Radiology

Long-term Risk of False-Positive Screening Results and Subsequent Biopsy as a Function of Mammography Use[1]

Karen Blanchard, BA
James A. Colbert
Daniel B. Kopans, MD
Richard Moore, MA
Elkan F. Halpern, PhD
Kevin S. Hughes, MD
Barbara L. Smith, MD, PhD
Kenneth K. Tanabe, MD
James S. Michaelson, PhD

Purpose: To retrospectively determine the long-term risk of false-positive mammographic assessments and to evaluate the effect of screening regularity on the risk of false-positive events.

Materials and Methods: Institutional review board approval was obtained, and informed consent was waived. Retrospective analysis was performed for the occurrence of false-positive assessments among 83 511 women who underwent 314 185 mammographic examinations from January 1, 1985, to February 19, 2002. Data were collected from a database that had been assembled prospectively. Two categories of false-positive events were examined: biopsies that did not reveal cancer and false-positive mammographic assessments. Rates of false-positive events were compared by using a χ^2 analysis, and 95% confidence limits were calculated. Because comparisons of multiple pairs were considered, all *P* values that demonstrated statistical significance exceeded the requirement of the Bonferroni correction.

Results: While the overall rates of biopsies that did not reveal cancer and of false-positive mammographic assessments were similar to those found in other studies, most of the burden of false-positive events was borne by women who underwent intermittent screening. Long-term rates of false-positive events were lower among women who underwent regular screening than among those who underwent intermittent screening. In the 5-year group, 2.9% of women who underwent five mammographic examinations over the next 5 years had biopsy results that did not reveal cancer, whereas 4.6% of women who underwent three mammographic examinations over the next 5 years had biopsy results that did not reveal cancer. For women who underwent regular screening, the risk of undergoing biopsies that did not reveal cancer declined over time to 0.25% per year after several years of screening, a value that is lower than the risk of these events among women who did not undergo screening. The rate of false-positive mammographic assessments was also lower for women who underwent regular screening than for those who underwent intermittent screening.

Conclusion: Prompt annual attendance for mammographic screening reduces the occurrence of false-positive mammographic results.

[1] From the Departments of Surgery (K.B., J.A.C., K.S.H., B.L.S., K.K.T., J.S.M.), Pathology (J.S.M.), and Radiology (D.B.K., R.M., E.F.H.), Massachusetts General Hospital, Yawkey 7939, 55 Fruit St, Boston, MA 02114; and Departments of Pathology (J.S.M.), Surgery (K.S.H., B.L.S., K.K.T.), and Radiology (D.B.K., E.F.H.), Harvard Medical School, Boston, Mass. Received January 21, 2005; revision requested March 23; revision received May 18; accepted June 21; final version accepted November 2. Supported by departmental funds from the Massachusetts General Hospital Division of Surgical Oncology. **Address correspondence to** J.S.M. (e-mail: *michaelj@helix.mgh.harvard.edu*).

While many lines of evidence suggest that mammographic screening reduces breast cancer death (1–3), a major drawback to screening is the occurrence of false-positive findings that are suggestive of cancer in women who are ultimately found to have no disease (4–7). Such false-positive findings involve additional costs (4,8,9), the use of medical services (10), anxiety (11,12), and biopsies performed in women who are ultimately found to have no disease (4–12). Understanding the actual occurrence of false-positive events and finding ways to minimize them are major research objectives in the field of breast cancer screening (13–17).

One major question has been the magnitude of the long-term occurrence of false-positive events, which might be expected to present a considerable burden over the course of each woman's screening history (which can span 30 years or more). A related question concerns the effect of the screening interval on the long-term occurrence of false-positive events. It has been long recognized that the availability of a previous mammogram can reduce the risk of a false-positive event (12–14). Because intermittent screening limits the availability of recent mammograms, it might be hypothesized that the frequency of screening may affect the incidence of false-positive events.

It has long been recognized that failure to undergo prompt screening is a widespread phenomenon (18–21). Thus, while several researchers have estimated the long-term accumulation of false-positive events (4,7), to our knowledge there are no actual measures of the long-term occurrence of these events. The large size and long span of time for the data contained within the Massachusetts General Hospital Avon Comprehensive Breast Center screening population database (20–22), which has information on more than 300 000 mammographic examinations performed in more than 83 000 women since 1985, made it possible for us to conduct this study, the purpose of which was to retrospectively determine the long-term risk of false-positive mammographic assessments and to evaluate the effect of screening regularity on the risk of false-positive events.

Materials and Methods

Database

The database of the Massachusetts General Hospital Avon Comprehensive Breast Center contains information on 314 185 mammograms obtained in 83 511 women from January 1, 1985, to February 19, 2002. This database also contains extracted information from hospital pathology reports on breast tissue examined during this period. Details on the demographic features of the population (eg, age, race, and cancer incidence) have been previously published (20–22,27,28).

