

Stephen A. Feig, MD

**Index terms:**

Breast neoplasms, 00.32  
Breast neoplasms, radiography, 00.32  
Breast radiography, utilization, 00.32  
Cancer screening, 00.32

**Radiology 2000;** 214:633–640

**Abbreviations:**

MGH = Massachusetts General Hospital  
PPV = positive predictive value  
RCT = randomized controlled trial  
UCSF = University of California, San Francisco

<sup>1</sup> From the Department of Radiology, The Mount Sinai Hospital, Mount Sinai School of Medicine, New York, NY. Received February 15, 1999; revision requested March 25; revision received May 3; accepted June 14. **Address reprint requests** to the author, 200 Locust St, Apt 21A, Philadelphia, PA 19106.

© RSNA, 2000

# Age-related Accuracy of Screening Mammography: How Should It Be Measured?<sup>1</sup>

The American Cancer Society, the American College of Radiology, and the American Medical Association now advise annual screening mammography for all women beginning at the age of 40 years (1–3), and the National Cancer Institute now recommends screening mammography every 1–2 years starting at the age of 40 years (4). These guidelines are based on the statistically significant reductions in breast cancer deaths among women aged 35–49 years (5–10) and aged 50 years and older (10–12) who were offered screening mammography in randomized controlled trials (RCTs). Because such trials are not subject to the biases inherent in studies of survival rates, only RCTs can provide direct proof of benefit from screening (13).

Until 1993, the National Cancer Institute had recommended screening mammography every 1–2 years for women aged 40–49 years on the basis of indirect evidence that such screening was likely to reduce deaths caused by breast cancer. Annual screening mammography was advised for women aged 50 years and older.

Indirect evidence of the efficacy of screening women aged 40–49 years included (a) similarly improved survival rates for screening-detected cancers among women aged 40–49 years and among women aged 50 years and older in survival rate studies such as the Breast Cancer Detection Demonstration Project (14); (b) substantial but not statistically significant ( $P > .05$ ) reductions in breast cancer mortality among women aged 40–49 years at entry into RCTs (15); (c) detection of earlier cancers in service screening programs (ie, screening intended primarily for clinical service rather than research) conducted in the 1990s than in the RCTs that began 1–3 decades earlier (15); and (d) use of a new statistical method for analysis of data from an RCT conducted by the Health Insurance Plan Study of Greater New York. This reanalysis of Health Insurance Plan data suggested but did not unequivocally establish a statistically significant benefit from screening women in their 40s (10).

In November 1993, the National Cancer Institute rescinded its previous recommendation for screening women aged 40–49 years because none of the RCTs had yet found a statistically significant benefit from screening women in their 40s (16). The National Cancer Institute was also concerned that recently published data from the National Breast Screening Study of Canada, the data of which were of questionable validity, did not show a benefit from screening women in this age group (17). Concomitantly, the American College of Radiology, the American Medical Association, and the American Cancer Society continued to advise that women in their 40s should undergo screening every 1–2 years. All organizations agreed that the findings from the RCTs clearly had shown a benefit from screening women aged 50 years and older (11,12,16).

At that time, the lack of an unequivocal statistically significant benefit from screening women in their 40s led some observers to suggest that screening younger women might be ineffective (16). This conclusion may have been welcomed by some third-party payers eager to identify a means of stemming health care costs. In addition, some women felt relieved to be told that they need not be screened because their risk of cancer was lower than they believed (18) and because they were also told that the chance of an unnecessary diagnostic imaging work-up and biopsy of screening-detected findings was relatively high (19–23). The fact that a statistically significant benefit for younger women was not found until after several more years of follow-up only served to perpetuate the screening controversy.

The idea that the natural history of breast cancer may be different in premenopausal compared with postmenopausal women led to the separate analyses of screening data for women younger and older than the age of 50 years. However, none of the RCTs were

adequately designed to specifically evaluate the relative effectiveness of screening among younger women (13). The number of younger women enrolled was insufficient. In addition, the screening intervals in most RCTs were too long (18–28 months) with respect to the shorter detection lead time in younger women, that is, the mean length of time by which screening advances the time of detection from that expected if screening had not been performed. Therefore, no statistically significant reduction in mortality was found during the early years of follow-up (1).

With longer follow-up of women enrolled in the RCTs, a statistically significant ( $P < .05$ ) benefit for younger women began to emerge in 1995 (6) and continued with successively longer term RCT follow-up data published from 1995 to 1997. These reductions in breast cancer mortality for women who began screening between the ages of 35 and 49 years include 18% for eight RCTs combined (5); 24% for seven population-based RCTs combined (6,7); 29% for five Swedish RCTs combined (5); 35% for the Malmö, Sweden, RCT (8); and 46% for the Gothenburg, Sweden, RCT (9). As a consequence, the National Cancer Institute in 1997 decided to again recommend screening mammography every 1–2 years for all women in their 40s (4). At the same time, the American College of Radiology and the American Cancer Society increased their advised frequency for screening from every 1–2 years to every year for all women in this age group. The American Medical Association did so in 1999 (3). These latter decisions were based on compelling indirect evidence that annual screening should result in substantially greater benefit than biennial screening (1–3,7, 24–27).

