AD_____

Award Number: DAMD17-99-1-9112

TITLE: Outcomes of Screening Mammography in Elderly Women

PRINCIPAL INVESTIGATOR: Rebecca Smith-Bindman, M.D.

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

REPORT DATE: October 2005

TYPE OF REPORT: Final

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.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

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 Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01-10-2005	Final	1 Oct 1999 – 30 Sep 2005

4. TITLE AND SUBTITLE

Outcomes of Screening Mammography in Elderly Women

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DAMD17-99-1-9112

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Rebecca Smith-Bindman, M.D.

Email: rebecca.smith-bindman@radiology@ucsf.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of California, San Francisco
San Francisco, CA 94143-0962

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. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
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 cancer, it may potentially harm other women through false positive diagnoses and the diction and surgical treatment of clinically insignificant lesions. Further it is not known how the use of mammography and breast cancer outcomes varies by demographic factors such as race and ethnicity. The research designed in this proposal was targeted to try to understand the balance between benefit and harm associated with mammography screening. Much of the research involved the design and implementation of data analyses of data from the Center for Medicaid and Medicare Services, data from the National Surveillance Epidemiology and End Results (SEER) program and data from the NCI funded Breast Cancer Surveillance System. Additional related projects were focused on assessing the quality of mammography and the outcomes associated with mammographic screening. Further 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, by age and race and ethnicity. The project involves defining whether Medicare billing claims data were accurate for assessment of mammography utilization and completion of the outlined aims once these data were shown to be reliable.

15. SUBJECT TERMS
Breast Cancer

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	105	19b. TELEPHONE NUMBER *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

	Page
Cover	1
SF 298	2
Table of Contents	3
Introduction	4
Body	4
Key Research Accomplishments	6
Reportable Outcomes	7
Conclusion	7
References	7
Appendices	8

I. 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. Further it is not known how the use of mammography and breast cancer outcomes varies by demographic factors such as race and ethnicity. The research designed in this proposal was targeted to try to understand the balance between benefit and harm associated with mammography screening. Much of the research involved the design and implementation of data analyses of data from the Center for Medicaid and Medicare Services, data from the National Surveillance Epidemiology and End Results (SEER) program and data from the NCI funded Breast Cancer Surveillance System. Additional related projects were focused on assessing the quality of mammography and the outcomes associated with mammographic screening.

II. BODY

The success on each task outlined in the Statement of Work (SOW) is provided below. The manuscripts are numbered in consecutive order, and these same numbers are used throughout this final progress report.

SOW #1: Obtain Health Care Financing Administration and SEER tumor registry data for the study period and perform data cleaning. Tasks completed during years 1 and 2.

SOW #2: Detailed study design and project development for Specific Aim #1, completed year 2.

SOW #3 Validate that Medicare billing claims can be used to determine mammographic screening history among elderly women. Project completed years 2-5. We found billing claims are accurate for assessment of mammography screening in comparison to data prospectively collected in a mammography registry. Please see attached manuscript for details

(#1) Smith-Bindman R, Quale C, Chu PW, et al. Can Medicare Billing Claims Data Be Used to Assess Mammography Utilization Among Women Age 65 and Older. *Medical Care* 2006 May;44(5):463-70

SOW #4: Evaluate breast cancer treatments as it varies by race, ethnicity and mammographic screening, several projects completed years 2-5. Please see attached manuscripts for details. In summary, utilization of screening mammography is lower than
suggested by self report and there are substantial differences by age, race/ethnicity in the use of screening mammography (#2); there are substantial differences in the adequacy of breast cancer treatment by age, race and ethnicity with older and minority women more likely to have inadequate care (#3) , and the use of screening mammography explains a substantial amount of the differences in breast cancer outcomes that have been seen by age, race and ethnicity (#4). The last manuscript was made available on the DOD web site.

(#2) Kagay C, Quale C, Smith-Bindman R. Mammography Use Among the American Elderly," *Am J Prev Med* 2006 Aug; 31(2):142-9.

(#3) Haggstrom DA, Quale C, Smith-Bindman R. Differences in the Quality of Breast Cancer Care Among Vulnerable Populations. *Cancer.* 2005 Dec 1; 104(11): 2347-58

(#4) Smith-Bindman R, Miglioretti D, Lurie N, Abraham L, et al. Does Utilization of Screening Mammography Explain Racial and Ethnic Differences in Breast Cancer? *Ann Intern Med,* 2006 Apr18; 144(8): 541-53.