Each of the mammographic entries contains one of 737 Massachusetts General Hospital assessment codes. These codes are entered by the radiologist at the time of the examination and make it possible to identify visits that correspond to a negative mammographic screening result. Codes can be converted into one of five Breast Imaging Reporting and Data System (BI-RADS) categories as follows: BI-RADS category 0, incomplete assessment and the need for additional imaging; BI-RADS category 1, negative findings; BI-RADS category 2, benign findings; BI-RADS category 3, probably benign findings and a recommendation for short-interval follow-up; BI-RADS category 4, suspicious abnormalities (biopsy should be considered); and BI-RADS category 5, highly suggestive of malignancy. Analysis of the database was performed by three authors (K.B., J.A.C., J.S.M.).

This retrospective study was compliant with the Health Insurance Portability and Accountability Act and had institutional review board approval; informed consent was waived.

Prior to 1993, data were not entered regarding whether mammography was performed for screening or for diagnostic purposes at the time of the examination. However, it was possible to determine which visits corresponded to a negative screening result by using information from the Massachusetts General Hospital assessment codes. Since 1993, all mammograms in the database have been classified as screening, diagnostic, or procedural at the time of the examination. Beginning in 1993, all screening mammograms were also specified as being either unilateral or bilateral.

Groups

The occurrence of false-positive events was determined among women in three overlapping groups. The 5-year group comprised all 16 853 women who received either a positive or a negative screening result in 1996 and whose experience was examined through December 31, 2000. The 8-year group comprised all 13 877 women who received either a positive or a negative screening result in 1993 and whose experience was examined through December 31, 2000. The 10-year group comprised all 12 972 women who received a negative screening result in 1991 and whose experience was examined from January 1, 1992, through December 31, 2001.

These three study groups allowed for complementary information. For example, data obtained from the 10-year group not only provided information over the longest period of time but also permitted an estimate of the frequency of false-positive events among women who did not undergo screening at the Massachusetts General Hospital Avon Comprehensive Breast Center. Because this group differed from the other two groups in that it comprised only women

Published online
10.1148/radiol.2402050107

Radiology 2006; 240:335–342

Abbreviation:
BI-RADS = Breast Imaging Reporting and Data System

Author contributions:
Guarantors of integrity of entire study, all authors; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; literature research, K.B., J.A.C., J.S.M.; clinical studies, K.S.H., B.L.S., K.K.T.; statistical analysis, K.B., J.A.C., E.F.H., J.S.M.; and manuscript editing, K.B., J.A.C., J.S.M.

Authors stated no financial relationship to disclose.

who had received a negative screening result in the 1st year, the 10-year group contained a subgroup of women for whom there was no record of additional screening. Thus, while the 10-year group provided information on the occurrence of false-positive events over the longest period of time, the relatively small number of women in each subgroup resulted in this group having little statistical power with regard to individual subgroup analysis. In contrast, the larger number of women in each subgroup for the 5-year group made such comparisons possible.

It was also possible to examine how the risk of false-positive findings could change over time in the 5- and 8-year groups because both groups comprised women who received any screening mammography result (positive or negative). Thus, in these two groups, the rate of false-positive findings could be measured in the 1st year; this was not possible in the 10-year group, because the nature of the record set in 1991 (see above) required it to comprise women who received a negative screening result in the 1st year. Women with biopsy-proved breast cancer prior to the index years of the study were excluded from analysis, which was performed by three authors (K.B., J.A.C., J.S.M.).

Events and Definitions

Two broad categories of false-positive events were examined: *(a)* biopsies that did not reveal cancer (ie, biopsy results that indicated neither invasive breast cancer nor ductal carcinoma in situ) and *(b)* false-positive mammographic assessments (ie, mammographic results that were classified as BI-RADS category 0, 3, 4, or 5 findings yet ultimately did not lead to a pathologic diagnosis of breast cancer). Note that the values for women with different BI-RADS assessments were not additive because many screening visits led to more than one assessment. For example, a woman who is given a BI-RADS category 0 assessment at screening mammography (incomplete assessment, needs additional imaging evaluation) may later be given a BI-RADS category 3, 4, or 5 assessment at further imaging.

False-positive events were analyzed among subgroups of women who differed according to the number of mammographic examinations they chose to undergo within a specific period of time. For example, among the women in the 10-year group, almost 20% never came back after receiving a negative result for their index mammogram in 1991, fewer that 5% returned for 10 bilateral screening examinations, and approximately 10% underwent anywhere from one to nine of the 10 possible screening examinations during the 10-year period (Table 1).

In assigning women to one of these subgroups, only bilateral screening mammograms were considered so as to avoid any overestimation of the use of screening among women who had undergone an extra diagnostic examination or 6-month follow-up. However, because the occurrence of a false-positive event could have originated from a screening mammogram, diagnostic mammogram, or nonmammographic finding, women were scored as having a false-positive finding for all three categories.