### COMPARISON OF BENEFITS WITH ADVERSE CONSEQUENCES

Considering the longevity of the screening controversy, which began in 1975 (28), it was no surprise that some observers did not agree immediately with the new 1997 guidelines (29–31). However, the focus of their concern shifted from the existence of a benefit to the absolute amount of benefit and to a comparison of benefits with adverse consequences. Those observers who still doubt that mammography reduces deaths from breast cancer among younger women mainly question whether much of the benefit from screening women who entered the RCTs in their 40s may have resulted from can-

**TABLE 1**  
Cancer Detection Rates at Mammographic Screening at UCSF and MGH according to Age

| Screening                                | Age (y) |       |       |       |       |
|--|---------|-------|-------|-------|-------|
|  | 30–39   | 40–49 | 50–59 | 60–69 | 70–79 |
| UCSF (first screening)*                  | 1.3     | 2.7   | 6.0   | 13.1  | 14.2  |
| MGH (first and subsequent screenings)†‡  | NA      | 2.4   | 3.0   | 3.9   | 5.0   |
| UCSF (second screening)*                 | 1.4     | 1.3   | 2.9   | 1.3   | 3.0   |
| UCSF (first and subsequent screenings)†§ | 1.9     | 3.4   | 5.4   | 7.5   | 9.5   |

Note.—Cancer detection rate = cases/1,000 women, NA = not applicable.

\* Women without personal or close family history of breast cancer; calculated from data in Kerlikowske et al (19).

† Includes women with or without a close family history of breast cancer.

‡ Data from Kopans et al (39,40).

§ Calculated from data of Sickles (43).

cers detected after the age of 50 years (32). Such skepticism is not supported by published data (9,33).

Most of the reluctance of those who still do not favor screening younger women is caused by the misperceptions that detection rates are substantially lower and that the “unnecessary” imaging procedures, false-positive biopsy results, and cost-effectiveness are much higher for women aged 40–49 years than for women in the succeeding decades (18–23,29–31,34–37). Their basic argument is that because of the lower incidence of breast cancer among younger women, the ratios of benefit to adverse consequences and costs are abruptly lower before the age of 50 years.

These arguments were the focus of an article by Kerlikowske et al (19) published in *JAMA*, entitled “Positive Predictive Value of Screening Mammography by Age and Family History of Breast Cancer,” which evaluated the results of screening mammography performed at the University of California, San Francisco (UCSF). These authors summarized their results as follows (19):

Ten cancers were diagnosed per 1000 first-screening examinations in women aged 50 years or older, with 14.8 diagnostic procedures per cancer diagnosed compared with two cancers per 1000 screening examinations and 48.3 diagnostic tests per cancer diagnosed in women younger than 50 years. . . . Five times as many cancers per 1000 first-screening mammographic examinations were diagnosed in women aged 50 years or older compared with women aged less than 50 years.

These statements were used by the authors to support their argument that “Efforts to promote screening mammography should focus on women aged 50

years and older and on women aged 40–49 years with a family history of breast cancer” (19). This idea has been reiterated recently by Gail and Rimer (36), who propose that screening should be performed only on selected women aged 40–49 years whose risk factors place them at the same risk as that of a 50-year-old woman with no risk factors. This recommendation ignores the fact that most breast cancers occur among women without any specific risk factors such as family history of breast cancer (38).

Two articles by Kopans et al (39,40), which are based on screening data from Massachusetts General Hospital (MGH), were subsequently published as a response to Kerlikowske et al (19). The major point of Kopans et al was that the change in measures such as detection rates or positive predictive values (PPVs) from any year of age to the next, such as 49–50 or 50–51, is exceedingly slight and that screening parameters do not suddenly become unfavorable in women younger than 50 years of age (39,40–42). Thus, I would like to offer further commentary showing that screening parameters for women aged 40–49 years are nearly as favorable as those for women aged 50–59 years, especially when incidence (subsequent) screening data among women undergoing regular annual screening, rather than prevalence (initial) screening data, are compared.

### CANCER DETECTION RATES

It is not apparent from a glance at the abstract quoted from Kerlikowske et al (19) that “women younger than 50 years” and “women aged 50 years or older” refer to those aged 30–49 years and 50–69 years, respectively. Including women aged

**TABLE 2**  
**Relative Cancer Detection Rates for Women Aged 50–59 Years versus 40–49 Years at Initial and Subsequent Screenings at UCSF and MGH**

| Screening  | No. of Women | Relative Detection Rate* |
|--|--------------|--------------------------|
| UCSF (first screening) <sup>†</sup>                  | 14,448       | 2.2 (6.0/2.7)            |
| UCSF (second screening) <sup>†</sup>                 | 5,055        | 2.2 (2.9/1.3)            |
| UCSF (first and subsequent screenings) <sup>‡§</sup> | 41,738       | 1.6 (5.4/3.4)            |
| MGH (first and subsequent screenings) <sup>§  </sup> | 40,104       | 1.2 (3.0/2.4)            |

\* Numbers in parentheses are the rates for women aged 50–59 y/40–49 y.

<sup>†</sup> Calculated from data in Kerlikowske et al (19); women without personal or close family history of breast cancer.

<sup>‡</sup> Calculated from data of Sickles (43).

<sup>§</sup> Includes women with or without a close family history of breast cancer.

<sup>||</sup> Calculated from data of Kopans et al (39,40).

30–39 years in the group younger than the age of 50 years is irrelevant to the screening guideline controversy because routine screening of women aged 30–39 years is not recommended by any medical organization (1–3). Moreover, grouping data in this manner may lead to erroneous conclusions.

