

AD _____

Award Number: DAMD17-99-1-9112

TITLE: Outcomes of Screening Mammography in Elderly Women

PRINCIPAL INVESTIGATOR: Doctor Rebecca Smith-Bindman

CONTRACTING ORGANIZATION: University of California, San Francisco
San Francisco, California 94143-0962

REPORT DATE: October 2004

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20050630 025

REPORT DOCUMENTATION PAGEForm Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE October 2004	3. REPORT TYPE AND DATES COVERED Annual (1 Oct 03-30 Sep 04)	
4. TITLE AND SUBTITLE Outcomes of Screening Mammography in Elderly Women			5. FUNDING NUMBERS DAMD17-99-1-9112	
6. AUTHOR(S) Doctor Rebecca Smith-Bindman				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of California, San Francisco San Francisco, California 94143-0962 E-Mail: Rebecca.Smith-Bindman@Radiology.ucsf.edu			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 Words) There is uncertainty about whether women older than age 65 should undergo screening mammography. Although screening mammography may benefit some elderly women through the detection of early breast cancers, it may harm other women through false positive diagnoses and the detection of clinically insignificant lesions. This research study involves the design and implementation of a data analysis of HCFA Medicare billing claims linked with national tumor registry data from the Surveillance Epidemiology and End Results (SEER) program. The specific aims of this research will evaluate 1) differences in breast cancer mortality, 2) differences in breast cancer treatment and 3) differences in breast cancer tumor attributes between women who were screened and those who were not. The project involves defining whether Medicare billing claims data are accurate for assessment of mammography utilization and completion of the outlined aims once these data were shown to be reliable.				
14. SUBJECT TERMS Breast Cancer			15. NUMBER OF PAGES 44	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

Table of Contents

Cover.....	1
SF 298.....	2
Table of Contents.....	3
Introduction.....	4
Body.....	4
Key Research Accomplishments.....	7
Reportable Outcomes.....	8
Conclusions.....	8
References.....	
Appendices.....	9

INTRODUCTION

There is uncertainty about whether women older than age 65 should undergo screening mammography. Although screening mammography may benefit some elderly women through the detection of early breast cancers, it may potentially harm other women through false positive diagnoses and the detection and surgical treatment of clinically insignificant lesions. This research study involves the design and implementation of a data analysis of HCFA Medicare billing claims linked with National tumor registry data from the Surveillance Epidemiology and End Results (SEER) program. The specific aims of this research validates that Medicare billing claims can be used to assess screening mammography use, evaluates the use of screening mammography in elderly women, and the outcomes associated with the use of screening mammography in elderly women.

The numbering below refers to the Revised Statement of Work.

STUDIES and RESULTS

SOW #1: Obtain Health Care Financing Administration/SEER Tumor Registry Data

The linked Medicare HCFA/SEER database describing Medicare claims through 1998 and breast cancer cases through 1996 was obtained, and data cleaning was completed in Years 1 and 2.

SOW #2: Detailed study Design and project development for Specific Aim #1

a) Develop Algorithm that will be used for determining the predictor variable of screening mammography utilization (in women with breast cancer)

Task completed in Year 2.

b) Develop mammography registry abstraction algorithm.

Task completed in Year 2.

SOW #3 Validating Algorithm for Determining Screening History

a) Analyze HCFA claims.

Task completed in Year 2.

b) Choose women on whom the algorithm will be validated and obtain mammography registry on these women.

Task completed in Year 2.

c) Perform Statistical Analysis.

The purpose of this analysis is to determine whether the SEER-Medicare data can be used to determine the use of screening mammography. In summary, Medicare administrative data are reliable for assessment of mammography utilization, and have become more accurate over time. Population trends in the use of mammography can be assessed using these data.

Completed Year 3-5.

d) Further refine the criteria for defining a screening mammogram.

Task completed in Year 3-5.

- e) Manuscript preparation describing the method of using Medicare data to determine whether or not a woman underwent screening mammography.

Two manuscripts were written and submitted for publication, but were rejected. A new manuscript combining the two has been written and submitted to *Medical Care*. This found that Medicare claims capture most mammography (>90%), that screening and diagnostic mammography can be distinguished using these data, and found that most women are correctly classified as not screened, screened, and regularly screened (at least 2 mammograms/spaced by 9-36 months). "Medicare Billing Claims Data Be Used To Assess Mammography Utilization Among Women Age 65 and Older" submitted to *Medical Care*.

SOW #4: Evaluate breast cancer treatments by mammographic screening

- a) Perform literature reviews on variables that are associated with breast cancer.
Task completed in Year 3.

- b) Perform statistical analyses to determine differences by screening.

Task completed in Year 4-5. The utilization of screening mammography is lower than suggested by self report, there are substantial differences by age, race/ethnicity in the use of screening mammography, and there are substantial differences in breast cancer treatments by age and race/ethnicity, particularly the percentage of women who undergo recommended radiation treatments as per professional guidelines.

- c) Manuscript preparation.

1. "Does the Utilization of Screening Mammography Explain Racial and Ethnic Differences in Breast Cancer?" Under review, *Annals of Internal Medicine*.
2. "Mammography Rates by Race and Ethnicity Using Medicare Data," under review, *Annals of Internal Medicine*.
3. "Racial and Ethnic Differences in the Treatment of Early-Stage Breast Cancer," will be submitted to *Cancer* within 1 month.

SOW: #5: Evaluate breast cancer tumor attributes by mammographic screening

- a) Perform statistical analyses to determine differences in tumor attributes by screening history.
Cancers detected through screening are, as expected, smaller and of lower stage than cancers detected clinically. The age and race differences in cancer stage at diagnosis are largely, but not entirely, explained by differences in the use of mammography screening. Additional analyses are underway and nearly completed adjusting for co-morbidities in this analysis.
- b) Manuscript in preparation, describing tumor attributes by screening history.
"Breast Cancer Characteristics At Detection as They Vary by the Utilization of Screening Mammography."

SOW #6: Evaluate outcomes of screening, adjusted by co-morbidities

- a) Perform literature review on co-morbidities and breast cancer and total mortality.
Task completed in Year 4.

- b) Development adaptation of Charlson/Deyo Index to use as a means to adjust for co-morbidities using inpatient and outpatient diagnoses.

We have used two methods to account for patient illness, including a modified Charlson Index (that is based on diagnoses noted at the time of all hospital admissions), during the seven years of the study, and a method that indicates illnesses noted in both inpatient and outpatient health claim files. Additionally we have developed a method to account for visits to health care providers as a way to account both for access to health care providers as well as indications of health. Comorbidities have been adjusted for using the latter method.

- c) Evaluate outcomes of breast cancer by race and ethnicity, use of screening mammography, and breast cancer treatment, and co-morbidities.

Currently completing survival analysis, and stratified multivariate analysis by screening group. We are approximately 6 months behind, from the outline in our revised Statement of Work. We expect the analysis to be completed and submitted for publication by March 2005.

Rebecca Smith-Bindman, MD

SOW #7: Evaluate breast cancer and total mortality by screening history

- a) Obtain National Death Registry Tapes and perform data linkages.

Task completed.

- b) Statistical analysis of breast cancer survival by mammographic screening, adjusting for co-morbidities and breast cancer treatment differences, and manuscript preparation. We are completing the analysis and expect analysis and manuscript to be completed and submitted for publication by March 2005.

Additional Work

Two additional and related papers were completed and the support of the DOD was acknowledged.

- "US-UK Comparison of Screening Mammography." Smith-Bindman R, Chu PW, Miglioretti DL, Sickles EA, Blanks R, Ballard-Barbash R, Bobo JK, Lee NC, Wallis MG, Patnick J, Kerlikowske K. *JAMA* 2003 Oct 22;290(16):2129-37. (Original article included in Appendices.)
- "Physician Predictors of Mammographic Accuracy." Smith-Bindman R, Chu P, Miglioretti D, Quale C, Rosenberg R, Cutter G, Geller B, Bacchetti P, Sickles E, Kerlikowske K. [In press, *JNCI*.] (Original abstract included in Appendices.)

SIGNIFICANCE

Medicare physicians' claims can be used to determine whether women have undergone screening mammography and thus Medicare data are a very reliable method that can be used to study screening mammography in elderly women.

