serve as a basic indicator of tumor biology but simply are limited by relatively ineffective mammographic techniques and approaches (18,32). In conclusion, advancing the time of diagnosis for invasive cancers by screening does indeed appear to diminish the propensity for axillary lymph node metastasis and hence reduces the likelihood for mortality equally in women of ages 40–49 as in older women.

There are other benefits of screening women aged 40–49, apart from those indicated by the surrogate-marker evidence cited above. These range from the reassurance gained from knowledge that a screening examination was normal to the greater likelihood of being eligible for breast conservation therapy and of avoiding breast radiation therapy and chemotherapy when cancer is detected by screening, versus usual care. However, the outcomes data from modern service screening programs do not provide evidence to document such benefits, and so discussion of these benefits is beyond the scope of this article.

**Risks of Screening With Modern Mammography**

Several other measures of performance are also reported for service screening mammography, even though they do not appear to be reliable surrogates for breast cancer mortality. These include positive predictive value (PPV), biopsy yield of cancer, and specificity. These measures do, however, provide useful indicators of the factors with which false-positive screening outcomes occur, thereby serving as surrogate measures for the risks (harms) of screening. It is important to note that PPV and biopsy yield are highly dependent on the prior probability of breast cancer, which increases steadily and substantially with advancing age, so that observed PPV and biopsy yield for women age 40–49 should be considerably lower than for older women.

Discussion of the nature and relative magnitudes of the risks of screening with mammography is also beyond the scope of this article. However, the outcomes data from modern service screening programs do provide relevant information, presented herein, on how some of these risks change with age. The risks of screening mammography should be considered separately for two specific populations of screened women.

The first population involves women recalled for additional noninterventional evaluation after abnormal mammography screening examinations. Outcomes data from modern service screening programs demonstrate that women age 40–49 are recalled for additional evaluation at approximately the same rate as women screened in later decades of life (7,33). When these data are examined by five-year age groupings, the same results are found (34). In the UCSF service screening program, there is essentially no difference in overall recall rate when comparing women age 40–49 with older women. Most recalled women will be found to have no clinically significant abnormalities. These women thus experience several negative outcomes of false-positive examination (anxiety, inconvenience, physical discomfort, cost). There also will be some women who eventually are found to have breast cancer, and because the incidence of breast cancer increases with advancing age, fewer women age 40–49 (than older women) will have true-positive examinations. Therefore, the PPV of screening will be lower for women age 40–49 than for older women (7,19,35). However, this age-dependent effect on true-positive examinations—that is, the prior probability of having breast cancer—is of very small overall magnitude, because less than 10% of recall examinations are true-positive examinations. Therefore, among women recalled for additional noninterventional evaluation after an abnormal screening examination, the overall risks are essentially age independent.

The second population involves women recalled for interventional evaluation (fine needle aspiration biopsy, core biopsy, or surgical biopsy) after abnormal diagnostic imaging examinations. Outcomes data from modern service screening programs demonstrate that women age 40–49 undergo these types of biopsy at approximately the same rate as women screened in later decades of life (7,11,12,33). When outcomes data are examined by five-year age groupings (and even by one year at a time), the same results are found (36). Most women undergoing biopsy will be found to have benign lesions. These women thus experience several negative outcomes of false-positive biopsy (anxiety, inconvenience, discomfort, scarring, cost), which are of greater magnitude than the risks described for recall examination. There also will be some women who eventually are found to have breast cancer, and because the incidence of breast cancer increases with advancing age, fewer women age 40–49 (than older women) will have true-positive biopsy. Therefore, the biopsy yield will be lower for women age 40–49 than for older women (7,11,12,33,36). However, this age-dependent effect on true-positive biopsy—again, the prior probability of having breast cancer—is of relatively small overall magnitude because only about one-third of biopsies are true-positive cases (11,12). Therefore, among women undergoing interventional evaluation after abnormal diagnostic imaging examinations, the overall risks are for the most part age independent.

Another point merits consideration concerning the biopsy of screen-detected lesions. In the United States, over the past five years, there has been a dramatic increase in the number of these lesions that undergo biopsy by percutaneous sampling (fine needle aspiration biopsy or core biopsy) rather than by surgical excision. Compared to surgical biopsy, percutaneous sampling is equally accurate but results in much less discomfort, produces
essentially no scarring, and is done at less than half the cost (37–40). When lesions undergoing percutaneous biopsy are found to be benign, in most cases surgical biopsy is averted, thereby resulting in substantially reduced morbidity. The cancer yield for surgical biopsy thus can be increased to between 60% and 75% (7,15,39,41). Because of the inherent advantages of percutaneous biopsy, the trend toward using this method rather than surgical biopsy for screen-detected lesions will probably continue at an accelerated rate as we proceed further into managed-care medicine.

One further useful piece of evidence can be derived from the UCSF service screening program. By comparing the outcomes from initial versus subsequent screening examinations, my colleagues and I demonstrated that the recall rate (frequency of abnormal screening interpretation) is substantially higher, the biopsy yield of cancer is considerably lower, and the surrogate measures of mortality reduction (tumor size, lymph node status, cancer stage) are less favorable for initial screening examinations than for subsequent examinations (42). Tables 3 and 4 present an update of UCSF screening program data on initial versus subsequent screening outcomes, demonstrating that the previously reported observations apply equally to women age 40–49 and to older women. It is important to note that ongoing service screening will involve many subsequent screening examinations but only one initial examination. Thus, outcomes data based either entirely or predominantly on initial screening will tend to underestimate the benefit and overestimate the risk of service screening.