It can be appreciated from Table 1 (1,39, 40,43) that detection rates for the first screening of women at UCSF (43) increase gradually with age, rather than abruptly at the age of 50 years. By using other groupings, it can be calculated from this table that 4.2 times as many cancers (13.7/3.3) per 1,000 screening examinations were detected among women aged 60–70+ years as among women aged 30–59 years. This grouping would falsely suggest not to begin screening until the age of 60 years. An alternative grouping of detection rates for women aged 40–70+ years versus those aged 30–39 years would yield 9.0 versus 1.3 cancers per 1,000 women and would imply that screening could begin at the age of 40 years. Screening women aged 40–70+ years versus those aged 30–39 years would enable detection of 6.9 times as many cancers and would seem more convincing than the 5 times higher detection yield for screening women aged 50–70+ years versus those aged 30–49 years.

A similarly inappropriate process of combining data for women aged 50–69 years for comparison with data for women aged 40–49 years was used in recent studies of screening detection rates by Harris and Leininger (34) and of cost-effectiveness of screening mammography by Salzmann et al (37). Antman and Shea (44) calculated that screening mammography performed every 18–28 months for a decade would save six lives for women aged 60–69 years versus only one life for women aged 40–49 years. Their conclu-

sion not only underestimates the benefit from screening women in their 40s but also excludes comparative data for women aged 50–59 years. None of these approaches is as valid as comparing results for a single decade with those for the next decade or assessing results for even smaller annual age increments, as Kopans et al (39,40) did.

Moreover, comparing data for initial (prevalence) screening is less meaningful than comparing data for subsequent (incidence) screening, where differences in detection rates by age group are much smaller, as evident from the UCSF and MGH data shown in Table 1. In fact, when UCSF data for women aged 60–70+ years and those aged 40–59 years are compared, detection rates on the second screening are similar (ie, 2.2 vs 2.1 cancers per 1,000 women). Detection rates for the first combined with subsequent screenings for these same two age groups at MGH were 4.5 versus 2.7 cancers per 1,000 women, respectively. Moreover, in a population undergoing annual screening, incidence screenings are much more representative because after the first prevalence screening, incidence screenings are the ones occurring over the remainder of the lifetime of the patient. Even when comparing 1 decade of data for women who begin annual screening at the age of 40 years versus those who begin at the age of 50 years, only 10% of the screenings are prevalence screenings versus 90% that are incidence screenings.

Claiming “a 5 times higher detection yield” on initial screening for UCSF women 50 years of age or older versus those younger than 50 years, as Kerlikowske et al (19) have done, misrepresents data with respect to the screening issue for those aged 40–49 years because on this initial screening, detection rates for women aged 50–59 years versus those

aged 40–49 years were 2.2 times higher. The authors could also have compared data for women aged 50–59 years versus those aged 40–49 years from the second screening, although the relatively smaller number of women aged 40–59 years in their second screening database could limit the accuracy of the results. The use of more recent data from UCSF by Sickles (43), which involved the first and subsequent screenings of many more women, indicates that the detection yield for women in their 50s was only 1.6 times higher than for those in their 40s. These results are similar to those of Kopans et al (40), which show a detection rate 1.2 times higher for women aged 50–59 years, compared with those in the earlier decade, when both prevalence and incidence screening data are included (Table 2). Therefore, the use of prevalence screening data alone may lead to erroneous conclusions.

The much smaller age-related differences in detection rates when subsequent rather than initial screenings are compared may be explained in terms of age-related differences in screening detection lead time. Detection lead times for women aged 40–49 years are shorter mainly because of their faster breast cancer growth rates when compared with those of older women (7). To an extent, depending on lead time, initial screening detects not only cancers that would otherwise have surfaced clinically in the next year but also in succeeding years as well. Detection rates for the second screening will be lower than those for the initial screening because cancers in excess of the annual breast cancer incidence will have been detected already on the initial screening.

Detection rates for the second and subsequent screenings will reflect more nearly the expected annual breast cancer incidence and will depend on lead time to a considerably lesser extent than do detection rates for the initial screening. In the absence of a sudden increase in screening sensitivity caused by improvement in mammographic technique or interpretation, detection rates for subsequent screenings should not exceed the annual incidence rate.

#### **RECALL RATES AND DIAGNOSTIC IMAGING PROCEDURES**

The same inappropriate methods of age group stratification and data selection were used by Kerlikowske et al (19) to arrive at their statement that there were “48.3 diagnostic procedures per cancer

detected below age 50 vs 14.8 diagnostic procedures per cancer detected above age 50." On the basis of their own published data from UCSF, I can show that the number of diagnostic procedures per cancer detected decreases steadily with increasing age (Table 3). According to the method used by these authors, one could lump other age group data together to show that there would be 42.3 procedures per cancer detected for ages 30–59 years versus 10.2 procedures for ages 60–69 years and therefore that women should not be screened below the age of 60 years. This difference is actually proportionately greater (4.1 times) than the difference between groups aged 30–49 years versus those aged 50–70+ years (3.3 times).

A different conclusion that supports screening women aged 40 years and older would be reached by comparing women aged 30–39 years with women aged 40–69 years, where there would be 61.1 and 25.3 procedures per cancer detected, respectively. (The authors calculate procedures per cancer by dividing the total number of procedures by the total number of cancers for the age groups combined, whereas I have calculated this parameter by averaging the ratios of procedures per cancer for each of the age group decades because this latter method will not be influenced by the age distribution of the UCSF population and would be applicable to a population of uniform age group distribution.)