SOW: #5: Evaluate breast cancer tumor attributes by mammographic screening. Completed year 5. We found that tumors found by screening tend to be smaller and of lower stage, and the racial and ethnic differences in tumor characteristics at diagnosis are largely the result of differences in screening.

(#4) Smith-Bindman R, Miglioretti D, Lurie N, Abraham L, et al. Does Utilization of Screening Mammography Explain Racial and Ethnic Differences in Breast Cancer? *Ann Intern Med,* 2006 Apr18; 144(8): 541-53.

SOW #6 and SOW #7: Evaluate outcomes of breast cancer (including survival and mortality) by race and ethnicity, use of screening mammography, breast cancer treatment, and co-morbidities. Measures of co-morbidity needed to be developed as part of these SOWs. Several measures of co-morbidity were developed and tested during years 4 and 5, and these results included in the manuscripts below (#2, #3, #5). Please see attached manuscripts for details. We have submitted to *Cancer* the results describing the multivariate analysis of factors that contribute to breast cancer survival, and the differences by race and ethnicity (#5). In summary, breast cancer screening, treatment, and biology all contribute to outcomes in approximately equal measure. Although breast cancer survival is substantially lower among African American women, after adjusting for these factors, the differences (among all stage breast cancer) are no longer present. Please see attached manuscripts for more details.

(#2) Kagay C, Quale C, Smith-Bindman R. Mammography Use Among the American Elderly," *Am J Prev Med* 2006 Aug; 31(2): 142-9.

(#3) Haggstrom DA, Quale C, Smith-Bindman R. Differences in the Quality of Breast Cancer Care Among Vulnerable Populations. *Cancer.* 2005 Dec 1; 104(11): 2347-58

(#5) Curtis E, Quale C, Haggstrom D, Smith-Bindman R. Racial and Ethnic Differences in Breast Cancer Survival: How Much Is Explained By Screening, Tumor Severity, Biology, Treatment, Co-morbidities and Demographics? Submitted to *Cancer*, included as attachment

Additional Related Work

The general area of research was to try to document the benefits and harms associated with screening mammography. Several related projects were completed with the DOD support that allowed Dr. Smith-Bindman to assess the quality of mammography. The support of the DOD was acknowledged in each of these publications

(#6) 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

(#7) Smith-Bindman R, Ballard-Barbash R, Miglioretti D, Patnick J, Kerlikowske K. Comparing the Performance of Mammography Screening in the United States and the United Kingdom. *J Med Screen* 2005 12(1): 50-54.

(#8) Smith-Bindman R, Chu PW, Miglioretti D, Quale C, et al. Physician Predictors of Mammographic Accuracy.
J Natl Cancer Inst 2005; 97:358-67.

III. KEY RESEARCH ACCOMPLISHMENTS

A. Determined that Medicare billing claims can be used to determine the use of mammography among elderly women. This study validates that Medicare data can be used to study breast cancer screening and associated process and outcomes of care.

(#1) Smith-Bindman R, Quale C, Chu PW, et al. Can Medicare Billing Claims Data Be Used to Assess Mammography Utilization Among Women Age 65 and Older. *Medical Care* 2006 May; 44(5): 463-70

B. Documented that mammography use is substantially lower among elderly women, and racial and ethnic minorities then widely thought. These results contrast with many widely held views. These research findings were presented at the DOD Era of Hope meeting and presented at several Institute of Medicine Meetings including

(#2) Kagay C, Quale C, Smith-Bindman R. Mammography Use Among the American Elderly," *Am J Prev Med* 2006 Aug; 31(2): 142-9.

C. Documented that there are persistent disparities in adequate breast cancer treatment. Specifically, radiation treatment is missing or incomplete in a high percent of African American women who have undergone breast conserving treatment and the adequacy of diagnostic evaluation is suboptimal among minority women

(#3) Haggstrom DA, Quale C, Smith-Bindman R. Differences in the Quality of Breast Cancer Care Among Vulnerable Populations. *Cancer*. 2005 Dec 1; 104(11): 2347-58

D. Documented that a substantial component of breast cancer disparity is due to differential use of screening mammography. These results, when published in the Annals of Internal Medicine, resulted in extensive media coverage and reconsideration of the need to continue to emphasize the need for improved access to screening mammography among all women.