KEY RESEARCH ACCOMPLISHMENTS

- Determined Medicare claims can be used to determine the use of mammography.
"Medicare Billing Claims Data Be Used To Assess Mammography Utilization Among Women Age 65 and Older" submitted to *Medical Care*.
- Found that the degree to which older women undergo regular mammography screening is much lower than suggested by self-reported surveys and that mammography should increase among elderly women.
"Screening Mammography Rates by Race and Ethnicity Using Medicare Data" under review, *Annals of Internal Medicine*.
- We found age, race/ethnic, and geographic differences in the use of mammography and breast cancer treatments among elderly women.
"Racial and Ethnic Differences in the Treatment of Early-Stage Breast Cancer" will be submitted to *Cancer* shortly.
- We found that most of the racial and ethnic differences in breast cancer can be explained by mammograms.
"Does Utilization of Screening Mammography Explain Racial and Ethnic Differences in Breast Cancer?" under review, *Annals of Internal Medicine*.
- We found substantial variation among physicians in the accuracy of mammograms.
(Original abstract included in Appendices.)
"Physician Predictors of Mammographic Accuracy" (Smith-Bindman R, Chu P, Miglioretti D, Quale C, Rosenberg R, Cutter G, Geller B, Bacchetti P, Sickles E, Kerlikowske K), accepted by the *Journal of the National Cancer Institute*, *In Press*.
- The U.S. is not doing as well in mammography, in comparison to the U.K. and this is an area that future research needs to address. (Original article included in Appendices.)
"U.S. – U.K. Comparison of Screening Mammography" (Smith-Bindman R, Chu PW, Miglioretti DL, Sickles EA, Blanks R, Ballard-Barbash R, Bobo JK, Lee NC, Wallis MG, Patnick J, Kerlikowske K. *JAMA* 2003 Oct 22;290(16):2129-37).

REPORTABLE OUTCOMES

None

CONCLUSIONS

The last year of the project has been successful and we achieved major goals outlined in the Statement of Work. Manuscript preparation will continue through the next several months.

APPENDIX

Comparison of Screening Mammography in the United States and the United Kingdom

Rebecca Smith-Bindman, MD ^{1,2}

Philip W. Chu, MS ¹

Diana L. Miglioretti, PhD ³

Edward A. Sickles, MD ¹

Roger Blanks, PhD ⁴

Rachel Ballard-Barbash, MD, MPH ⁵

Janet K. Bobo, PhD ⁶

Nancy C. Lee, MD ⁶

Matthew G. Wallis, MB, ChB, FRCR ⁷

Julietta Patnick, BA, FFFHM ⁸

Karla Kerlikowske, MD ^{2,9,10}

¹ Department of Radiology, University of California, San Francisco

² Department of Epidemiology and Biostatistics, University of California, San Francisco

³ Center for Health Studies, Group Health Cooperative and Dept. of Biostatistics, University of Washington, Seattle, WA

⁴ Cancer Screening Evaluation Unit, Institute of Cancer Research, University of London, U.K.

⁵ Applied Research Program, National Cancer Institute, NIH, Bethesda, MD

⁶ Division of Cancer Prevention and Control, Centers for Disease Control and Prevention

⁷ Warwickshire, Solihull, and Coventry Breast Screening Service, U.K.

⁸ NHS Cancer Screening Programs, U.K.

⁹ Department of Medicine, University of California, San Francisco

¹⁰ General Internal Medicine Section, Department of Veterans Affairs, University of California, San Francisco

Dr. Bobo is currently at the Battelle Centers for Public Health Research and Evaluation, Seattle, WA

Corresponding Author: Rebecca Smith-Bindman, MD. Department of Radiology, University of California San Francisco, UCSF / Mt. Zion Campus, 1600 Divisadero Street, San Francisco, CA 94115, Box 1667. Phone (415) 885 - 7464, Fax (415) 885 - 7876.

Email: Rebecca.Smith-Bindman@Radiology.UCSF.edu

This work was supported in part by the NCI (CA86032 [Rebecca Smith-Bindman], Breast Cancer Surveillance Consortium co-operative agreement U01CA63740 [Karla Kerlikowske], and an NCI-funded Statistical Coordinating Center grant, U01CA86076 [Diana Miglioretti]) and The Department of Defense DAMD179919112 (Rebecca Smith-Bindman, Philip Chu).

Key words: Screening Mammography, International Comparison, Accuracy

Word count: 4269

Reprint requests should be addressed to the corresponding author.

ABSTRACT

Context: The provision of screening mammography differs between the United States (U.S.) and United Kingdom (U.K.), and a direct comparison may suggest methods to improve the practice of screening mammography.

Objective: To compare screening mammography performance between the U.S. and U.K. among similar aged women.

Design, Setting, and Participants: We studied women aged 50 and older who underwent 5.5 million mammograms from 1996 to 1999 within three large-scale mammography registries or screening programs: in the U. S., the Breast Cancer Surveillance Consortium (BCSC [978,591]) and National Breast and Cervical Cancer Early Detection Program (NBCCEDP [613,388]), and in the U.K., the National Health Service Breast Cancer Screening Program (NHSBSP ([3.94 million])). A total of 27,612 women were diagnosed with breast cancer (invasive or ductal carcinoma *in situ*) within 12 months of screening among the three groups.

Main Outcome Measures: Recall rates (the recommendation for further evaluation including diagnostic mammography, ultrasound, clinical examination, or biopsy) and cancer detection rates (invasive or ductal carcinoma *in situ*) were calculated for first and subsequent examinations, and within 5-year age groups. Cancer was identified through active case follow up or by linkage to a pathology database or tumor registry.

Results: Recall rates were approximately twice as high in the U.S. as in the U.K. for all age groups, yet cancer rates were similar. For example, among women aged 50–54 who underwent a first screening mammogram, 14.6 percent of those in the BCSC and 12.5 percent of those in the NBCCEDP were

recalled for further evaluation versus only 7.6 percent in the U.K.; cancer detection rates per 1,000 screens were 5.8, 5.9 and 6.3, in the BSCS, NBCCEDP and U.K. respectively. Recall rates were lower for subsequent examinations in all three settings, but remained twice as high in the U.S. as in the U.K. A similar percentage of women underwent biopsy in each setting, but rates of percutaneous biopsy were lower, and rates of open surgical biopsy higher in the U.S. Open surgical biopsies that did not result in a diagnosis of cancer (negative biopsies) were twice as high in the U.S. as in the U.K. Based on a 10-year time period of screening, the estimated recall rates for women aged 50 to 59 years were 49.4% and 43.3% in the U.S versus 17.4% in the U.K., and for women aged 60 to 69 years 41.4% and 33.4% in the U.S versus 13.3% in the U.K. The estimated cancer detection rates (per 1000 women) for women aged 50 to 59 years were 24.5 and 23.8 in the U.S. versus 19.4 in the U.K., and for women aged 60 to 69 years 31.5 and 26.6 in the U.S. versus 27.9 in the U.K.

Conclusion: Recall and negative open surgical biopsy rates are twice as high in the U.S. settings as in the U.K., but cancer detection rates are similar. Efforts to improve mammographic screening in the U.S. should target lowering the recall rate without reducing the cancer detection rate.

INTRODUCTION

The provision of screening mammography differs greatly between the United States (U.S.) and United Kingdom (U.K.). In the U.S. screening is provided in diverse settings, such as private practice, health maintenance organizations, and academic medical centers,¹ whereas in the U.K. a single organized screening program run by the National Health Service (NHS) provides virtually all mammographic screening for women aged 50 or older.^{2,3} There are differences also between the ages of women screened, the recommended interval between mammographic examinations, the proportion of women recalled for additional imaging examinations (such as diagnostic mammography or ultrasound), and the methods used to further evaluate examinations considered suspicious for cancer.⁴⁻⁶ However, it is not clear if there are actual differences in the performance and outcomes of screening mammography between the two countries. Comparing the performance of screening mammography between the two countries may suggest methods to improve the practice of screening mammography.

The goal of this study is to compare recall (the percentage of mammograms where there is a recommendation for prompt additional testing, clinical evaluation, or percutaneous biopsy), surgical biopsy, and cancer detection rates for screening mammography, among similarly aged women, between the U.S. and the U.K.

METHODS

Data Sources

Data on screening mammography in the U.S. was obtained from the Breast Cancer Surveillance Consortium (BCSC)⁷ and the National Breast and Cervical Cancer Early Detection Program (NBCCEDP).^{8,9} In the U.K. data came from the National Health Service Breast Screening Program (NHSBSP).^{3,10} Results of all screening mammograms in women aged 50 or older performed within

each of these settings between 1996 and 1999 were included. More than one screening examination was included if the examinations occurred more than nine months apart. We excluded mammograms obtained to further evaluate a mass detected by clinical breast examination, because of a breast symptom, or to follow up an abnormal mammogram. The study was approved by the University of California, San Francisco Institutional Review Board.