We lack information on the fraction of women in the Massachusetts General Hospital Avon Comprehensive Breast Center screening population who underwent mammography elsewhere or who continued their medical care at another hospital. However, there are a number of indications that most women who began screening at the Massachusetts General Hospital Avon Comprehensive Breast Center remained in the data set. In previous studies on the whole population of women who underwent screening mammography in 1992 (20,21), 82% eventually returned; for a subset of this population, which included women who underwent previous screening mammography (ie, prior to 1996) within 1.1 years of the 1996 visit, 93% returned.

The categorizations reported here for the 10-year group are roughly comparable to those reported by Elmore et al (4) for a Health Maintenance Organization population. There are, however, minor differences. Our definition of a false-positive event was slightly broader than the definition used by Elmore et al in that both false-positive mammographic assessments and biopsy results that did not reveal cancer were counted as false-positive events, regardless of whether they were the result of screening mammography, were elicited from a diagnostic mammogram, or were determined by means of other factors, such as the detection of a palpable mass. Indeed, the occurrence of events that were not associated with mammography could be seen among women in the 10-year group who never returned for additional mammography after receiving a negative result for their index mammogram in 1991.

Statistical Analysis

The rates of false-positive events were compared by using a χ^2 analysis, and 95% confidence limits were calculated (Excel 2000, Microsoft, Redmond, Wash, and SAS, SAS Institute, Cary, NC). Because comparisons of multiple pairs were considered, all *P* values that demonstrated statistical significance exceeded the requirement for the Bonferroni correction. *P* values of less than .05 were considered to indicate a significant difference.

Results

Biopsy Results That Did Not Reveal Cancer

Of the women who received a negative screening result in 1991, 8.04% underwent biopsy that did not reveal cancer within the next 10 years. However, an analysis of subgroups for these women, who differed according to the number of screening procedures they chose to undergo, indicated that much of the burden of false-positive results was borne by women who underwent intermittent screening (Tables 1–3). Thus, while there were 26 women in whom biopsy results did not reveal cancer per 1000 mammograms in the population as a whole, there were only nine women in whom biopsy results did not reveal cancer per 1000 mammograms in the subpopulation of women who chose to undergo 10 mammographic examinations during the 10-year period (Fig 1). In contrast, 57 women in whom biopsy

Table 1

History of False-Positive Events in 10-year Group for Women Who Underwent Screening Beginning in 1991

Parameter	Population of Women Who Underwent Screening in 1991	No. of Mammographic Screening Procedures over 10-year Period										
		0	1	2	3	4	5	6	7	8	9	10
No. of women	12 972	2546	1658	1207	995	1010	1125	1069	1102	1060	838	359
False-positive assessment												
BI-RADS 0	14.33 (13.7, 14.9)	1.6 (1.1, 2.0)	6.4 (5.2, 7.6)	9.9 (8.2, 11.4)	14.3 (12.1, 16.4)	15.9 (13.7, 18.2)	18.9 (16.6, 21.2)	23.8 (21.3, 26.4)	22.6 (20.1, 25.1)	25.8 (23.2, 28.5)	23.9 (21.0, 26.7)	27.6 (22.9, 32.2)
BI-RADS 3	5.35 (5.0, 5.7)	1.0 (0.6, 1.4)	2.8 (2.0, 3.6)	4.8 (3.6, 6.0)	5.0 (3.7, 6.4)	8.1 (6.4, 9.8)	8.9 (7.2, 10.5)	9.6 (7.9, 11.4)	8.4 (6.8, 10.1)	7.7 (6.1, 9.3)	5.4 (3.8, 6.9)	2.5 (0.9, 4.1)
BI-RADS 4	3.93 (3.6, 4.3)	1.0 (0.6, 1.4)	2.5 (1.8, 3.3)	3.6 (2.5, 4.6)	3.8 (2.6, 5.0)	3.6 (2.4, 4.7)	5.8 (4.4, 7.1)	7.1 (5.6, 8.6)	6.1 (4.7, 7.5)	5.2 (3.8, 6.5)	5.2 (3.7, 6.8)	4.7 (2.5, 6.9)
BI-RADS 5	0.03 (0, 0.06)	0 (0, 0)	0.1 (0, 0.3)	0 (0, 0)	0.1 (0, 0.3)	0 (0, 0)	0 (0, 0)	0.1 (0, 0.3)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
BI-RADS 0, 3, 4, or 5	17.14 (16.5, 17.8)	2.7 (2.0, 3.3)	8.9 (7.7, 10.0)	12.8 (11.4, 14.1)	17.2 (15.6, 18.7)	19.2 (17.6, 20.8)	23.3 (21.5, 25.0)	27.7 (25.8, 29.5)	27.0 (25.2, 28.9)	29.1 (27.3, 31.0)	26.0 (24.2, 27.8)	29.2 (27.4, 31.1)
BI-RADS 3, 4, or 5	8.41 (7.9, 8.9)	1.9 (1.4, 2.5)	4.6 (3.6, 5.7)	7.5 (6.0, 8.9)	7.9 (6.3, 9.6)	10.6 (8.7, 12.5)	13.0 (11.0, 14.9)	14.7 (12.6, 16.8)	13.1 (11.1, 15.1)	12.5 (10.5, 14.5)	9.8 (7.8, 11.8)	7.2 (4.6, 9.9)
Biopsies that did not reveal cancer	8.04 (7.6, 8.5)	3.5 (2.8, 4.2)	5.7 (4.6, 6.8)	6.8 (5.4, 8.2)	8.1 (6.4, 9.8)	9.9 (8.1, 11.7)	9.5 (7.8, 11.2)	11.3 (9.4, 13.2)	12.2 (10.3, 14.2)	10.7 (8.9, 12.6)	10.3 (8.2, 12.3)	9.2 (6.2, 12.2)