The authors' statement that "women younger than 50 years had a threefold higher number of diagnostic tests than women aged 50 years or older" (19) actually compared women aged 30–49 years with those aged 50–69 years. A more appropriate comparison would be that women aged 40–49 years required 2 times as many diagnostic tests per cancer diagnosed than those aged 50–59 years, just as women aged 50–59 years needed 2 times as many diagnostic tests per cancer diagnosed as women aged 60–69 years (Table 3).

In the UCSF study, most "diagnostic procedures" required for evaluation of abnormal screening mammograms involved nothing more than additional "diagnostic" mammographic views, ultrasonography (US), and follow-up mammography, rather than biopsy. Moreover, for women aged 40–49 years, only 6.3% of the mammograms on the first screening and 2.0% of the mammograms on each subsequent screening were interpreted as abnormal. Only 25% of all abnormal mammograms required biopsy,

**TABLE 3**  
Measures of Specificity of Screening Mammography at UCSF and MGH according to Age

| Measure and Screening                                     | Age (y) |       |       |       |       |
|---|---------|-------|-------|-------|-------|
|   | 30–39   | 40–49 | 50–59 | 60–69 | 70–79 |
| Mammographic interpretation abnormal (%)                  |         |       |       |       |       |
| UCSF (first screening)*†                                  | 4.0     | 6.3   | 6.6   | 7.8   | 7.3   |
| UCSF (second screening)*†                                 | 2.3     | 2.0   | 1.9   | 2.0   | 1.4   |
| MGH (first and subsequent screenings)‡§                   | NA      | 7.0   | 6.9   | 6.0   | 5.6   |
| Diagnostic procedures/breast cancer diagnosed             |         |       |       |       |       |
| UCSF (first screening)*                                   | 61.1    | 43.9  | 21.9  | 10.2  | NA    |
| Abnormal screening examinations/breast cancer             |         |       |       |       |       |
| UCSF (first screening)*                                   | 31      | 22.5  | 11.1  | 6.0   | NA    |
| UCSF (second screening)*                                  | 16      | 15.3  | 6.3   | 14.5  | 4.5   |
| PPV <sub>1</sub> : cancers/abnormal screening examination |         |       |       |       |       |
| UCSF (first screening)*                                   | 0.03    | 0.04  | 0.09  | 0.17  | NA    |
| UCSF (second screening)*                                  | 0.06    | 0.07  | 0.16  | 0.07  | 0.22  |
| MGH (first and subsequent screenings)‡#                   | NA      | 0.03  | 0.05  | 0.07  | 0.09  |

Note.—NA = not applicable.

\* Women without breast lumps or personal or family history of breast cancer.

† Data from Kerlikowske et al (19).

‡ Includes women with or without personal or family history of breast cancer.

§ Calculated from data of Kopans et al (39).

|| Calculated from data of Kerlikowske et al (19).

# Calculated from data of Kopans et al (39,40).

that is, 1.6% of the mammograms on the first screening and 0.5% or less of the mammograms on each subsequent screening. Furthermore, when a patient had undergone both additional mammographic views and US studies, these were counted by Kerlikowske et al (19) as two diagnostic procedures. When a patient underwent excisional biopsy following needle localization for a nonpalpable lesion, these were also counted as two procedures rather than one. Across all age groups at the initial screening at UCSF, there were approximately two diagnostic procedures for every abnormal screening mammogram (Table 3).

The percentage of screening studies at UCSF that were interpreted as abnormal showed slight, if any, variation with patient age, especially for the second screening, where only about one-third as many studies were interpreted as abnormal compared with the first screening (Table 3). Instead of summarizing their results with the statement that on the first screening, there were "48.3 diagnostic procedures/cancer diagnosed in women younger than 50 years compared with 14.8 diagnostic procedures in women age women age 50 or older" (19), it would have been more accurate if the authors had stated that on the second screening, there were 15.3 abnor-

mal mammograms for every cancer diagnosed among women aged 40–49 years versus 10.4 abnormal mammograms per cancer diagnosed among women aged 50–69 years (Table 3). I have calculated these results from data in their article (19).

Unfortunately, the authors did not include any data on the number of procedures per cancer diagnosed for the second screening. However, it is possible to evaluate the effect of age on the ratio of the number of cancers divided by the number of abnormal screening examinations, otherwise known as the PPV<sub>1</sub>. It can be appreciated from Table 3 that the PPV<sub>1</sub> of data from both UCSF and MGH increases steadily with age. There is also no abrupt change at the age of 50 years.

Because stability in the mammographic findings at periodic examinations supports a benign diagnosis, the proportion of screening cases recalled for additional diagnostic imaging work-up will be lower at incidence screening because previous mammograms are available for comparison. Therefore, the PPV<sub>1</sub> values for women of all ages will be higher on incidence screening than on prevalence screening (Table 3). However, this increase in the PPV<sub>1</sub> values will be partially restrained by the fact that detection rates are lower on

**TABLE 4**  
**Positive Predictive Value for Biopsies Performed (PPV<sub>3</sub>) among Women Screened at UCSF and MGH according to Age**

| Measure and Screening                    | Age (y) |       |       |       |       |
|--|---------|-------|-------|-------|-------|
|  | 30-39   | 40-49 | 50-59 | 60-69 | 70-79 |
| PPV <sub>3</sub> : cancers/biopsy        |         |       |       |       |       |
| UCSF (first screening)*†                 | 0.15    | 0.17  | 0.30  | 0.46  | NA    |
| UCSF (first and subsequent screenings)‡§ | 0.16    | 0.26  | 0.35  | 0.43  | 0.55  |
| MGH (first and subsequent screenings)‡   | NA      | 0.17  | 0.24  | 0.32  | 0.40  |
| Biopsies/cancer: 1/PPV <sub>3</sub>      |         |       |       |       |       |
| UCSF (first screening)*#                 | 6.7     | 5.9   | 3.3   | 2.2   | NA    |
| UCSF (first and subsequent screenings)‡§ | 6.3     | 3.8   | 2.9   | 2.3   | 1.8   |
| MGH (first and subsequent screenings)‡   | NA      | 5.9   | 4.2   | 3.1   | 2.5   |

Note.—NA = not applicable.