Breast Cancer Surveillance Consortium (BCSC)

The BCSC is a National Cancer Institute (NCI)-funded consortium of mammography registries in San Francisco, California; Colorado; New Hampshire; New Mexico; North Carolina; Seattle, Washington; and Vermont.⁷ The primary purpose of the consortium is to collect data pertaining to mammography performance in a uniform fashion across diverse settings and populations.¹¹ Women are included if they self refer or are referred by a physician for a mammogram, to one of the 202 contributing facilities. Data are obtained for individual women from self-administered questionnaires¹² and radiologist reports (medical records), and mammography results are reported using the categories of the American College of Radiology's Breast Imaging Reporting and Data Systems (BI-RADS¹³). Cases of cancer are ascertained through active case follow up and through linkages with state tumor registries, Surveillance Epidemiology and End Result (SEER) programs, or pathology databases and cancer ascertainment has been found to be 94% complete.¹⁴ While all facilities and radiologists in the U.S. must follow the Mammography Quality Standards Act/Mammography Quality Standards Reauthorization Act (MQSA/QMSRA) regulations,¹⁵ the BCSC offers no specific guidelines for, nor has authority in advising, how mammograms should be interpreted.

National Breast and Cervical Cancer Early Detection Program (NBCCEDP)

The NBCCEDP, which is funded by the Centers for Disease Control and Prevention (CDC), provides breast and cervical cancer screening to poor, uninsured women throughout the U.S.^{8,9} Screening mammography for women aged 40 and older have been provided in all 50 states, tribes and territories since 1996, although funding limitations have allowed only 15-20 percent of eligible women to be served. The CDC funds each state, which in general contract for the mammographic screening through diverse settings. Data are collected for individual women from self-administered questionnaires and reports (medical records) from primary providers and radiologists. Mammography results are reported to the programs using the categories of BI-RADS.¹³ Cancer occurrences are ascertained primarily through active follow-up of abnormal mammograms and review of pathology reports, but some programs also link to state tumor registries or SEER programs. The NBCCEDP offers no specific guidelines on how mammograms should be interpreted, but it works with all of the state programs to improve program performance including mammography.

National Health Service Breast Screening Program (NHSBSP)

The government-funded National Health Service Breast Screening Program provides free breast cancer screening in the U.K. for women 50 or older.^{3,10} Women aged 50-64 years are invited by postcard to attend breast screening every three years through a system that relies on centralized computer databases; from age 65 onwards, women are encouraged to self-refer. By 1995, the NHSBSP achieved national coverage so that screening mammography was available to all eligible women. The program is currently organized into 95 separate breast-screening programs that coordinate the provision of screening services and cancer ascertainment. Data are collected and analyzed locally as well as centrally in the Department of Health and the Cancer Screening Evaluation Unit, University of London. Women specifically concerned about breast problems are referred to hospital breast clinics for diagnostic mammography, and the results of such testing are not included in this report.

Positive Mammogram

For the BCSC and NBCCEDP, a mammogram was classified as positive (recall) if the assessment was abnormal or incomplete (BI-RADS¹³ categories 0,3,4,5) and there was a recommendation for prompt diagnostic imaging, clinical evaluation, or biopsy (including fine needle aspiration [FNA], core biopsy, and open surgical biopsy). Clinical evaluation was the reason for a positive examination in a small minority of cases (< 2% of the recalls) but was included to be consistent with the NHSBSP. For the NHSBSP, a mammogram was classified as positive (recall) if there was a recommendation for further work-up including diagnostic imaging, clinical examination, or pathologic evaluation. Any additional views that were recommended contributed to the recall rate. Mammograms that were recommended for short interval follow-up only were not considered positive.

First vs. Subsequent Mammogram

Because recall and cancer detection rates vary by whether women have undergone previous mammography,^{9, 16, 17} all analyses were sub-divided by whether women were undergoing a first or subsequent screening examination (screening cycle). For the BCSC and NBCCEDP data, a mammogram was classified as "first" if the woman had no prior mammogram in the database and self-reported no prior mammogram within five years. For the NHSBSP, the first mammogram that a woman underwent in the program was considered "first." Information on race, ethnicity, socioeconomic status, and cancer risk factors (such as menopausal status and use of hormone therapy) are not collected by the NHSBSP or NBCCEDP, and were not included.

Cancer Detected

Women were considered to have breast cancer detected if active case follow-up or reports from a pathology database, SEER program, or state tumor registry showed invasive carcinoma or ductal

carcinoma *in situ* (DCIS) within 12 months of a positive screening mammogram. Cancers that occurred after a negative mammogram (false negative examinations) were not included in this analysis.

Statistical Analysis

Recall, non-invasive work-up, and biopsy rates were calculated per 100 screening mammograms and stratified by first or subsequent examinations and by five-year age groups, or age adjusted to a standard age distribution. The standard age distribution was the mean of the age distributions of the three data sources, where each data source was weighted equally. The recall rate was calculated as the number of positive mammograms per 100 screening mammograms. The non-invasive work-up rate was calculated as the number of recommendations for only non-invasive work-up, including ultrasound, diagnostic mammography, other non-invasive tests, or breast examination per 100 mammograms. Each mammogram was counted one time when calculating the non-invasive work-up rate, even if more than one test was recommended. The biopsy rate (any type of biopsy) was calculated as the number of mammograms with a recommendation for fine needle aspiration (FNA), core biopsy, biopsy where the type was not specified, or open surgical biopsy, per 100 mammograms. Each mammogram was counted one time when calculating the overall biopsy rate, even if more than one biopsy was recommended. The percutaneous biopsy rate was calculated as the number of FNAs or core biopsies per 100 mammograms. The open surgical biopsy rate was calculated as the number of open surgical biopsies per 100 mammograms. Women could have contributed to both the percutaneous biopsy rate and the open surgical biopsy rate, and thus these numbers do not sum to the overall biopsy rate. The open surgical biopsy rate was subdivided into two groups; those that resulted in a diagnosis of cancer (the positive open surgical biopsy rate), and those that did not result in a diagnosis of cancer (the negative open surgical biopsy rate). The specific method of biopsy could not be determined for the NBCCEDP data and for three of the seven BCSC sites and thus the percutaneous and open surgical biopsy rates could not

be calculated for these sites. The cancer detection rate was calculated as the number of breast cancers detected per 1,000 examinations. The rate of invasive cancer by tumor size (<10mm, 10-20mm, >20 mm) was calculated per 1000 examinations using the standard age distribution.

Because mammographic screening is recommended^{3, 18-21} and performed^{3, 22} more frequently in the U.S. than the U.K., one would expect fewer cancers to be diagnosed per subsequent screening examination in the U.S. To compare cancer detection rates for a similar time period of screening, we used four years of actual data to estimate the numbers of cancers detected and women recalled per 1000 women undergoing screening mammography over a 10-year period. For these estimates, we assumed that screening started at age 50 (or age 60) and continued for 10 years using an estimated screening interval for each setting. For the BCSC and NBCCEDP, the estimated screening interval was 18 and 19 months, respectively, and was based on the mean time between mammograms that women obtained between 1998 and 1999. These estimates are similar to those reported by others.^{22, 23} For the NHSBSP, screening occurred about every three years²⁴ and correspondingly, the interval was estimated at 36 months. To calculate 10-year estimates of cancer detection and recall for each program, a 50 year-old woman was assumed to have undergone a single "first" mammogram and several "subsequent" examinations. The age-specific recall rate and cancer rate of these first and subsequent examinations were those reported in Tables 2, 3 and 4. We assumed that women aged 60 or older underwent only "subsequent" examinations so only age-specific recall and cancer rates for subsequent screens in Tables 2, 3 and 4 were used to calculate 10-year estimates. We also assumed that the likelihood of recall and cancer detection was independent from one exam to the next, and that a woman could be recalled or have cancer detected only once. Thus we estimated the chance of at least one recalled examination or cancer diagnosis over a 10-year period for a 50 year-old woman who underwent routine screening and a 60 year-old woman who underwent routine screening in each setting. To estimate the variability of these 10-year estimates, we used the 95 percent

confidence interval for the recall rates (Tables 2, 3) and cancer detection rates (Table 4), and varied the screening interval from 16 to 20 months (BCSC), 17 to 21 months (NBCCEDP) and 33 to 39 months (NHSBSP). The lower estimate for the range in the cancer rate was calculated by assuming the lower bound of the 95 percent confidence interval for cancer detection and screening interval.

RESULTS

This analysis included 5.5 million mammograms: 978,591 from the BCSC, 613,388 from the NBCCEDP, and 3.94 million from the NHSBSP, which led to the diagnosis of 27,612 cases of breast cancer among women ages 50 and older (Table 1).

Recall rates were similar between the BCSC and the NBCCEDP for both first and subsequent examinations, Table 2. Recall rates in these two U.S. settings were approximately twice as high as those in the U.K. for all age groups, for first as well as subsequent examinations. For example, among first screening mammograms for women ages 50 – 54 years, 14.6 percent of women in the BCSC and 12.5 percent in the NBCCEDP, versus only 7.6 percent of women in the U.K., were recalled for further evaluation including diagnostic mammography, ultrasound, clinical examination, or biopsy. Biopsy rates were similar across all settings: 2.3 - 3.4 percent of first screening mammograms, and 0.84 – 1.7 percent of subsequent screening examinations were followed by a recommendation for biopsy. Thus, the higher recall rate in the U.S. was primarily due to a higher rate of diagnostic mammography, ultrasound, and clinical evaluation.