Note.—Unless otherwise indicated, data are the percentage of women, with 95% confidence intervals in parentheses.

Table 2

History of False-Positive Events in 8-year Group for Women Who Underwent Screening Beginning in 1993

Parameter	Population of Women Who Underwent Screening in 1993	No. of Mammographic Screening Procedures over 8-year Period							
		1	2	3	4	5	6	7	8
No. of women	13 877	2421	1642	1586	1790	1898	1907	1802	892
False-positive assessment									
BI-RADS 0	16.01 (15.4, 16.6)	7.2 (6.2, 8.2)	11.6 (10.1, 13.2)	13.7 (12.0, 15.4)	16.0 (14.3, 17.7)	20.7 (18.9, 22.5)	21.6 (19.8, 23.4)	19.9 (18.0, 21.7)	22.7 (19.8, 25.6)
BI-RADS 3	6.23 (5.8, 6.6)	4.5 (3.6, 5.3)	5.7 (4.5, 6.8)	6.2 (5.0, 7.4)	7.9 (6.6, 9.1)	8.2 (6.9, 9.4)	8.3 (7.1, 9.6)	4.6 (3.6, 5.6)	3.3 (2.0, 4.8)
BI-RADS 4	4.49 (4.1, 4.8)	2.2 (1.6, 2.8)	3.5 (2.6, 4.4)	4.8 (3.8, 5.9)	5.1 (4.1, 6.2)	6.4 (5.3, 7.5)	4.9 (4.0, 5.9)	4.7 (3.7, 5.7)	5.1 (3.6, 6.6)
BI-RADS 5	0.05 (0.01, 0.09)	0 (0, 0)	0.1 (0, 0.3)	0.1 (−0.1, 0.2)	0.1 (0, 0.2)	0.1 (0, 0.2)	0 (0, 0)	0.1 (0, 0.2)	0 (0, 0)
BI-RADS 0, 3, 4, or 5	19.18 (18.5, 19.8)	10.2 (9.0, 11.4)	15.1 (13.4, 16.8)	16.8 (15.0, 18.6)	20.2 (18.3, 22.1)	24.7 (22.8, 26.6)	24.7 (22.8, 26.6)	22.3 (20.4, 24.2)	23.9 (21.0, 26.8)
BI-RADS 3, 4, or 5	9.82 (9.3, 10.3)	6.4 (5.4, 7.4)	8.3 (7.0, 9.6)	9.8 (8.3, 11.3)	11.2 (9.7, 12.7)	13.1 (11.6, 14.6)	12.5 (11.0, 14.0)	8.9 (7.6, 10.2)	8.2 (6.3, 10.1)
Biopsies that did not reveal cancer	6.95 (6.5, 7.4)	3.3 (2.6, 4.0)	5.6 (4.5, 6.7)	6.7 (5.5, 8.0)	7.5 (6.3, 8.8)	9.9 (8.6, 11.2)	8.2 (6.9, 9.4)	7.7 (6.5, 8.9)	8.2 (6.3, 10.1)

Note.—Unless otherwise indicated, data are the percentage of women, with 95% confidence intervals in parentheses.

Table 3

History of False-Positive Events in 5-year Group for Women Who Underwent Screening Beginning in 1996

Parameter	Population of Women Who Underwent Screening in 1996	No. of Mammographic Screening Procedures over 5-year Period				
		1	2	3	4	5
No. of women	16 853	3218	3056	3777	4446	2352
False-positive assessment						
BI-RADS 0	11.39 (10.9, 11.9)	6.2 (5.3, 7.0)	10.8 (9.7, 11.9)	12.1 (11.0, 13.1)	13.0 (12.0, 14.0)	15.1 (13.6, 16.5)
BI-RADS 3	4.53 (4.2, 4.8)	4.1 (3.4, 4.8)	6.5 (5.6, 7.3)	5.7 (4.9, 6.4)	3.7 (3.2, 4.3)	2.2 (1.6, 2.8)
BI-RADS 4	3.23 (3.0, 3.5)	2.3 (1.8, 2.8)	4.0 (3.3, 4.7)	3.9 (3.3, 4.5)	3.1 (2.6, 3.6)	2.5 (1.9, 3.2)
BI-RADS 5	0.01 (0, 0.03)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
BI-RADS 0, 3, 4, or 5	14.29 (13.8, 14.8)	9.6 (8.6, 10.6)	15.2 (14.0, 16.5)	15.3 (14.2, 16.4)	15.1 (14.0, 16.2)	16.4 (14.9, 17.9)
BI-RADS 3, 4, or 5	7.24 (6.8, 7.6)	5.9 (5.1, 6.7)	9.8 (8.7, 10.9)	8.9 (8.0, 9.8)	6.4 (5.7, 7.1)	4.6 (3.8, 5.4)
Biopsies that did not reveal cancer	3.77 (3.5, 4.1)	2.9 (2.3, 3.5)	3.7 (3.1, 4.4)	4.6 (4.0, 5.3)	4.1 (3.1, 4.4)	2.9 (2.2, 3.6)