\* Women without breast lumps or personal or family history of breast cancer.

† Data from Kerlikowske et al (19).

‡ Includes women with or without personal or family history of breast cancer.

§ Calculated from data of Sickles (43).

|| Calculated from data of Kopans et al (39,40).

# Calculated from data of Kerlikowske et al (19).

incidence screening than on prevalence screening. For example, Frankel et al (45) found that the frequency of abnormal interpretation was 2.3 times greater than the frequency of cancer detection for initial screening versus 1.5 times greater for subsequent screening at UCSF. Therefore, for all age groups combined, the PPV<sub>1</sub> was 1.5 times higher (2.3/1.5) on subsequent screening.

Because of faster breast cancer growth rates and the consequent shorter detection lead times for younger women, differences in detection rates and, hence, differences in PPV<sub>1</sub> values between women aged 40-49 years and those aged 50-59 years will be less for subsequent (incidence) screenings compared with the initial (prevalence) screening. For example, it can be calculated that PPV<sub>1</sub> ratios for women aged 50-59 years versus those aged 40-49 years are 2.2 times higher (0.09/0.04) for initial screening at UCSF versus 1.7 times higher (0.05/0.03) by using data from both initial and subsequent screenings at MGH (Table 3).

A misperception about diagnostic procedures resulting from screening women aged 40-49 years can be found in a recent editorial in *JAMA* by Davis and Love (21). The authors stated that a National Cancer Institute International Workshop on Screening of Breast Cancer had found that "for every 1,000 women under 50 years of age screened with mammography in a decade, 700 would require some sort of diagnostic procedure." It is hard to imagine how the authors (21) arrived at this estimate, which is at marked variance from the recall rate of 6.3% at the first screening and 2% at the second screening

for women aged 40-49 years at UCSF. For 10 annual screenings, this might result in a recall of 250 women for additional work-up for every 1,000 screened (6.3% + [2% × 9] = 24%). Probably even fewer women would be recalled because callback rates would still decline slightly at the third and subsequent screenings. Possibly the overestimate by Davis and Love (21) resulted from a misassumption that recall rates of 6%-7% on the initial screening remain unchanged on subsequent screenings (ie, 10 × 0.07 × 1,000 = 700).

Such overestimates of the number of diagnostic procedures that result from screening are compounded by the fact that many readers equate diagnostic procedures with biopsies. Nothing could be farther from the truth. In reality, nearly all diagnostic procedures are noninvasive, and most are nothing more than supplementary mammographic views obtained at different angles or with more focal breast compression or more magnified than routine mammographic views. More recently, a similar overestimation of cumulative 10-year recall rates was made by Elmore et al (46) on the basis of extrapolation from observed rates for women undergoing mammography on average every 2-3 years (47).

#### BIOPSY POSITIVE PREDICTIVE VALUES

The positive biopsy rate, which varies according to the age of the patient, represents a major consideration in the determination of screening guidelines. The positive biopsy rate, or PPV<sub>3</sub>, is the ratio

of the number of cancers detected divided by the number of biopsies performed (48). In contradistinction, PPV<sub>2</sub> is the number of cancers divided by the number of biopsies advised, and PPV<sub>1</sub> is the number of cancers divided by the number of abnormal screening examinations (48). To an epidemiologist, PPV is synonymous with PPV<sub>1</sub> (48), whereas to a radiologist, PPV is synonymous with PPV<sub>3</sub> (positive biopsy rate) (48). These differences in semantics contribute to misinterpretation of screening data.

In their final table, Kerlikowske et al (19) summarized the UCSF results at the first screening as showing "2 cancers for every 13 biopsies per thousand women screened below age 50" (PPV<sub>3</sub> = 0.15) versus "10 cancers for 25 biopsies per thousand women screened age 50 and over" (PPV<sub>3</sub> = 0.4) (19). This table has been widely quoted and in fact was reproduced in the *Annals of Internal Medicine* in a subsequent editorial by Sox (22), where it represented a major argument against screening women younger than 50 years old.

Although superficially persuasive, this oversimplified conclusion is misleading. It is not apparent from the table that to the authors, women younger and older than 50 years old refers to those aged 30-49 years and aged 50-69 years, respectively. There are two reasons why comparing combined data for women in their 30s and 40s with combined data for women in their 50s and 60s is inappropriate. First, this practice will accentuate differences between data for women in the 40s and 50s. Second, screening should not be routinely performed on women aged 30-39 years.

The PPV<sub>3</sub> values at the first screening at UCSF increase with age for each successive decade (Table 4). The PPV<sub>3</sub> values for the combined first and subsequent screenings, which are based on more inclusive UCSF data published by Sickles (43) and the MGH data from Kopans et al (39,40), are also shown in Table 4. It can be appreciated that the differences in PPV<sub>3</sub> by decade are much smaller when data for those combined first and subsequent screenings are considered. Moreover, those combined PPV<sub>3</sub> values are much higher than those for the first screening alone.