Although the biopsy rates were similar between the two countries, biopsies were more likely to be open surgical biopsies in the U.S. (Table 3). For example, for 100 first screening mammograms, 1.1% in the U.S. compared with 2.4% in the U.K. resulted in a recommendation for percutaneous biopsy, and for 100 subsequent screening mammograms, 0.4% in the U.S. compared with 0.8% in the U.K. resulted in a recommendation for percutaneous biopsy (age adjusted values). On the other hand, for 100 first screening mammograms, 1.15% in the U.S. compared with 0.72% in the U.K. resulted in a recommendation for open surgical biopsy (age adjusted values, Table 3). Most of the difference in open surgical biopsy rates was attributed to procedures among women who did not have breast cancer (Table

3), with negative open surgical biopsy rates 2–3 times as high in the U.S. compared with the U.K. For example, for 100 first screening examinations, 0.82 percent resulted in negative open surgical biopsy in the U.S., compared with 0.36 percent in the U.K. Positive surgical biopsy rates were more similar between the two countries but tended to be slightly higher in the U.K.

The cancer detection rates increased with age and were 2-3 times as high for first compared with subsequent mammograms in both countries (Table 4). Despite substantially higher recall rates in the U.S., cancer detection rates were similar across settings, particularly for first screening examinations. For example, for 1000 first examinations among women ages 50 – 54 years, 5.8, 5.9, and 6.3 cancers were diagnosed in the BCSC, NBCCEDP, and U.K., respectively (Table 4). Differences in cancer detection rates between the two countries were greater for subsequent examinations, likely reflecting more frequent screening in the U.S.

The estimated numbers of cancers detected per 1000 women screened over 10 years were also similar between both countries (Table 5). If 1000 women ages 50-59 underwent regular mammographic screening over 10 years, approximately 24.5 cancers would be detected in the BCSC, 23.8 in the NBCCEDP, and 19.4 in the U.K. If 1000 women ages 60-69 underwent regular mammographic screening over 10 years, approximately 31.5 cancers would be detected in the BCSC, 26.6 in the NBCCEDP, and 27.9 in the U.K. While invasive cancer detection rates are more similar between the two countries, the *in situ* cancer rates are higher in the U.S. Among women ages 50-59, approximately 5.8, 7.4, and 3.8 *in situ* cancers would be detected in the BCSC, NBCCEDP, and U.K. respectively. The higher frequency of screening in the U.S. magnifies the difference in the estimated recall rates between the countries when projected over 10 years. For example, after 10 years of screening 1000 women aged 50 to 59 years, 494 women in the BCSC and 433 women in the NBCCEDP, compared to

175 women in the U.K., would have been recalled for additional work-up. After 10 years of screening women aged 60 to 69 years, 414 women in the BCSC and 334 women in the NBCCEDP, compared to 133 women in the U.K., would have been recalled for additional work-up.

For first screening mammograms, there were slightly fewer invasive cancers diagnosed per 1000 examinations in the U.S. in most size categories (Figure 1A). For subsequent examinations, there were lower rates of invasive cancer in all size categories in the U.S. compared to the U.K. (Figure 1B). The absolute difference in cancer rates between the U.S. and U.K. were highest for invasive tumors 10 to 20 mm. (Figures 1A, 1B).

DISCUSSION

The recall and negative open surgical biopsy rates associated with screening mammograms were twice as high in U.S. settings than they were in the U.K., and yet cancer detection rates were similar in the two countries. This means that in the U.K. half as many women are recalled for diagnostic tests and half as many women without breast cancer undergo open surgical biopsies as in the U.S. These results observed in large numbers of women are similar to recent findings from a series of 60 test cases evaluated by both U.S. and U.K. physicians in whom false positive rates were higher among U.S. physicians, but cancer detection rates were not.²⁵ These are important findings because the goal of any cancer screening effort is to obtain high cancer detection rates while avoiding unnecessary diagnostic evaluation following false positive results, which are costly and associated with ongoing psychological morbidity.²⁶

There are several possible explanations for the striking differences in recall rates between the U.S. and U.K. Much higher rates of malpractice lawsuits in the U.S. that focus on missed breast cancer diagnoses

²⁷ provide a strong incentive to increase sensitivity at the expense of specificity, possibly leading U.S. radiologists to recall women when they identify a finding with a low likelihood of cancer. ^{25, 28} In addition, U.S. physicians must read only 480 mammograms annually to fulfill MQSA requirements ¹⁵ while U.K. radiologists are required to read at least 5000 mammograms annually, ²⁹ and on average, U.K. radiologists interpret 5-7 times their U.S. counterparts. Thus very experienced radiologists read mammograms in the U.K. Furthermore, although over 90 percent of programs in the U.K. employ double reading, this practice is much less common in the U.S. While the exact impact of double reading remains uncertain, some evidence shows that double reading by consensus or arbitration, as used in the U.K., raises detection rates and decreases recall rates. ^{30, 31} Also to be considered is the greater centralization of mammogram reading in the U.K., as well as a less mobile population, which might make prior mammograms more readily available for comparison when interpreting mammograms. Still, while the latter difference might reduce recall rates for subsequent mammograms, ^{32, 33} it does not account for higher recall rates for first screening mammograms.

Perhaps most importantly, quality assurance standards for the NHSBSP programs are set nationally and are regularly monitored through a quality assurance network. Ranges of acceptable values for recall, biopsy, and cancer detection rates have been established, and an organized program operates at the local and national levels to monitor and achieve these target values. ^{5, 29} All U.K. screening programs receive data that enables them to compare their recall and cancer detection rates with other programs. Both programs and individual radiologists below a minimum standard are subject to quality assurance scrutiny. In contrast, the U.S. has only voluntary guidelines ³⁴ and there is no national organization to collect or monitor data to promote high levels of performance. Finally, an organized program of professional development in the U.K. specifically provides instruction and individual feedback regarding recall and cancer detection rates, using a set of test mammography cases called PERFORMS. ³⁵

• Although not compulsory, 85 percent of U.K. mammographers participated in this test in 2001.

Continuing medical education is a requirement for U.S. radiologists, but the actual content is not uniformly organized, often does not specifically address screening interpretation, and almost never targets specific recall or cancer detection rates.

The NHSBSP has integrated quality assurance into all clinical aspects of its program,^{5, 29, 36} and as a result, the U.K. has seen dramatic improvements in the performance of screening mammography since the program began in 1988. For example, cancer detection rates have increased dramatically for both first and subsequent screening examinations, as has the positive predictive value of mammography.³⁶ The U.K. saw a rapid 50 percent decline in the open surgical biopsy rate between 1996 and 1999 as a result of a coordinated effort to increase the use of percutaneous biopsy, and to decrease the percentage of women without breast cancer who underwent open surgical biopsy.^{3, 37} The well-documented improvements in the U.K.⁵ demonstrate that implementation of quality control can be efficient and feedback mechanisms effective. Despite the differences between the two countries in the provision and funding of screening mammography, mammography technology is very similar between the two countries, and one would think that similar targets for mammography outcomes, including specific recall and biopsy rates, could be established in the U.S. Success in reaching technical targets set by the MQSA/MQSRA demonstrates how a coordinated quality assurance program can work in the U.S.³⁸

Screening mammography is performed more frequently in the U.S. than the U.K. Thus over a 10-year period, women 50 and older will undergo approximately seven mammograms in the U.S versus only three in the U.K. More frequent screening likely translates into smaller average cancer size at diagnosis, as evidenced by the slightly lower rates of invasive cancer 10 mm or greater, and the higher rates of *in situ*, cancer diagnosed in the U.S. Additionally, screening in the U.S. tends to begin at an earlier age

than in the U.K. From the results reported here, it cannot be determined whether these differences affect breast cancer mortality.

We compared the cancer detection rates between the U.S. and U.K. as a measure of mammography performance. Cancer detection rates are widely used as a measure of mammography performance^{36, 39, 40} as they approximate the total cancer rates and can be readily measured for quality assurance purposes. We found the breast cancer detection rates in the U.S. and the U.K. are similar. Given the overall age-adjusted breast cancer incidence rates are slightly higher in the U.S.⁴¹, one would expect that the U.S. would have similar or higher cancer detection rates than in the U.K. This suggests it is unlikely that the U.K. is missing cancers despite a much lower recall rate than the U.S. This is not unexpected. It has been shown that at high recall rates, cancer detection rates levels off.⁴² Thus despite recalling more women, more cancers are not detected in the U.S.