Note.—Unless otherwise indicated, data are the percentage of women, with 95% confidence intervals in parentheses.

results did not reveal cancer per 1000 mammograms underwent only one mammographic procedure within this 10-year period ($P < .001$).

The magnitude of the reduction in the rate of biopsies that did not reveal cancer per mammogram that was associated with regular screening was strong enough to result in a lower long-term rate of these events in women who underwent prompt annual screening compared with those who underwent intermittent screening. In the 5-year group, 2.9% of women who underwent five mammographic examinations over the next 5 years had biopsy results that did not reveal cancer, whereas 4.6% of women who underwent three mammographic examinations over the next 5 years had biopsy results that did not reveal cancer ($P < .001$, Table 3).

Women are at risk for undergoing biopsies that do not reveal cancer whether screening is performed or not because nonmammographic signs, such as benign palpable masses, may also lead to biopsy. This was seen in the finding that, of the 2546 women who had a negative mammographic result in 1991 and never returned for mammography during the 10-year period, 3.5% had biopsy results that did not reveal cancer (approximately 0.35% per year) (Table 1). In contrast, 9.2% of the women who underwent screening each year during the 10-year period had biopsy results that did not reveal cancer ($P < .001$, Table 1). Thus, it is reasonable to conclude that the actual 10-year mammography-associated burden for biopsies that do not reveal cancer is approximately 6% among women who undergo regular screening.

Figure 1

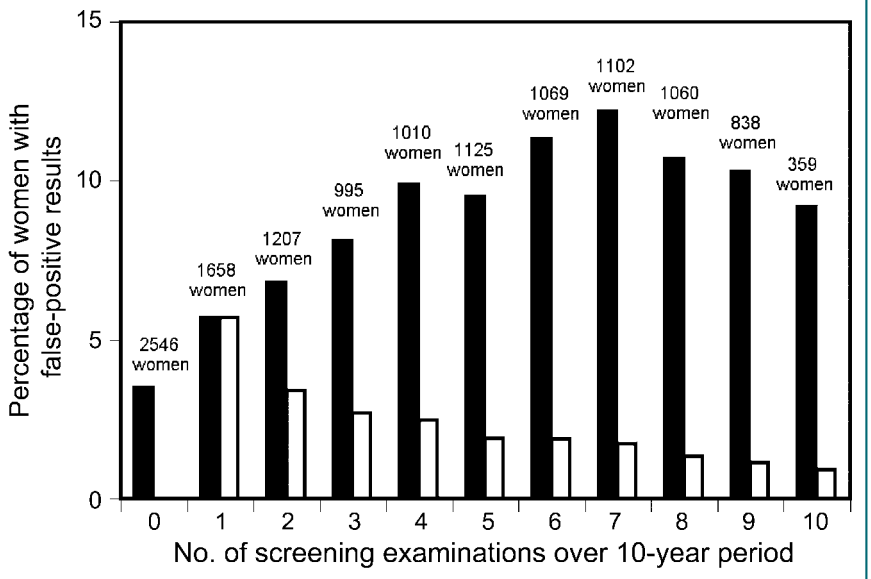

Figure 1: Ten-year rates for biopsies that did not reveal cancer (black bars) compared with those for biopsies that did reveal cancer (white bars) in women with negative screening results in 1991 (see Table 1 for values). The number of women who underwent biopsies that did not reveal cancer declined from 57 women per 1000 mammograms for those who underwent one mammographic procedure to nine women per 1000 mammograms for those who underwent 10 mammographic procedures. While a small reduction might be expected due to simple random sampling without replacement, which would be expected to have reduced the number from 57 women per 1000 mammograms to 44 women per 1000 mammograms, the actual reduction to nine women far exceeds this value.

Among the women who underwent regular screening, the risk of undergoing biopsies that do not reveal cancer declined markedly over time (Fig 2). Thus, for the 829 women who underwent screening mammography in 1993

(approximately one-third of whom underwent their first mammographic procedure in the index year of 1993, which perhaps reflects the rapid expansion of mammography use during this period) and who used all eight opportunities for screening within the 8-year period, there was a 3% risk of undergoing biopsies that did not reveal cancer in the first visit. This risk declined over subsequent visits until eventually reaching a more than fivefold reduction in the risk of negative events (0.25%) in the 7th and 8th years (Fig 2).