The summary by Kerlikowske et al (19) of PPV<sub>3</sub> data on the initial screening for women aged 30-49 years versus those 50-69 years implies that biopsy PPVs for women older than 50 years old are 2.7 times higher (0.40/0.15). However, data from Sickles (43) and Kopans et al (39,40)

indicate that when more appropriate combined first and subsequent screening data for single decades are compared, there is only a 1.3–1.4 times increase in  $PPV_3$  (0.35/0.26 to 0.24/0.17) from women aged 40–49 years to those aged 50–59 years. This is similar to the increase from women aged 50–59 years to those aged 60–69 years or from women aged 60–69 years to those aged 70 years and older.

## BIOPSY-CANCER RATIOS

Screening results may also be evaluated by means of the ratio of the number of biopsies performed divided by the number of cancers found at biopsy (the inverse of  $PPV_3$ ). This ratio indicates the number of biopsies that need to be performed to diagnose one cancer. Lower biopsy-cancer ratios mean that fewer biopsies are needed to find the same number of cancers. In another report of the same UCSF data, Esserman and Kerlikowske (20) stated the following: “Compared with women age 50 and older, women under age 50 underwent 2.5 times more biopsies to diagnose one-fifth as many cancers.” This statement implies that the biopsy-cancer ratio for women in their 40s should be 12.5 times higher ( $2.5 \times 5$ ) than that for older women. As such, it would represent a powerful argument against screening these younger women. Although this interpretation would be reasonable on the basis of the way that the authors’ sentence is phrased, it would not be correct.

The authors derived 2.5 by comparing biopsy-cancer ratios for women aged 30–49 years with those for women aged 50–69 years (6.2/2.5) and obtained their cancer detection rates for women in the same age groups. However, even if one accepts the practice of comparing these two 20-year age groups and of using data only from the initial screening, their statement is still not internally consistent. It would have been more reasonable to state that on the initial screening, compared with women aged 50–69 years, women aged 30–49 years required 2.5 times as many biopsies per cancer detected and that one-fifth as many cancers were detected in the younger age group (Tables 1, 4).

Such a statement would clearly indicate that the biopsy-cancer ratio of the younger age group was 2.5 times rather than 12.5 times higher! However, even that statement would represent a gross exaggeration of the age-related difference in biopsy-cancer ratios. For the reasons discussed in relation to detection rates

and examination specificity, the only completely valid comparison with regard to the screening controversy for women in their 40s is one which (a) compares women aged 40–49 years with those aged 50–59 years, (b) does not also include women aged 30–39 years or 60–69 years, and (c) is based on incidence screening data, not prevalence screening data only or even combined prevalence and incidence screening data.

Relative biopsy-cancer ratios for women aged 40–49 years versus those aged 50–59 years can be calculated from the data in Table 4. Compared with biopsy-cancer ratios based on prevalence screening data only, these are lower when incidence screening data are combined with prevalence screening data. For example, the relative biopsy-cancer ratio is 1.8 (5.9/3.3) for the initial screening at UCSF (women with personal or family history of breast cancer excluded) on the basis of data in the article by Kerlikowske et al (19); this ratio is 1.3 (3.8/2.9) for the combined initial and subsequent screenings at UCSF on the basis of calculations from data from a larger number of screenings reported by Sickles (43) and is 1.4 (5.9/4.2) for the combined initial and follow-up screening data at MGH reported by Kopans et al (39,40). Compared with results from prevalence screening alone, combining results from both prevalence and incidence screenings allows a more valid assessment of relative risk of false-positive biopsies for women aged 40–49 years.

## BIOPSY RATES

There is an unsubstantiated concern about excessive biopsy rates from screening women in their 40s. In a letter to the editor of the *Journal of the National Cancer Institute*, Harris claimed that if annual screening were performed for a decade on 1,000 women during their 40s, “about 300 women (30%) would have had an invasive procedure (eg, biopsy or aspiration)” (35). Harris did not indicate how he derived these data. On the basis of published biopsy rate data for women in their 40s, this represents a gross overestimate of the intervention rates.

As opposed to the biopsy rate estimated by Harris (35), Kopans et al (40) have reported a biopsy rate of approximately 1.2% for women, regardless of age, and a  $PPV_3$  of 0.2 for women in their 40s, which translates to 120 biopsies and 24 cancers if 1,000 women were screened annually for 10 years beginning at the age of 40 years at MGH. Sickles (43) reports a simi-

lar biopsy rate of 1.2% and a  $PPV_2$  of 0.26, which would mean 130 biopsies performed and 34 cancers detected at UCSF if 10 annual screenings were performed on 1,000 women beginning at the age of 40 years.

## FACTORS AFFECTING AGE-RELATED ACCURACY

Factors that affect the accuracy of mammography include the breast cancer incidence and the mammographic parenchymal density. Because the incidence of breast cancer increases with age, there will be a concomitant increase in the prior probability of disease (prevalence and incidence of cancer in a population over a given period of time), the screening detection rates, and the  $PPV_1$ ,  $PPV_2$ , and  $PPV_3$  values, as well as specificity.

With increasing age, there may also be a gradual replacement of fibroglandular tissue by fatty tissue (49), which affords better visualization and characterization of mammographic abnormalities caused by an improvement in mammographic exposure and contrast (50). However, breast density cannot be used to accurately predict patient age because some younger women have fatty breasts, while some older women may have dense breasts. More important, differences in breast density between women aged 40–49 years and those aged 50–59 years are relatively small (49). Although there is probably a correspondingly small effect on differences in mammographic accuracy between these age groups, these differences provide no reason as to why screening mammography should not be almost as effective for women in their 40s as for those in the next decade.