The main limitation of our study is that we cannot be certain that our definition of "screening mammography" was the same across all three settings. Specifically, we cannot be sure there were not a higher proportion of diagnostic examinations among the U.S. women, which might account for a higher recall rate. However, more diagnostic mammograms should produce a substantially higher cancer rate^{43, 44} in the two U.S. settings, which we did not find. We should also note that our estimation of the total cancers detected over 10 years was based on only four years of screening data, and the assumptions of the model were simplistic. Still, when we used different values for these assumptions our results did not appreciably change. Additionally, our estimated recall rates are similar to those results found by others.^{32, 45} There is likely a small degree of overlap between the two U.S. data sources, but this is estimated to be less than 3 percent of the mammograms described in this report. Additionally, by pooling data within each program, we have ignored variations by region, physician and other variables in each program.^{36, 46}

• Lastly, although the data from the U.K. includes virtually all mammographic screening performed in that country, the data from the U.S. reflects only a small percentage of mammography done in this country. Because mammograms from all 50 U.S. states were included, and the results from the BCSC and NBCCEDP were broadly similar, these results probably provide the best current evidence of the performance of mammography screening in the U.S.

We did not focus on differences between the BCSC and NBCCEDP (such as the slightly higher diagnostic imaging rate and slightly lower biopsy rate in the BCSC) because the differences between the two U.S. data sources were small compared with the differences between the U.S. and U.K., and these programs describe different populations, where breast cancer rates, as well as tumor characteristics, might be different.

Given the results reported here, what should U.S. women do? Women undergoing screening mammography should consider going to facilities where physicians read a large number of mammography examinations,^{28,40} where the radiologists devote a large percentage of their practice to mammography,⁴⁰ and where comprehensive auditing of outcomes is undertaken on a routine basis.¹³ Additionally, women should return to the same facility for repeat screening or ensure that comparison films are available to radiologists at the time of imaging interpretation if they change facilities.^{32,33} Lastly, if they do have an abnormal examination and an open surgical biopsy is recommended, they should discuss all options with a radiologist or surgeon, and consider getting a second opinion if they want to avoid a surgical biopsy.

CONCLUSION

•

In the U.K., the NHSBSP has set and reached targets that emphasize high rates of cancer detection and low recall. Recall rates in the U.K. are now substantially lower than in the U.S. with no substantial reduction in cancer detection. We believe this success stems primarily from a centralized program of continuous quality improvement. A large portion of the costs associated with mammographic screening comes from frequent screening ⁴⁷ and the relatively high percentage of women who undergo additional testing. ⁴⁸ Screening women aged 50-69 biennially and reducing recall rates could substantially reduce the cost of mammography, as well as associated anxiety caused by false-positive diagnoses. ⁴⁹ Efforts to improve mammographic screening in the U.S. should be targeted to lowering the recall rate without substantially lowering the cancer detection rate.

ACKNOWLEDGEMENT

We thank Travis Seawards for his invaluable assistance in preparing the tables and manuscript, Kim Lovy for her analysis of the U.K. data, Sandy Thames for her help obtaining and interpreting the data from the CDC, and Dr. Deborah Grady, Dr. Virginia Ernster and Dr. William Barlow for their helpful comments on an earlier version of the manuscript.

TABLES

Table 1: Mammography registries and programs included in this report, and number of mammograms obtained between 1996-1999 among women age 50 and older.

*After 1997, most women aged ≥ 65 years were ineligible for the NBCCEDP, as Medicare began to cover all costs for screening mammography. Thus, most of the mammograms for the age group in NBCCEDP were performed in 1996-1997.

Table 2: Recommendations for further assessment per 100 screening mammograms by age, setting and screening cycle.

Footnotes Table 2

*Values in parenthesis are 95 percent confidence intervals.

†Recall includes any recommendation for further workup, including non-invasive imaging (ultrasound, diagnostic mammography, other tests), breast examination, or pathologic evaluation (fine needle aspiration, core biopsy, surgical biopsy, or biopsy type not specified). Each mammogram contributed once to the recall rate even if multiple tests were recommended.

‡ Non-invasive workup (ultrasound, diagnostic mammography, other tests, or breast exam), but not a recommendation for pathological evaluation.

§Biopsy including any recommendation for pathologic evaluation, including fine needle aspiration, core biopsy, surgical biopsy, or biopsy type not specified. Each mammogram contributed once to the biopsy rate even if multiple biopsies were recommended.

Table 3: Recommended open surgical biopsy rates per 100 screening mammograms, by age, setting and screening cycle.

Footnotes Table 3

*For the NHSBSP and for four of the BCSC sites (Colorado, North Carolina, Seattle, Vermont) we were able to differentiate the type of biopsy (percutaneous, including fine needle aspiration or core biopsy, versus open surgical biopsy). Values in parenthesis are 95% confidence intervals. The positive and negative open surgical biopsy rate may not sum due to rounding.

†Open surgical biopsies that yielded a diagnosis of cancer per 100 mammograms.

‡Open surgical biopsies that did not yield a diagnosis of cancer per 100 mammograms.

Table 4: Cancers detected per 1000 screening mammograms, by age, setting and screening cycle.

Footnote Table 4

*Values in parenthesis are 95 percent confidence intervals.

Table 5. Estimated number of women with at least one recalled examination, cancer diagnosis, or biopsy over ten years.

Footnote Table 5

*The time between mammograms was assumed to be the mean interval observed in each setting.

Figures in parenthesis are estimated intervals of variation around these estimates and were calculated by taking the 95 percent confidence interval for the cancer detection, recall, and biopsy rates and varying the screening interval from 16 to 20 months (BCSC), 17 to 21 months (NBCCEDP) and 33 to 39 months (NHSBSP)

† Open surgical biopsies are a subset of all biopsies

††The type of biopsy could not be determined from the NBCCEDP data and for three of the BCSC sites.

• **Figure 1:** Rate per 1,000 screening examinations of DCIS and Invasive Breast Cancer by Size (<10 mm, 10-20 mm, >20 mm) for First (1A) and Subsequent (1B) Screening Mammograms, by setting. Results are age adjusted.

Footnote Figure 1

Detailed information on the size of the invasive breast cancers is only available in the U.K. from 1998 - 1999, and thus the comparison of tumor size was limited to 1998-1999. In the U.K. tumors that measured 10 mm were included with those that measured 11-20 mm, and we used this same grouping of 10-20 mm for all three data sources as these are the data available. As this cut-point is different than typically used in the U.S., the size distributions reported here might be slightly different than reported by others in the U.S.

REFERENCES

1. Houn F, Brown M. Current practice of screening mammography in the United States: data from the National Survey of Mammography Facilities. *Radiology* 1994; 190:209-215.
2. Blanks R, Moss S, Patnick J. Results from the United Kingdom NHS breast screening programme 1994-1999. *J Med Screen* 2000; 7:195-198.
3. U.K. National Health Service Breast Screening Web site.
www.cancerscreening.nhs.uk/breastscreen.
4. Brown M, Houn F, Sickles E, Kessler L. Screening mammography in community practice: positive predictive value of abnormal findings and yield of follow-up diagnostic procedures. *AJR* 1995; 165:1373-1377.
5. Blanks R, Bennett R, Wallis M, Moss S. Does Individual programme size affect screening performance? Results from the United Kingdom NHS breast screening programme. *J Med Screen* 2002; 9:11-14.
6. Shapiro S, Coleman E, Broeders M, et al. Breast cancer screening programmes in 22 countries: current policies, administration and guidelines. International Breast Cancer Screening Network (IBSN) and the European Network of Pilot Projects for Breast Cancer Screening. *Int J Epidemiology* 1998; 27:735-742.
7. Ballard-Barbash R, Taplin SH, Yankaskas BC, al. e. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. *AJR* 1997; 169:1001-1008.
8. Henson R, Wyatt S, Lee N. The National Breast and Cervical Cancer Early Detection Program: a comprehensive public health response to two major health issues for women. *J Public Health Manag Pract* 1996; 2 Spring:36-47.