(#4) Smith-Bindman R, Miglioretti D, Lurie N, Abraham L, et al. Does Utilization of Screening Mammography Explain Racial and Ethnic Differences in Breast Cancer? *Ann Intern Med*, 2006 Apr18; 144(8): 541-53.

E. Documented how different factors, such as screening, treatment and biology, contribute to breast cancer outcomes.

(#5) Curtis E, Quale C, Haggstrom D, Smith-Bindman R. Racial and Ethnic Differences in Breast Cancer Survival: How Much Is Explained By Screening, Tumor Severity, Biology, Treatment, Co-morbidities and Demographics? Submitted to *Cancer*, included as attachment

F. Documented in several papers that the quality of mammography is woefully inadequate and there is a pressing need to improve the quality of mammography screening.

(#6) 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

(#7) Smith-Bindman R, Ballard-Barbash R, Miglioretti D, Patnick J, Kerlikowske K. Comparing the Performance of Mammography Screening in the United States and the United Kingdom. *J Med Screen* 2005 12(1): 50-54.

(#8) Smith-Bindman R, Chu PW, Miglioretti D, Quale C, et al. Physician Predictors of Mammographic Accuracy. *J Natl Cancer Inst* 2005; 97:358-67.

IV. REPORTABLE OUTCOMES

Overall 7 manuscripts were published, and one manuscript has been submitted for publication acknowledging support of the DOD. Additionally, the results were presented at two Institute of Medicine meetings, and the support of the DOD was acknowledged in those presentations. There were not adverse outcomes.

(#1) Smith-Bindman R, Quale C, Chu PW, et al. Can Medicare Billing Claims Data Be Used to Assess Mammography Utilization Among Women Age 65 and Older. *Medical Care* 2006 May; 44(5): 463-70
(#2) Kagay C, Quale C, Smith-Bindman R. Mammography Use Among the American Elderly," *Am J Prev Med* 2006 Aug; 31(2): 142-9.
(#3) Haggstrom DA, Quale C, Smith-Bindman R. Differences in the Quality of Breast Cancer Care Among Vulnerable Populations. *Cancer*. 2005 Dec 1; 104(11): 2347-58
(#4) Smith-Bindman R, Miglioretti D, Lurie N, Abraham L, et al. Does Utilization of Screening Mammography Explain Racial and Ethnic Differences in Breast Cancer? *Ann Intern Med*, 2006 Apr18; 144(8): 541-53.
(#5) Curtis E, Quale C, Haggstrom D, Smith-Bindman R. Racial and Ethnic Differences in Breast Cancer Survival: How Much Is Explained By Screening, Tumor Severity, Biology, Treatment, Co-morbidities and Demographics? Submitted to *Cancer*,
(#6) 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
(#7) Smith-Bindman R, Ballard-Barbash R, Miglioretti D, Patnick J, Kerlikowske K. Comparing the Performance of Mammography Screening in the United States and the United Kingdom. *J Med Screen* 2005 12(1): 50-54.
(#8) Smith-Bindman R, Chu PW, Miglioretti D, Quale C, et al. Physician Predictors of Mammographic Accuracy. *J Natl Cancer Inst* 2005; 97:358-67.

2004 Institute of Medicine (IOM) Conference, Saving Women's Lives, Washington D.C. *Accuracy and Access to Screening Mammography*
2005 Institute of Medicine (IOM) Conference Improving Mammographic Quality Standards, *Participant and external reviewer of final published report.*

V. CONCLUSIONS

The project was successful and we achieved all major goals outlined in the original application

VI. REFERENCES

(#1) Smith-Bindman R, Quale C, Chu PW, et al. Can Medicare Billing Claims Data Be Used to Assess Mammography Utilization Among Women Age 65 and Older. *Medical Care* 2006 May; 44(5): 463-70
(#2) Kagay C, Quale C, Smith-Bindman R. Mammography Use Among the American Elderly," *Am J Prev Med* 2006 Aug; 31(2): 142-9.
(#3) Haggstrom DA, Quale C, Smith-Bindman R. Differences in the Quality of Breast Cancer Care Among Vulnerable Populations. *Cancer*. 2005 Dec 1; 104(11): 2347-58
(#4) Smith-Bindman R, Miglioretti D, Lurie N, Abraham L, et al. Does Utilization of Screening Mammography Explain Racial and Ethnic Differences in Breast Cancer? *Ann Intern Med*, 2006 Apr18; 144(8): 541-53.
(#5) Curtis E, Quale C, Haggstrom D, Smith-Bindman R. Racial and Ethnic Differences in Breast Cancer Survival: How Much Is Explained By Screening, Tumor Severity, Biology, Treatment, Co-morbidities and Demographics? Submitted to *Cancer*,
(#6) 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