The observation that there is a much greater age-dependent variation in detection rates on the initial screening than on the second and subsequent screenings (Tables 1, 2) reflects the fact that the age-related differences in lead time are mainly caused by age-related differences in breast cancer growth rates, rather than differences in the mammographic breast density.

There is evidence from histologic evaluation of tumors detected at screening trials that tumor growth rates, along with progression in tumor grade and nodal status, are faster for women in their 40s than for older women (7). Therefore, screening detection lead times and sojourn times are shorter for women in their 40s than for older women (7). Cancers detected through screening women

aged 40–49 years have a similar stage of disease and prognosis as those detected through screening older women. However, interval cancers (cancers that surface between screenings) have a worse prognosis than screening-detected cancers and are more common before the age of 50 years than after (25). Therefore, interval cancers lead to a lower screening detection rate and a lower screening sensitivity for women younger than the age of 50 years, compared with older women (25). As a consequence, annual screening intervals are especially important for women in their 40s (1,7,24,25).

## CONCLUSION

Pooling data for women aged 40–49 years with data from younger women, pooling data for women aged 50–59 years with data from older women, and the exclusive use of data from the initial (prevalence) screening result in an inaccurate portrayal of screening outcomes for women in their 40s. Such improper assessment will lead to the erroneous impression that screening women younger than 50 years old will detect only 20% as many cancers per 1,000 women screened, will require 4 times as many diagnostic procedures per cancer detected, and will cause 2.5 times as many false-positive biopsies for each cancer detected, compared with screening older women.

Proper assessment of the accuracy of screening mammography for women aged 40–49 years requires comparison of data from that age group only with data for women aged 50–59 years. The use of data from initial (prevalence) screening alone may be misleading. The use of data from subsequent (incidence) screening alone is preferred, but combined data from prevalence and incidence screenings may also be used. Such assessment will indicate that screening women aged 40–49 years will detect at least 63%–80% as many cancers (Table 1), require 1.7 times as many diagnostic imaging procedures (Table 3), and result in 1.3–1.4 times as many false-positive biopsy results for cancers detected (Table 4).

The increased sensitivity and specificity of screening at ages 50–59 years compared with ages 40–49 years are similar to those at ages 60–69 years compared with ages 50–59 years and are similar to those at age 70+ years compared with ages 60–69 years. Although mammography becomes more accurate with increasing age, there is no abrupt change in accuracy at the age of 50 years. Comparison of ben-

efits with risks is almost as favorable for women in their 40s as for those in their 50s and supports screening in both age groups. Moreover, analysis of published data indicates that detection rates and the need for further diagnostic imaging and breast biopsy from annual screening of women aged 40–49 years can be kept well within acceptable limits.