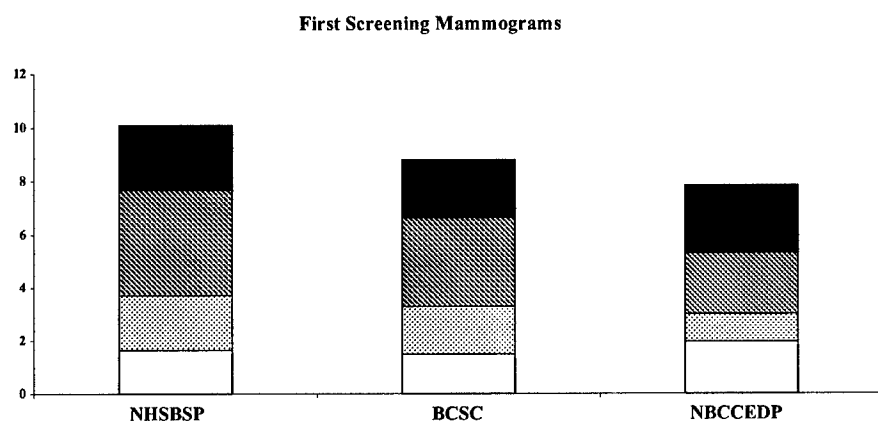
9. May D, Lee N, Nadel M, al. e. The National Breast and Cervical Cancer Early Detection Program: report on the first 4 years of mammography provided to medically underserved women. *AJR* 1998; 170:97-104.
10. Working Group chaired by Professor Sir Patrick Forrest. Breast cancer screening: report to Health Ministers of England, Wales, Scotland and Northern Ireland. London: HMSO, 1986.
11. Carney P, Miglioretti D, Yankaskas B, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the performance of screening mammography. *Annals Int Med* 2003; 138:168-173.
12. Breast Cancer Surveillance Consortium 2002 mammography questionnaire web site.
<http://www.breastscreening.cancer.gov/elements.html#questionnaires>
13. American College of R. Breast imaging reporting and data system (BI-RADS), 3rd ed. Reston, VA: American College of Radiology, 1998.
14. Ernster VL, Ballard-Barbash R, Barlow WE, al e. Detection of DCIS in women undergoing screening mammography. *J Natl Cancer Inst* 2002; 94:1151-1159.
15. Public Law 102-539, The Mammography Quality Standards Act of 1992.
16. Kerlikowske K, Grady D, Barclay J, al. e. Likelihood ratios for modern screening mammography. Risk of breast cancer based on age and mammographic interpretation. *JAMA* 1996; 276:39-43.
17. Sickles E, Ominsky S, Sollitto R, al. e. Medical audit of a rapid-throughput mammography screening practice: methodology and results of 27,114 examinations. *Radiology* 1990; 175:323-327.
18. Leitch AM DG, Costanza M, et al. American Cancer Society guidelines for the early detection of breast cancer. Update 1997. *CA Cancer J Clinicians* 1997; 47:150-153.
19. Feig S, D'orsi C, Hendrick R, ea. American College of Radiology guidelines for breast cancer screening. *AJR* 1998; 171:29-33.

20. U.S. Preventive Services Task Force. Screening for breast cancer: recommendations and rationale. see <http://www.ahrq.gov/clinic/3rduspstf/breastcancer/> 2002.
21. U.S. Preventive Services Task Force. The Guide to Clinical Preventive Services. Washington, DC: Office of Disease Prevention and Health Promotion, 1996.
22. Yood M, McCarthy B, Lee N, Jacobsen G, Johnson C. Patterns and characteristics of repeat mammography among women 50 years and older. *Cancer Epidemiol, Biomarkers Prev* 1999; 8:595-599.
23. 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; 270:2444-50.
24. Department of Health, U.K., Statistical Bulletin, Breast Screening Programme, England:1999-2000: National Statistics, March 2000.
25. Esserman L, Cowley H, Eberle C, et al. Improving the accuracy of mammography: volume and outcome relationships. *J Natl Cancer Inst* 2002; 94.
26. Moss M. Spotting breast cancer: doctors are weak link. *New York Times*. NY, 2002.
27. Physician Insurers Association of America. Breast cancer study: June 1995. Washington, DC: Physician Insurers Association of America, 1995.
28. Kan L, Olivotto I, Burhenne LW, Sickles E, Coldman A. Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. *Radiology* 2000; 215:563-567.
29. NHS Breast Screening Radiologists Quality Assurance Committee. Quality Assurance Guidelines for Radiologists. Sheffield, UK: NHSBSP, Publication No 15, (Revised May 1997). 1997.
30. Kopans D. Double reading. *Radiol Clin North Am* 2000; 38:719-24.

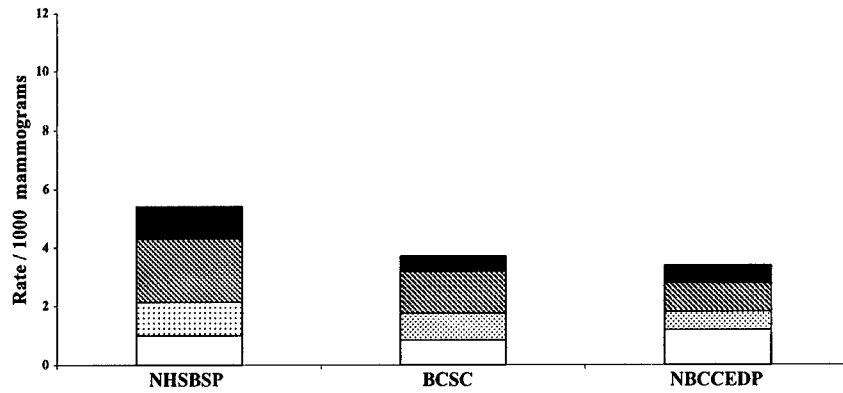
31. Williams L, Hartswood M, Prescott R. Methodological issues in mammography double reading studies. *J Med Screen* 1998; 5:202-6.
32. Christiansen C, Wang F, Barton M, et al. Predicting the cumulative risk of false-positive mammograms. *J Natl Cancer Inst* 2000; 92:1657-66.
33. Burnside E, Sickles E, Sohlich R, Dee K. The differential value of comparison with previous examinations in diagnostic versus screening mammography. In press, *AJR* 2002.
34. Bassett L, Hendrick R, Bassford T, al. e. Quality determinants of mammography, clinical practice guideline #13. Rockville, MD: Agency for Health Care Policy and Research (AHCPR publication no. 95-0632). 1994.
35. Gale A, Cowley H. Breast cancer screening: comparison of radiologists' performance in a self-assessment scheme and in actual breast screening. In: Krupinski EA. *Medical Imaging 1999: image perception and performance*. Proc SPIE1 #999:3663:157-68.
36. Blanks R, Moss S, Wallis M. Monitoring and evaluating the UK National Health Service Breast Screening Programme: evaluating the variation in radiological performance between individual programmes using PPV-referral diagrams. *J Med Screen* 2001; 8:24-28.
37. NHS Breast Screening Programme and British Association of Surgical Oncology. An audit of screen detected breast cancers for the year of screening, April 1988-March 1999. See <http://www.cancerscreening.nhs.uk/breastscreen>.
38. Suleiman O, Spelic D, McCrohan J, Symonds G, Houn F. Mammography in the 1990's: the United States and Canada. *Radiology* 1999; 210:345-51.
39. Klabunde CN, Sancho-Gardnie H, Taplin S, et.al. Quality assurance in follow-up and initial treatment for screening mammography programs in 22 countries. *Int J Qual Health Care* 2002; 14:449-61.

40. Sickles E, Wolverton D, Dee K. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology* 2002; 224:861-9.
41. International Agency for Research on Cancer Web site. <http://www.iarc.fr/>.
42. Yankaskas BC, Cleveland RJ, Schell MJ, Kozar R. Association of recall rates with sensitivity and positive predictive values of screening mammography. *AJR Am J Roentgenol* 2001; 177:543-9.
43. Sohlich R, Sickles E, Burnside E, Dee K. Interpreting data from audits when screening and diagnostic mammography outcomes are combined. *AJR* 2002; 178:681-686.
44. Dee K, Sickles E. Medical audit of diagnostic mammography examinations: comparison with screening outcomes obtained concurrently. *AJR* 2001; 176:729-33.
45. 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; 338:1089-96.
46. Beam CA, Sullivan DC, Layde PM. Effect of human variability on independent double reading in screening mammography. *Academic Radiology* 1996; 3:891-7.
47. Salzmann 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; 127:955-65.
48. Burnside E, Belkora J, Esserman L. The impact of alternative practices on the cost and quality of mammographic screening in the United States. *Clin Breast Cancer* 2001; 2:145-52.
49. Brett J, Austoker J. Women who are recalled for further investigation for breast screening: psychological consequences 3 years after recall and factors affecting re-attendance. *J Public Health Med* 2001; 23:292-300.

Figure 1: Rate of Breast Cancer by Size (<10 mm, 10-20 mm, >20 mm) for 1000 First (1A) and Subsequent (1B) Screening Mammograms, for each program. Results are age adjusted.



Subsequent Screening Mammograms



Footnote Figure 1

Detailed information on the size of the invasive breast cancers is only available in the U.K. from 1998 - 1999, and thus the comparison of tumor size was limited to 1998-1999. In the U.K. tumors that measured 10 mm were included with those that measured 11-20 mm, and we used this same grouping of 10-20 mm for all three data sources as these are the data available. As this cut-point is different than typically used in the U.S., the size distributions reported here might be slightly different than reported by others in the U.S.