Figure 2

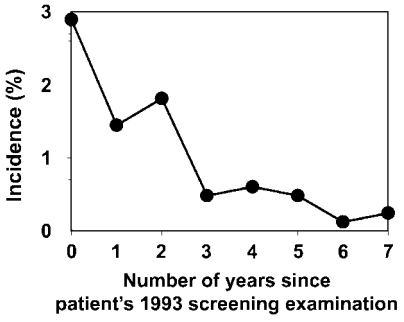

Figure 2: Graph demonstrates a reduction over time in the occurrence of biopsies that did not reveal cancer in women who underwent regular screening. The incidence of biopsies that did not reveal cancer is shown as a function of the amount of time since the patients' screening mammogram in 1993 for women who promptly returned each year for the next 7 years.

False-Positive Assessments

Most of the false-positive assessments proved to be BI-RADS category 0 findings, with a smaller number of false-positive assessments for BI-RADS category 3 and BI-RADS category 4 findings (Tables 1–3, Fig 3). False-positive assessments for BI-RADS category 5 findings were exceedingly rare. For example, of the 12 972 women at the Massachusetts General Hospital who had a negative screening result in 1991, 1859 (14.33%) had a false-positive BI-RADS category 0 assessment, 694 (5.35%) had a false-positive BI-RADS category 3 assessment, 510 (3.93%) had a false-positive BI-RADS category 4 assessment, and only four (0.03%) had a false-positive BI-RADS category 5 assessment during the next 10 years.

Women who underwent screening yearly had a lower risk of false-positive assessments per mammogram than did those who underwent screening intermittently (Tables 1–3, Fig 3). For BI-RADS category 3 and 4 assessments, the overall risk of false-positive findings per mammogram decreased with increasing screening use to a magnitude that resulted in an overall long-term decrease in the risk of false-positive findings for women who underwent screening each year (Tables 1–3, Fig 3). The per mammogram risk of false-positive BI-RADS category 0 findings also decreased with increasing screening use (Tables 1–3, Fig 3), although not to a degree that lead to an absolute decrease in long-term risk. Thus, the 10-year incidence per woman of false-positive mammographic assessments for BI-RADS category 3, 4, or 5 findings peaked at 14.7% for women who underwent six screening examinations within the 10-year period. This number declined to 7.2% ($P < .001$) for women who underwent all 10 of the annual mammographic examinations that were available within the 10-year period (Table 1, Fig 3). The 10-year incidence per woman of false-positive mammographic assessments for BI-RADS category 0, 3, 4, or 5 findings was 29.2% for women who underwent all 10 of the annual mammographic examinations that were available within the 10-year period (Table 1, Fig 3).

Discussion

The data presented here reveal that much of the burden of false-positive events can be ascribed to the failure of

Figure 3

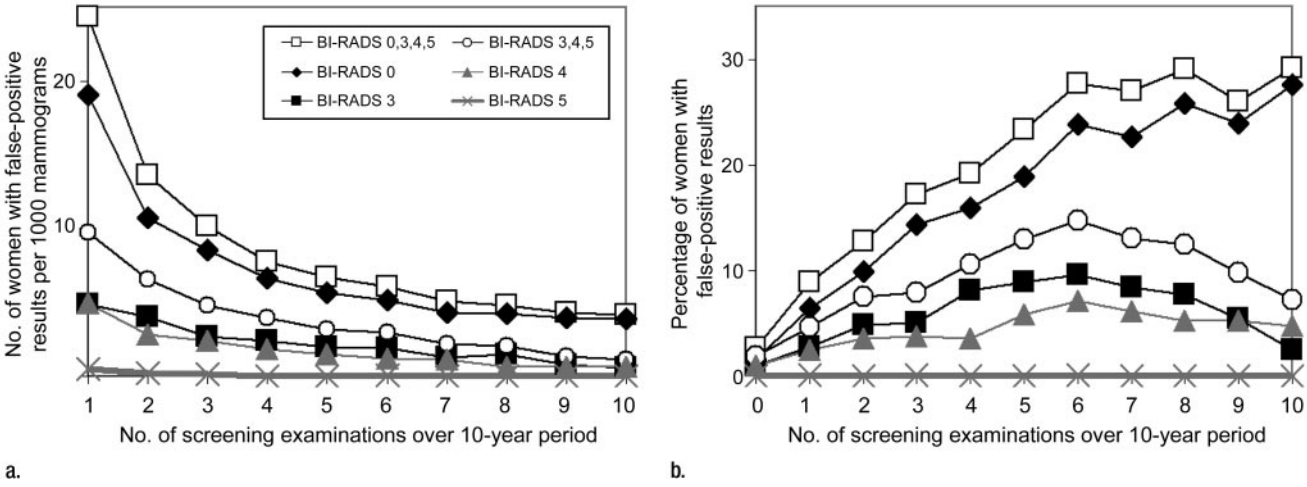

Figure 3: **(a, b)** Ten-year rates for false-positive mammographic assessments in women who received a negative screening result in 1991 (see Table 1 for values). Note that, in **a**, the decline in the occurrence of false-positive assessments with increased screening use (as was the case for biopsies that did not reveal cancer) is greater than would be expected for simple random sampling without replacement.