(#7) Smith-Bindman R, Ballard-Barbash R, Miglioretti D, Patnick J, Kerlikowske K. Comparing the Performance of Mammography Screening in the United States and the United Kingdom. *J Med Screen* 2005 12(1): 50-54.
(#8) Smith-Bindman R, Chu PW, Miglioretti D, Quale C, et al. Physician Predictors of Mammographic Accuracy. *J Natl Cancer Inst* 2005; 97:358-67.

VII. APPENDICES

Each of the publications cited is included as an appendix

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

Rebecca Smith-Bindman, MD
Philip W. Chu, MS
Diana L. Miglioretti, PhD
Edward A. Sickles, MD
Roger Blanks, PhD
Rachel Ballard-Barbash, MD, MPh
Janet K. Bobo, PhD
Nancy C. Lee, MD
Matthew G. Wallis, MB, ChB, FRCR
Julietta Patnick, BA, FFPH
Karla Kerlikowske, MD

Context Screening mammography differs between the United States and the United Kingdom; a direct comparison may suggest methods to improve the practice.

Objective To compare screening mammography performance between the United States and the United Kingdom among similar-aged women.

Design, Setting, and Participants Women aged 50 years or older were identified who underwent 5.5 million mammograms from January 1, 1996, to December 31, 1999, within 3 large-scale mammography registries or screening programs: the Breast Cancer Surveillance Consortium (BCSC, n=978591) and National Breast and Cervical Cancer Early Detection Program (NBCCEDP, n=613388) in the United States; and the National Health Service Breast Screening Program (NHSBSP, n=3.94 million) in the United Kingdom. A total of 27612 women were diagnosed with breast cancer (invasive or ductal carcinoma in situ) within 12 months of screening among the 3 groups.

Main Outcome Measures Recall rates (recommendation for further evaluation including diagnostic imaging, ultrasound, clinical examination, or biopsy) and cancer detection rates were calculated for first and subsequent mammograms, and within 5-year age groups.

Results Recall rates were approximately twice as high in the United States than in the United Kingdom for all age groups; however, cancer rates were similar. Among women aged 50 to 54 years who underwent a first screening mammogram, 14.4% in the BCSC and 12.5% in the NBCCEDP were recalled for further evaluation vs only 7.6% in the NHSBSP. Cancer detection rates per 1000 mammogram screens were 5.8, 5.9, and 6.3, in the BCSC, NBCCEDP, and NHSBSP, respectively. Recall rates were lower for subsequent examinations in all 3 settings but remained twice as high in the United States. A similar percentage of women underwent biopsy in each setting, but rates of percutaneous biopsy were lower and open surgical biopsy higher in the United States. Open surgical biopsies not resulting in a diagnosis of cancer (negative biopsies) were twice as high in the United States than in the United Kingdom. Based on a 10-year period of screening 1000 women aged 50 to 59 years, 477, 433, and 175 women in the BCSC, NBCCEDP, and NHSBSP, respectively, would be recalled; and for women aged 60 to 69 years, 396, 334, and 133 women, respectively. The estimated cancer detection rates per 1000 women aged 50 to 59 years were 24.5, 23.8, and 19.4, respectively, and for women aged 60 to 69 years, 31.5, 26.6, and 27.9, respectively.

Conclusions Recall and negative open surgical biopsy rates are twice as high in US settings than in the United Kingdom but cancer detection rates are similar. Efforts to improve US mammographic screening should target lowering the recall rate without reducing the cancer detection rate.