TABLE 1. Mammography registries and programs included in this report, and number of mammograms obtained between 1996-1999 among women age 50 and older.

Data Sources	Abbreviation	Years	Mammograms	Mammograms by Age Group (Yrs.)				Breast Cancers
				50-54	55-59	60-64	≥ 65+	
Breast Cancer Surveillance Consortium, National Cancer Institute	BCSC	1996-1999	978,591	264,229	198,122	160,039	356,201	4,232
National Breast and Cervical Cancer Early Detection Program, Centers for Disease Control and Prevention	NBCCEDP	1996-1999	613,388	196,407	163,026	147,815	106,140*	2,711
UK National Breast Screening Program	NHSBSP	1996-1999	3,939,329	1,581,190	1,239,908	1,056,997	61,234	20,669
Total			5,531,308	2,041,826	1,601,056	1,364,851	523,575	27,612

*After 1997, most women aged ≥ 65 years were ineligible for the NBCCEDP, as Medicare began to cover all costs for screening mammography. Thus, most of the mammograms for the age group in NBCCEDP were performed in 1996-1997.

TABLE 2. Recommendations for further assessment per 100 screening mammograms by age, setting and screening cycle.*

Age (years)	First Screening Mammogram			Subsequent Screening Mammogram		
	BCSC USA	NBCEDP USA	NHSBP UK	BCSC USA	NBCEDP USA	NHSBP UK
Recall†						
50 – 54	14.6 (14.2-15.1)	12.5 (12.3-12.8)	7.6 (7.6-7.7)	8.7 (8.6-8.8)	8.0 (7.8-8.1)	3.9 (3.8-3.9)
55 – 59	13.7 (13.1-14.3)	12.0 (11.6-12.3)	7.0 (6.8-7.2)	8.3 (8.2-8.4)	7.0 (6.9-7.2)	3.6 (3.5-3.6)
60 – 64	12.6 (12.0-13.3)	11.4 (11.1-11.8)	6.7 (6.5-6.9)	7.9 (7.7-8.0)	6.7 (6.5-6.8)	3.4 (3.4-3.5)
≥ 65	12.3 (11.9-12.8)	8.3 (7.9-8.7)	7.6 (6.9-8.4)	6.9 (6.8-7.0)	5.1 (4.9-5.2)	3.7 (3.5-3.8)
All Ages‡	13.3 (13.1-13.6)	11.2 (11.1-11.4)	7.2 (7.0-7.3)	8.0 (7.9-8.1)	6.8 (6.8-6.9)	3.6 (3.6-3.7)
Non-invasive diagnostic workup‡						
50 – 54	12.8 (12.4-13.3)	9.3 (9.1-9.6)	5.3 (5.2-5.3)	7.6 (7.5-7.8)	6.3 (6.1-6.4)	3.0 (3.0-3.0)
55 – 59	11.8 (11.2-12.4)	8.8 (8.5-9.1)	4.6 (4.5-4.8)	7.2 (7.1-7.4)	5.6 (5.5-5.7)	2.6 (2.6-2.7)
60 – 64	10.5 (9.9-11.1)	8.1 (7.8-8.4)	4.1 (4.0-4.3)	6.8 (6.7-6.9)	5.3 (5.1-5.4)	2.4 (2.4-2.5)
≥ 65	9.8 (9.4-10.2)	5.6 (5.3-6.0)	3.9 (3.4-4.4)	5.9 (5.8-6.0)	3.9 (3.8-4.0)	2.3 (2.2-2.5)
All Ages‡	11.2 (11.1-11.4)	8.1 (8.0-8.3)	4.5 (4.4-4.6)	6.9 (6.9-7.0)	5.4 (5.3-5.5)	2.6 (2.6-2.6)
Any biopsy§ (percutaneous or open surgical biopsy)						
50 – 54	2.3 (2.1-2.5)	3.2 (3.1-3.4)	2.4 (2.3-2.4)	.97 (94-1.0)	1.7 (1.6-1.8)	.84 (82-86)
55 – 59	2.3 (2.0-2.6)	3.1 (3.0-3.3)	2.3 (2.2-2.4)	.99 (94-1.0)	1.4 (1.4-1.5)	.90 (88-91)
60 – 64	2.5 (2.2-2.8)	3.4 (3.2-3.6)	2.5 (2.4-2.6)	1.0 (99-1.1)	1.4 (1.4-1.5)	.97 (95-99)
≥ 65	2.9 (2.6-3.1)	2.6 (2.4-2.9)	3.4 (2.9-3.9)	1.1 (1.0-1.1)	1.2 (1.1-1.2)	1.3 (1.2-1.4)
All Ages‡	2.4 (2.3-2.6)	3.1 (3.0-3.2)	2.5 (2.5-2.6)	.99 (97-1.02)	1.4 (1.4-1.5)	.96 (94-98)

*Values in parenthesis are 95 percent confidence intervals.

†Recall includes any recommendation for further workup, including non-invasive imaging (ultrasound, diagnostic mammography, other tests), breast examination, or pathologic evaluation (fine needle aspiration, core biopsy, surgical biopsy, or biopsy type not specified). Each mammogram contributed once to the recall rate even if multiple tests were recommended.

‡Adjusted to a standard age distribution

§Non-invasive workup (ultrasound, diagnostic mammography, other tests, or breast exam), but not a recommendation for pathological evaluation.

§Biopsy including any recommendation for pathologic evaluation, including fine needle aspiration, core biopsy, surgical biopsy, or biopsy type not specified. Each mammogram contributed once to the biopsy rate even if multiple biopsies were recommended.

TABLE 3. Recommended open surgical biopsy rates per 100 screening mammograms, by age, setting, and screening cycle. *

Age (years)	First		Subsequent	
	BCSC USA	NHSBSP UK	BCSC USA	NHSBSP UK
Open surgical biopsy				
50 - 54	1.1 (.91-1.3)	.64 (.62-.65)	.30 (.27-.33)	.26 (.25-.27)
55 - 59	1.2 (.90-1.5)	.70 (.65-.75)	.31 (.27-.35)	.28 (.27-.29)
60 - 64	.94 (.67-1.3)	.76 (.69-.83)	.36 (.32-.41)	.30 (.29-.31)
≥ 65	1.5 (1.2-1.7)	.88 (.65-1.2)	.42 (.39-.45)	.33 (.29-.38)
All Ages ^f	1.15 (1.1-1.2)	.72 (.67-.77)	.33 (.32-.35)	.28 (.27-.29)
Positive open surgical biopsy ^g				
50 - 54	.36 (.25-.49)	.25 (.24-.26)	.08 (.06-.10)	.15 (.14-.16)
55 - 59	.19 (.09-.35)	.33 (.29-.37)	.11 (.09-.14)	.18 (.17-.19)
60 - 64	.25 (.13-.45)	.43 (.38-.48)	.12 (.10-.15)	.21 (.20-.22)
≥ 65	.50 (.37-.67)	.53 (.36-.77)	.18 (.16-.20)	.22 (.18-.26)
All Ages ^f	.31 (.27-.36)	.36 (.32-.39)	.11 (.11-.12)	.18 (.17-.19)
Negative open surgical biopsy ^g				
50 - 54	.74 (.59-.93)	.39 (.38-.40)	.22 (.20-.25)	.11 (.10-.12)
55 - 59	.98 (.73-1.3)	.37 (.33-.41)	.20 (.17-.23)	.10 (.09-.11)
60 - 64	.69 (.46-.98)	.34 (.29-.38)	.24 (.21-.28)	.09 (.09-.10)
≥ 65	.95 (.77-1.2)	.36 (.21-.56)	.24 (.22-.27)	.11 (.09-.14)
All Ages ^f	.82 (.75-.89)	.36 (.33-.39)	.22 (.21-.23)	.10 (.10-.11)

*For the NHSBSP and for the four of the BCSC sites (Colorado, North Carolina, Seattle, Vermont) we were able to differentiate the type of biopsy (percutaneous, including fine needle aspiration or core biopsy, versus open surgical biopsy). Values in parenthesis are 95 percent confidence intervals. The positive and negative open surgical biopsy rate may not sum due to rounding.

^fAdjusted to a standard age distribution

^gOpen surgical biopsies that yielded a diagnosis of cancer per 100 mammograms.

*Open surgical biopsies that did not yield a diagnosis of cancer per 100 mammograms.

TABLE 4. Cancers detected per 1000 screening mammograms, by age, setting and screening cycle.