women to return for annual mammography. In the 5-year group, there was a 50% increase in the number biopsies that did not reveal cancer in women who made use of only three opportunities for mammography compared with those who made use of all five opportunities for mammography. Clearly, finding ways to encourage regular attendance at screening may provide a straightforward way to reduce the occurrence of false-positive events.

The large size (more than 300 000 visit records) and long time scale (1985 onward) of the Massachusetts General Hospital Avon Comprehensive Breast Center database made it possible for us to measure the actual long-term occurrence of false-positive events. In previous studies, which relied on smaller datasets, researchers had to estimate the cumulative occurrence of such events (4–7). Fortunately, our data reveal the reassuring finding that the occurrence of false-positive events among women who undergo regular screening is lower than previously suspected. For example, while Elmore et al (4) reported that "the estimated cumulative risk of a false-positive result was 49.1% after 10 mammograms," we have been able to measure the actual occurrence of these false-positive assessments among women who underwent 10 consecutive mammographic procedures over a 10-year period. Thus, our results reveal a substantially lower value (29.2%) than that obtained by Elmore et al (49.1%).

Similarly, while Elmore et al "estimate that among women who do not have breast cancer, 18.6% will undergo a biopsy after 10 mammograms" (4), the actual number of biopsies that did not reveal cancer in women who underwent 10 consecutive mammographic procedures in 10 years in the Massachusetts General Hospital population was much smaller (9.2%). Furthermore, because 3.5% of the women in the Massachusetts General Hospital population who did not return for mammography within the 10-year period had biopsy results that did not reveal cancer, the burden of negative biopsy results that was associated with regular screening within a 10-year period would appear to be about 6%. Thus, the actual risk of mammography-generated biopsies that do not reveal cancer in women who attend regular screening appears to be less than one-third of the value estimated by Elmore et al and is well within the limits that women find acceptable (23).

The data presented here also reveal that much of the risk of false-positive events occurs early in a woman's screening experience, if she attends screening regularly. Indeed, within a few years of regular screening, the risk of undergoing biopsies that do not reveal cancer decreases to levels that may be as low as (and possibly even slightly lower than) the levels of such events among women who do not undergo screening. Thus, these data suggest that special attention should be focused on trying to find ways to reduce the occurrence of false-positive events in the first few years of screening. These data also suggest that the possibility of an unacceptable cumulative burden of false-positive findings that might occur over a woman's screening lifetime, which can span 30 years or more, should be avoidable by encouraging prompt attendance at screening.

The major limitation of this study is that it examines the experience of a single screening population. Furthermore, being a large urban tertiary care facility, the Massachusetts General Hospital Avon Comprehensive Breast Center may not be representative of mammographic screening in the United States as a whole. However, the overall number of biopsies that did not reveal cancer at the Massachusetts General Hospital Avon Comprehensive Breast Center (8.04% of the women in the screening population) proved to be remarkably similar to that found in other populations (4,7); in the study by Elmore et al (4), the overall number of biopsies that did not reveal cancer was 7.8% for a group of 2400 women in a Health Maintenance Organization population. Nonetheless, we think that it would be valuable to determine whether the risk of false-positive events is also influenced by the regularity of screening use in other populations of patients.

Failure to undergo prompt screening would appear to result not only in a higher level of false-positive events, as indicated by the data outlined here, but also in a higher level of breast cancer deaths, as indicated by several other lines of evidence (24). Computer-simulated modeling of breast cancer growth, spread, and detection (25,26), which is based on quantitative estimates of the relative size at which invasive cancer is detected during screening versus in the of absence screening (27), the rate of tumor growth (28) and the relationship between tumor size and patient survival (29), suggests that each year of delay degrades the life-sparing potential of screening by about one-third (21,25,26). Similar results have been found by using a Markov model of screening (30).

The benefits of prompt annual screening are also suggested by the fact that, while the appearance of larger palpable tumors is reduced in the year after a negative mammographic result, such tumors begin to accumulate at a regular rate from about 1 year onward (31). These findings are also in agreement with measurements of breast cancer growth rate and sojourn time (28). Thus, the data reported here suggest that the use of interventions to encourage a prompt return to screening should have a substantial effect, not only in terms of a reduction in the number of deaths due to breast cancer but also in terms of a reduction in false-positive events.

References

1. Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2002;137:347–360.

2. Kopans DB. Updated results of the trials of screening mammography. Surg Oncol Clin N Am 1997;6:233–263.

3. Smith RA. Screening fundamentals. J Natl Cancer Inst Monogr 1997;22:15–19.

4. 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.