JAMA. 2003;290:2129-2137 www.jama.com

THE PROVISION OF SCREENING mammography differs greatly between the United States and the United Kingdom. In the United States, screening is provided in diverse settings, such as private practice, health maintenance organizations, and academic medical centers[1]; whereas in the United Kingdom, a single organized screening program run by the National Health Service provides virtually all mammographic screening for women aged 50 years or older.[2,3] There are also differences 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 findings 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 2 countries. Comparing the performance of screening mammography between the 2 countries may suggest methods to improve mammography practice.

We compared recall (the percentage of mammograms in which 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 United States and the United Kingdom.

Author Affiliations are listed at the end of this article. **Corresponding Author and Reprints:** Rebecca Smith-Bindman, MD, Department of Radiology, University of California–San Francisco, Mt Zion Campus, 1600 Divisadero St, San Francisco, CA 94115 (e-mail: rebecca.smith-bindman@radiology.ucsf.edu).

©2003 American Medical Association. All rights reserved.

Downloaded from www.jama.com at UCSF/Library, on April 12, 2007

METHODS

Data Sources

Data on US screening mammography 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 United Kingdom, data were obtained from the National Health Service Breast Screening Program (NHSBSP).[3,10] Results of all screening mammograms in women aged 50 years or older conducted within each of these settings between January 1, 1996, and December 31, 1999, were included. More than 1 screening examination was included if the examinations occurred more than 9 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 institutional review board at the University of California, San Francisco.

Breast Cancer Surveillance Consortium

The BCSC is a National Cancer Institute–funded consortium of mammography registries in San Francisco, Calif; Colorado; New Hampshire; New Mexico; North Carolina; Seattle, Wash; 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 1 of 202 contributing facilities. Data are obtained for individual women from self-administered questionnaires[12] and radiologist reports (medical records). Mammography results are reported using the categories of the American College of Radiology's Breast Imaging Reporting and Data Systems.[13] Cases of cancer are ascertained through active case follow-up and through linkages with state tumor registries, Surveillance, Epidemiology and End Result programs, or pathology databases, and cancer ascertainment has been found to be 94% complete.[14] Although all US facilities and radiologists must follow the Mammography Quality Standards Act/Mammography Quality Standards Reauthorization Act regulations,[15] the BCSC offers no specific guidelines for, or has authority in advising, how mammograms should be interpreted.

National Breast and Cervical Cancer Early Detection Program

The NBCCEDP, which is funded by the Centers for Disease Control and Prevention, provides breast and cervical cancer screening to poor uninsured women throughout the United States.[8,9] Although funding limitations have allowed only 15% to 20% of eligible women to be served, screening mammography for women aged 40 years or older have been provided in all 50 states, tribes, and territories since 1996. The Centers for Disease Control and Prevention funds each state, which in general contracts for the mammographic screening through diverse settings. Data are collected for individual women from self-administered questionnaires and medical records from primary physicians and radiologists. Mammography results are reported to the programs using the categories of Breast Imaging Reporting and Data Systems.[13] Cancer occurrences are ascertained primarily through active follow-up of mammograms with abnormal findings and review of pathology reports but some programs also link to state tumor registries or Surveillance, Epidemiology and End Result 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

The government-funded NHSBSP provides free breast cancer screening in the United Kingdom for women 50 years or older.[3,10] Women aged 50 to 64 years are invited by postcard to attend breast screenings every 3 years through a system that relies on centralized computer databases. From age 65 years onward, 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, London, England. 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 study.

Positive Mammogram

For the BCSC and NBCCEDP, a mammogram was classified as positive (recall) if the assessment was abnormal or incomplete (Breast Imaging Reporting and Data Systems[13] categories 0, 3, 4, and 5) and a recommendation for prompt diagnostic imaging, clinical evaluation, or biopsy (including fine-needle aspiration, core biopsy, and open surgical biopsy) was given. 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 pathological 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 subdivided into 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

©2003 American Medical Association. All rights reserved.

Downloaded from www.jama.com at UCSF/Library, on April 12, 2007

self-reported no prior mammogram within 5 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, Surveillance, Epidemiology and End Result program, or state tumor registry showed invasive carcinoma or ductal carcinoma in situ within 12 months of a positive screening mammogram. Cancers that occurred after a mammogram with negative findings (false-negative examinations) were not included in this analysis.