## References

1. Feig SA, D'Orsi CJ, Hendrick RE, et al. American College of Radiology guidelines for breast cancer screening. *AJR Am J Roentgenol* 1998; 171:29–33.
2. Leitch AM, Dodd GD, Costanza M, et al. American Cancer Society guidelines for the early detection of breast cancer: update 1997. *CA Cancer J Clin* 1997; 47:150–153.
3. Stapleton S. House: annual mammograms at 40. *American Medical News* 1999; 42(no. 27):27, 29.
4. Eastman P. NCI adopts new mammography screening guidelines for women. *J Natl Cancer Inst* 1997; 89:538–540.
5. Hendrick RE, Smith RA, Rutledge JH III, Smart CR. Benefit of screening mammography for women aged 40–49: a new meta-analysis of randomized controlled trials. *Monogr Natl Cancer Inst* 1997; 22:87–92.
6. Smart CR, Hendrick RE, Rutledge JH III, Smith RA. Benefit of mammography screening in women ages 40–49 years: current evidence from randomized controlled trials. *Cancer* 1995; 75:1619–1626 [Erratum: *Cancer* 1995; 75:2788.]
7. Falun Meeting Organizing Committee and Collaborators. Breast-cancer screening with mammography in women aged 40–49 years: Swedish Cancer Society and the Swedish National Board of Health and Welfare—report of the organizing committee and collaborators, Falun Meeting, Falun, Sweden (21 and 22 March 1996). *Int J Cancer* 1996; 68:693–699.
8. Andersson I, Janzon L. Reduced breast cancer mortality in women under age 50: updated results from Malmö Mammographic Screening Program. *Monogr Natl Cancer Inst* 1997; 22:63–67.
9. Bjurstam N, Bjorneld L, Duffy SW, et al. The Gothenburg breast screening trial: first result on mortality, incidence, and mode of detection for women ages 39–49 years at randomization. *Cancer* 1997; 80:2091–2099.
10. Chu KC, Smart CR, Tarone RE. Analysis of breast cancer mortality and stage distribution by age for the Health Insurance Plan clinical trial. *J Natl Cancer Inst* 1988; 80:1125–1132.
11. Shapiro S, Venet W, Strax P, Venet L. Periodic screening for breast cancer: the Health Insurance Plan Project and its sequelae, 1963–1986. Baltimore, Md: John Hopkins University Press, 1988.
12. 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.
13. Feig SA. Methods to identify benefit from mammographic screening of women aged 40–49 years. *Radiology* 1996; 201:309–316.
14. Smart CR, Hartmann WH, Beahrs OH, Garfinkel L. Insights into breast cancer screening of younger women: evidence from the 14-year follow-up of the Breast Cancer Detection Demonstration Project. *Cancer* 1993; 72:1449–1456.
15. Sickles EA, Kopans DB. Deficiencies in the analysis of breast cancer screening data. *J Natl Cancer Inst* 1993; 85:1621–1624.
16. Fletcher SW, 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.
17. Kopans DB, Feig SA. The Canadian National Breast Screening Study: a critical review. *AJR Am J Roentgenol* 1993; 161:755–760.
18. Black WC, Nease RF Jr, Tosteson ANA. Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. *J Natl Cancer Inst* 1995; 87:720–726.
19. 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; 270:2444–2450.
20. Esserman L, Kerlikowske K. Should we recommend screening mammography for women aged 40–49? *Oncology* 1996; 10:357–365.
21. Davis DL, Love SM. Mammographic screening (editorial). *JAMA* 1994; 271:152–153.
22. Sox HC. Screening mammography in women younger than 50 years of age (editorial). *Ann Intern Med* 1995; 122:550–552.
23. Kerlikowske K, Barclay J. Outcomes of modern screening mammography. *Monogr Natl Cancer Inst* 1997; 22:105–111.
24. Feig SA. Determination of mammographic screening intervals with surrogate measures for women aged 40–49 years. *Radiology* 1994; 193:351–358.
25. Feig SA. Estimation of currently attainable benefit from mammographic screening of women aged 40 years. *Cancer* 1995; 75:2412–2419.
26. Feig SA. Strategies for improving the sensitivity of screening mammography for women aged 40–49 years. *JAMA* 1996; 276:73–74.
27. Feig SA. Increased benefit from shorter screening mammography intervals for women ages 40–49 years. *Cancer* 1997; 80:2035–2039.
28. Feig SA. Mammography screening: published guidelines and actual practice. *Recent Results Cancer Res* 1987; 105:78–84.
29. National Institutes of Health Consensus Development Conference Statement: breast cancer screening for women ages 40–49, January 21–23, 1997. *J Natl Cancer Inst* 1997; 89:1015–1026.
30. Wells J. Mammography and the politics of randomized controlled trials. *Br Med J* 1998; 317:1224–1229.
31. Berry DA. Benefits and risks of screening mammography for women in their forties: a statistical reappraisal. *J Natl Cancer Inst* 1998; 90:1431–1439.
32. De Koning HJ, Boer R, Warmerdam PG, Beemsterboer PM, van der Maas PJ. Quantitative interpretation of age-specific mortality reductions from Swedish breast cancer-screening trials. *J Natl Cancer Inst* 1995; 87:1217–1223.
33. Tabar L, Duffy SW, Chen HH. Re: Quantitative interpretation of age-specific mortality reductions from the Swedish Breast Cancer-Screening Trials (letter). *J Natl Cancer Inst* 1996; 88:52–55.

34. Harris R, Leininger L. Clinical strategies for breast cancer screening: weighing and using the evidence. *Ann Intern Med* 1995; 122:539–547.
35. Harris R. Re: Efficacy of screening mammography for women in their forties (letter). *J Natl Cancer Inst* 1994; 86:1722–1724.
36. Gail M, Rimer B. Risk-based recommendations for mammographic screening for women in their forties. *J Clin Oncol* 1998; 16:3105–3114.
37. Salzman P, Kerlikowske K, Phillips K. Cost-effectiveness of extending screening mammography guidelines to include women 40 to 49 years of age. *Ann Intern Med* 1997; 127:955–965.
38. American College of Radiology Breast Cancer Task Force. Risk of risk-based mammography screening, ages 40 to 49 (letter). *J Clin Oncol* 1999; 17:735–737.
39. Kopans DB, Moore RH, McCarthy KA, et al. Positive predictive value of breast biopsy performed as a result of mammography: there is no abrupt change at age 50 years. *Radiology* 1996; 200:357–360.
40. Kopans DB, Moore RB, McCarthy KA, et al. Biasing the interpretation of mammography screening data by age grouping: nothing changes abruptly at age 50. *Breast J* 1998; 4:139–145.
41. Kopans DB. The use of mammography for screening (letter). *JAMA* 1994; 271:982.
42. Kerlikowske K. The use of mammography for screening—reply (letter). *JAMA* 1994; 271:982–983.
43. Sickles EA. Auditing your practice. In: Kopans DB, Mendelson EB, eds. *Syllabus: a categorical course in breast imaging*. Oak Brook, Ill: Radiological Society of North America, 1995; 81–91.
44. Antman K, Shea S. Screening mammography under age 50. *JAMA* 1999; 281:1470–1472.
45. Frankel SD, Sickles EA, Curpen BN, Sollito RA, Ominsky SH, Galvin HB. Initial versus subsequent screening mammography: comparison of findings and their prognostic significance. *AJR Am J Roentgenol* 1995; 164:1107–1109.
46. 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; 338:1089–1096.
47. Feig SA. A perspective on false positive screening mammograms. *ACR Bull* 1998; 54:8, 13.
48. American College of Radiology BI-RADS Committee. *Breast imaging reporting and data system*. 3rd ed. Reston, Va: American College of Radiology, 1998.
49. Stomper PC, D'Souza DJ, DiNitto PA, Arredondo MA. Analysis of parenchymal density on mammograms in 1353 women 25–79 years old. *AJR Am J Roentgenol* 1996; 167:1261–1265.
50. Jackson VP, Hendrick RE, Feig SA, Kopans DB. Imaging of the radiographically dense breast. *Radiology* 1993; 188:297–301.