5. Christiansen CL, Wang F, Barton MB, et al.

Predicting the cumulative risk of false-positive mammograms. J Natl Cancer Inst 2000; 92:1657–1666.

6. Fletcher SW, Elmore JG. Mammographic screening for breast cancer. N Engl J Med 2003;348:1672–1680.
7. Smith-Bindman R, Chu PW, Miglioretti DL, et al. Comparison of screening mammography in the United States and the United Kingdom. JAMA 2003;290:2129–2137.
8. Brown ML, Hiun F, Sickles EA, Kessler LG. Screening mammography in community practice. AJR Am J Roentgenol 1995;165: 1373–1377.
9. Lidbrink E, Elfving J, Jonsson E. Neglected aspects of false positive findings of mammography in breast cancer screening: analysis of false positive cases from the Stockholm trial. BMJ 1996;312:273–275.
10. Barton MB, Moore S, Polk S, Shtatland E, Elmore JG, Fletcher SW. Increased patient concern after false-positive mammograms: clinician documentation and subsequent ambulatory visits. J Gen Intern Med 2001;16: 150–156.
11. Lampic C, Thurfjell E, Bergh J, Sjoden PO. Short- and long-term anxiety and depression in women recalled after breast cancer screening. Eur J Cancer 2001;37:463–469.
12. Gilbert FJ, Cordiner CM, Zeck RA, Hood DB, Mathieson D, Walker LG. Breast screening: the psychological sequelae of false-positive recall in women with and without a family history of breast cancer. Eur J Cancer 1998;34:2010–2014.
13. Callaway MP, Boggis CR, Astley SA, Hutt I. The influence of previous films on screening mammographic interpretation and detection of breast carcinoma. Clin Radiol 1997;52: 527–529.
14. Sickles EA. Successful methods to reduce false-positive mammography interpretations. Radiol Clin North Am 2000;38:693–700.
15. Burnside ES, Sickles EA, Sohlich RE, Dee KE. Differential value of comparison with previous examinations versus screening mammography. AJR Am J Roentgenol 2002; 179:1173–1177.
16. Esserman L, Cowley H, Eberle C, et al. Improving the accuracy of mammography: volume and outcome relationships. J Natl Cancer Inst 2002;94:369–375.
17. Elmore JG, Miglioretti DL, Reisch LM, et al. Screening mammograms by community radiologists: variability in false positive rates. J Natl Cancer Inst 2002;94:1373–1380.
18. Ulcickas Yood M, McCarthy BD, Lee NC, Jacobsen G, Johnson CC. Patterns and characteristics of repeat mammography among women 50 years and older. Cancer Epidemiol Biomarkers Prev 1999;8:595–599.
19. Sabogal F, Merrill SS, Packel L. Mammography rescreening among older California women. Health Care Financ Rev 2001;22: 63–75.
20. Michaelson J, Satija S, Moore R, et al. The pattern of breast cancer screening utilization and its consequences. Cancer 2002;94:37–43.
21. Blanchard K, Colbert JA, Puri D, et al. Mammographic screening: patterns of use and estimated impact on breast carcinoma survival. Cancer 2004;101(3):495–507.
22. Michaelson JS, Satija S, Moore R, Weber G, Garland G, Kopans DB. Properties of the invasive breast cancers and features of screening utilization seen over the last decade at the Massachusetts General Hospital Breast Imaging Division. J Womens Imaging 2001;3:99–104.
23. Schwartz LM, Woloshin S, Fowler FJ, Welch HG. Enthusiasm for cancer screening in the United States. JAMA 2004;291:71–78.
24. Michaelson JS, Kopans DB, Cady B. The breast cancer screening interval is important. Cancer 2000;88:1282–1284.
25. Michaelson JS, Halpern E, Kopans D. A computer simulation method for estimating the optimal intervals for breast cancer screening. Radiology 1999;212:551–560.
26. Michaelson JS. Using information on breast cancer growth, spread, and detectability to find the best ways to use screening to reduce breast cancer death. J Womens Imaging 2001;3:54–57.
27. Michaelson JS, Satija S, Moore R, et al. Estimates of the sizes at which breast cancers become detectable on mammographic and on clinical grounds. J Womens Imaging 2003;5:3–10.
28. Michaelson JS, Satija S, Moore R, Weber G, Garland G, Kopans DB. Estimates of the breast cancer growth rate and sojourn time from screening database information. J Womens Imaging 2003;5:11–19.
29. Michaelson JS, Silverstein M, Wyatt J, et al. The prediction of breast cancer survival from tumor size. Cancer 2002;95:713–723.
30. Tabar L, Duffy SW, Vitak B, Chen HH, Prevost TC. The natural history of breast carcinoma: what have we learned from screening? Cancer 1999;86:449–462.
31. Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. J Natl Cancer Inst 2000;92:1081–1087.