Benefits and Risks of Screening With Modern Mammography, Applied to Populations of Women at Higher Than Average Risk for Breast Cancer

The RCTs were not designed to provide separate data on subpopulations of women at higher than average risk for developing breast cancer, and therefore no evidence on mortality reduction can be expected. However, outcomes data from the UCSF service screening program, using surrogate measures of screening performance, do provide the following indirect evidence on the benefits and risks of screening for women age 40–49 who have a strong or very strong family history of breast cancer: (a) the PPV of screening is higher in high-risk women age 40–49 than in the remainder of screened women in this age group (26%), again likely due to the increased incidence of breast cancer in high-risk women (greater prior probability of cancer); (b) the biopsy yield of cancer is higher in high-risk women age 40–49 (36%) than in the remainder of screened women in this age group (26%), again likely due to the increased incidence of breast cancer in high-risk women (greater prior probability of cancer); (c) the sensitivity of initial screening mammography is somewhat lower in high-risk women age 40–49 than in the remainder of initially screened women in this age group (28), likely due to a more rapid growth rate of cancers in younger high-risk women—these women do have a five-times greater risk of dying from breast cancer than younger average-risk women (43), suggesting that a greater proportion of cancers among younger high-risk women are aggressive and thus grow rapidly; (d) there are essentially no differences in the size, lymph node status, and stage of screen-detected breast cancers in comparing high-risk women age 40–49 with the remainder of screened women in this age group; and (e) had screening among women age 40–49 been limited to the 12% of these women at high risk by family history, this strategy would have detected only 19% of the extant cancers (18,35).

Thus, the overall conclusion to be drawn from the UCSF experience is that, for the age range 40–49, modern service screening mammography appears to detect breast cancer somewhat less effectively in women at high risk for developing breast cancer, but that the accompanying increased incidence of breast cancer will increase the positive predictive value and biopsy yield, thereby improving the cost-effectiveness of screening these high-risk women. However, the more cost-effective strategy of screening only high-risk women will relinquish to usual-care detection more than 80% of the cancers in the entire age 40–49 population.

Directions for Future Research

There have been numerous advances in conventional mammography over the past 10 years, involving equipment, technical imaging factors, quality assurance procedures, education of personnel, and mammographic interpretation performance. Continued advances are expected as we enter the 21st century. There is also promising and very important research involving digital mammography, high-resolution breast ultrasound, magnetic resonance imaging, and isotope scanning of the breast. Among these imaging techniques, digital mammography may provide increased sensitivity and/or specificity when used for breast cancer screening. All techniques may permit increased sensitivity

Table 4. Surrogate measures of breast cancer screening efficacy as a function of type of screening and age at screening*

<table>
<thead>
<tr>
<th>Type of Screening</th>
<th>Age</th>
<th>Subsequent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40–49</td>
<td>50+</td>
</tr>
<tr>
<td>Screening exams</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>Recalls for imaging</td>
<td>73</td>
<td>82</td>
</tr>
<tr>
<td>Biopsies performed</td>
<td>17</td>
<td>27</td>
</tr>
<tr>
<td>Cancers detected</td>
<td>4</td>
<td>12</td>
</tr>
</tbody>
</table>

*Based on data from the UCSF service screening program, involving 29,694 initial screening examinations and 42,451 subsequent screening examinations.

Table 3. Clinical outcomes of service screening mammography as a function of type of screening and age at screening*

<table>
<thead>
<tr>
<th>Type of Screening</th>
<th>Initial</th>
<th>Subsequent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age</td>
<td>Age</td>
</tr>
<tr>
<td></td>
<td>40–49</td>
<td>50+</td>
</tr>
<tr>
<td>Screening exams</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>Recalls for imaging</td>
<td>73</td>
<td>82</td>
</tr>
<tr>
<td>Biopsies performed</td>
<td>17</td>
<td>27</td>
</tr>
<tr>
<td>Cancers detected</td>
<td>4</td>
<td>12</td>
</tr>
</tbody>
</table>

*Based on data from the UCSF service screening program, involving 29,694 initial screening examinations and 42,451 subsequent screening examinations.
and/or specificity in the “diagnostic” setting (i.e., providing noninterventional evaluation of screen-detected abnormalities).

In contrast to breast imaging, which has undergone (and continues to undergo) many improvements, very little change has occurred in the practice of breast physical examination, other than the realization that it appears to be more accurate when performed with diligence by specially trained practitioners (44). Unfortunately, there currently is little enthusiasm either within or outside the medical community for improving the current state of breast physical examination in the United States. Two approaches that are likely to reap considerable benefit are (a) the recruitment, training, and deployment of large numbers of paramedical personnel to perform breast physical examination in screening centers and (b) federal legislation mandating quality assurance practices for breast physical examination, to parallel the provisions of the Mammography Quality Standards Act of 1992 (which has resulted in considerably improved delivery of high-quality mammography services).

The National Cancer Institute has funded a multisite Breast Cancer Surveillance Consortium, which is currently collecting outcomes data from more than one million women on many aspects of modern breast cancer screening. This research will provide valuable direction into methods of improving breast cancer screening in the United States. However, there is urgent need to go beyond this effort by creating a national cancer registry to permit collection of meaningful outcomes data for all American women. To be truly effective, such a cancer registry must permit low-cost data linkage by individual breast cancer screening practices, so that complete rather than partial outcomes data are available to service providers for the purpose of continuous quality improvement.

References