Age (years)	First			Subsequent		
	BCSC USA	NBCEDP USA	NHSBSP UK	BCSC USA	NBCEDP USA	NHSBSP UK
Total						
50-54	5.8 (4.5, 7.3)	5.9 (5.0, 6.8)	6.3 (6.1, 6.5)	2.6 (2.4, 2.9)	2.8 (2.4, 3.1)	3.8 (3.5, 4.0)
55-59	7.4 (5.4, 9.8)	8.1 (6.8, 9.3)	9.2 (8.2, 10.2)	3.6 (3.3, 4.0)	3.5 (3.1, 3.9)	4.9 (4.7, 5.1)
60-64	10.1 (7.5, 13.2)	11.9 (10.2, 13.6)	11.9 (10.6, 13.4)	3.9 (3.5, 4.4)	3.7 (3.3, 4.2)	5.9 (5.6, 6.1)
≥ 65	14.4 (12.3, 16.8)	8.8 (6.3, 11.3)	16.6 (12.2, 22.0)	5.2 (4.9, 5.6)	4.4 (3.5, 5.3)	8.7 (7.7, 9.8)
All Ages [†]	8.6 (7.9, 9.4)	8.3 (7.7, 8.7)	10.1 (9.4, 10.7)	3.6 (3.5, 3.7)	3.4 (3.3, 3.6)	5.4 (5.2, 5.5)
Invasive						
50-54	4.5 (3.4, 5.9)	4.6 (3.8, 5.4)	4.9 (4.7, 5.1)	1.9 (1.6, 2.1)	1.7 (1.4, 2.0)	3.0 (2.8, 3.2)
55-59	6.8 (4.9, 9.1)	6.0 (4.9, 7.1)	7.7 (6.8, 8.6)	2.9 (2.6, 3.3)	2.4 (2.1, 2.8)	3.9 (3.8, 4.1)
60-64	7.7 (5.5, 10.5)	8.9 (7.5, 10.4)	9.5 (8.3, 10.8)	3.0 (2.6, 3.4)	2.6 (2.2, 2.9)	4.9 (4.7, 5.1)
≥ 65	12.4 (10.4, 14.6)	7.1 (4.9, 9.4)	14.9 (10.7, 20.0)	4.2 (3.9, 4.5)	2.9 (2.2, 3.6)	6.9 (6.1, 7.9)
All Ages [†]	7.2 (6.5, 7.8)	6.3 (5.9, 6.7)	8.4 (7.8, 9.0)	2.8 (2.7, 2.9)	2.3 (2.2, 2.4)	4.3 (4.2, 4.5)
In-situ						
50-54	1.3 (0.7, 2.1)	1.3 (90, 1.7)	1.4 (1.3, 1.5)	.77 (0.6, 0.9)	1.1 (.86, 1.3)	.70 (.60, .80)
55-59	.63 (0.20, 1.6)	2.1 (1.4, 2.7)	1.4 (1.0, 1.8)	.73 (0.6, 0.9)	1.1 (.83, 1.3)	.90 (.90, 1.0)
60-64	2.4 (1.2, 4.1)	3.0 (2.1, 3.8)	2.2 (1.6, 2.9)	.96 (0.8, 1.2)	1.2 (.93, 1.5)	1.0 (.90, 1.0)
≥ 65	2.0 (1.3, 3.1)	1.7 (.6, 2.8)	1.8 (.60, 4.1)	1.0 (0.9, 1.2)	1.6 (1.0, 2.1)	1.7 (1.3, 2.2)
All Ages [†]	1.5 (1.2, 1.8)	1.9 (1.7, 2.2)	1.6 (1.4, 1.9)	.83 (.77, .90)	1.2 (1.1, 1.3)	.99 (.92, 1.1)

*Values in parenthesis are 95 percent confidence intervals.
†Adjusted to a standard age distribution

TABLE 5. Estimated number of women with at least one recalled examination, cancer diagnosis, or biopsy among 1000 women screened over ten years.*

	BCSC USA	NBCCEDP USA	NHSBSP UK
Women screened	1000	1000	1000
Average time between screening exams (months)	18	19	36
Ages 50 – 59 years			
Cancer detected	24.5 (19.9-30.7)	23.8 (19.1-28.9)	19.4 (17.5-21.4)
<i>In situ</i>	5.8 (5.0-8.3)	7.4 (5.2-9.9)	3.8 (3.3-4.5)
Invasive	19.0 (15.0-24.1)	16.4 (12.9-20.7)	15.3 (13.9-17.0)
Women recalled	494.1 (459.3-533.7)	432.8 (402.4-469.1)	174.5 (164.9-183.8)
Biopsy	79.0 (69.2-89.8)	113.2 (102.7-129.7)	49.3 (45.8-52.0)
Open surgical biopsy [†]	29.0 (23.5-35.8)	††	14.5 (13.4-15.6)
Ages 60 – 69 years			
Cancer detected	31.5 (26.3-37.8)	26.6 (20.7-34.2)	27.9 (24.1-32.3)
<i>In situ</i>	7.0 (5.3-9.2)	9.2 (5.9-12.9)	5.2 (4.0-6.7)
Invasive	24.7 (20.5-30.0)	18.0 (13.4-23.5)	22.7 (19.7-26.6)
Women recalled	413.8 (380.7-451.5)	333.7 (302.7-365.5)	132.6 (122.6-144.8)
Biopsy	71.7 (63.0-82.0)	83.9 (74.8-95.4)	43.4 (36.9-48.7)
Open surgical biopsy [†]	26.9 (22.4-32.8)	††	12.4 (10.9-14.4)

*The time between mammograms was assumed to be the mean interval observed in each setting. Figures in parenthesis are estimated intervals in variation around these estimates and were calculated by taking the 95 percent confidence interval for the cancer detection, recall, and biopsy rates and varying the screening interval from 16 to 20 months (BCSC), 17 to 21 months (NBCCEDP), and 33 to 39 months (NHSBSP).

[†]Open surgical biopsies are a subset of all biopsies.

^{††}The type of biopsy could not be determined from the NBCCEDP data and for three of the BCSC sites.

PHYSICIAN PREDICTORS OF MAMMOGRAPHIC ACCURACY

ABSTRACT

Background: The impact of physician experience on the accuracy of screening mammography in actual community practice has not been well studied.

Objective: To identify physician characteristics and practice patterns associated with the accuracy of screening mammography.

Data Sources: Data were obtained from mammography registries in Colorado, New Mexico, San Francisco, and Vermont that participate in the Breast Cancer Surveillance Consortium (BCSC), a National Cancer Institute funded consortium that prospectively collects mammographic interpretations and cancer outcomes. Information on physicians was obtained from the American Medical Association Master File.

Subjects: A total of 209 physicians who interpreted 1,220,046 screening mammograms between January 1, 1995 and December 31, 2000, of which 7,143 (5.9/1,000) were associated with breast cancer within 12 months of screening.

Results: Physicians interpreted a mean of 6,011 screening mammograms during the study period, including 34 from women diagnosed with breast cancer. Physicians varied widely in screening accuracy. The mean sensitivity of mammography was 77% (SD 12%), and this varied from 29% to 97%. The mean false positive rate was 10% (SD 5%), and this varied from 1% to 29%. In multivariable logistic regression analysis, patient characteristics (including age, breast density and whether the mammogram was a first or subsequent examination) strongly predicted sensitivity, specificity, and overall accuracy. After adjusting for each physician's mixture of patients, physician characteristics also strongly predicted specificity; a lower false positive rate was associated with at least 25 years (versus less than 10 years) since receipt of medical degree,

Comment [DM1]: Note: you sometimes use a comma and sometimes do not. Make sure you are consistent throughout.

interpretation of 2,500-4,000 (versus 481-750) screening mammograms annually, and a high clinical focus on screening as compared with diagnostic mammography. Only a high focus on screening mammography predicted slightly lower sensitivity. Mammographic accuracy was higher (reflecting a shift to an improved ROC curve) among more experienced physicians and those with a high clinical focus on screening. When compared with physicians who interpreted the minimum number of mammograms annually and had a low focus on screening, physicians who interpreted 2,500-4,000 mammograms annually and had a high focus on screening had approximately 50% fewer false positive examinations and detected a few less cancers; one less cancer for every 2,673 mammograms interpreted. For a high volume and high screening focused physician who interprets 3,000 mammograms annually, this would translate into 182 fewer false positive examinations and 1 missed cancer per year.

Conclusion: There is wide variation in sensitivity and false positive rates among physicians.

Physicians who interpret over 2,500-4,000 mammograms annually have substantially lower false positive rates than physicians who interpret fewer examinations and only miss on average about 1 additional cancer per year. Physicians who have a high focus on screening as opposed to diagnostic mammography or are 25-35 years since receipt of medical degree, have significantly improved accuracy. Women's age and breast density have a larger effect on the accuracy of mammography than physician characteristics.