Statistical Analysis

Recall, noninvasive work-up, and biopsy rates were calculated per 100 screening mammograms and stratified by first or subsequent examinations and by 5-year age groups, or age-adjusted to a standard age distribution. The standard age distribution was the mean of the age distributions of the 3 data sources, in which each data source was weighted equally. The recall rate was calculated as the number of mammograms with positive findings per 100 screening mammograms. The noninvasive work-up rate was calculated as the number of recommendations for only noninvasive work-up, including ultrasound, diagnostic imaging, other noninvasive tests, or breast examination per 100 mammograms. Each mammogram was counted 1 time when calculating the noninvasive work-up rate, even if more than 1 test was recommended. The biopsy rate (any type of biopsy) was calculated as the number of mammograms with a recommendation for fine-needle aspiration, core biopsy, biopsy where the type was not specified, or open surgical biopsy per 100 mammograms. Each mammogram was counted 1 time when calculating the overall biopsy rate, even if more than 1 biopsy was recommended. The percutaneous biopsy rate was calculated as the number of fine-needle aspirations 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, hence these numbers do not sum to the overall biopsy rate. The open surgical biopsy rate was subdivided into 2 groups: those that resulted in a diagnosis of cancer (positive open surgical biopsy rate) and those that did not (negative open surgical biopsy rate). The specific method of biopsy could not be determined for the NBCCEDP data and for 3 of the 7 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 1000 examinations. The rate of invasive cancer by tumor size (<10 mm, 10-20 mm, or >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 United States than the United Kingdom, one would expect fewer cancers to be diagnosed per subsequent screening examination in the United States. To compare cancer detection rates for a similar period of screening, we used 4 years of actual data to estimate the number of cancers detected and women recalled per 1000 women undergoing screening mammography during a 10-year period. For these estimates, we assumed that screening started at age 50 years (or 60 years) 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 3 years[24] 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, and the age-specific recall rate and cancer rate of these first and subsequent examinations were those reported herein. We assumed that women aged 60 years or older underwent only subsequent examinations so only age-specific recall and cancer rates for subsequent screens were used to calculate 10-year estimates. We also assumed that the likelihood of recall and cancer detection were independent from one examination to the next, and that a woman could be recalled or have cancer detected only once. We estimated the chance of at least 1 recalled examination or cancer diagnosis during 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% confidence interval for the recall rates and cancer detection rates, 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% confidence interval for cancer detection and screening interval. SAS version 8.2 (SAS Institute Inc, Cary, NC) was used for all statistical analyses.

RESULTS

This analysis included 5.5 million mammograms: 978591 from the BCSC, 613388 from the NBCCEDP, and 3.94 million from the NHSBSP, which led to the diagnosis of 27612 cases of breast cancer among women aged 50 years or 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 2 US settings were approximately twice as high as those in the United Kingdom for all age groups,

©2003 American Medical Association. All rights reserved.

Downloaded from www.jama.com at UCSF/Library, on April 12, 2007

for first as well as subsequent examinations. Among first screening mammograms for women aged 50 to 54 years, 14.4% of women in the BCSC and 12.5% in the NBCCEDP vs only 7.6% in the NHSBSP were recalled for further evaluation, including diagnostic imaging, ultrasound, clinical examination, or biopsy. Biopsy rates were similar across all settings: 2.3% to 3.4% of first screening mammograms and 0.84% to 1.7% of subsequent screening examinations were followed up with a recommendation for biopsy. The higher US recall rate was primarily because of a higher rate of diagnostic imaging, ultrasound, and clinical evaluation.

Although the biopsy rates were similar between the 2 countries, biopsies were more likely to be open surgical biopsies in the United States (TABLE 3). For 100 first screening mammograms, 1.1% in the United States compared with 2.4% in the United Kingdom resulted in a recommendation for percutaneous biopsy; for 100 subsequent screening mammograms, 0.4% in the United States compared with 0.8% in the United Kingdom resulted in a recommendation for percutaneous bi-

Table 1. Mammography Registries and Programs and Number of Mammograms Obtained Between 1996-1999

	Data Sources			
	Breast Cancer Surveillance Consortium (BCSC), US National Cancer Institute	National Breast and Cervical Cancer Early Detection Program (NBCCEDP), US Centers for Disease Control and Prevention	UK National Health Service Breast Screening Program (NHSBSP)	Total
No. of mammograms	978 591	613 388	3 939 329	**5 531 308**
Age group, y				
50-54	264 229	196 407	1 581 190	**2 041 826**
55-59	198 122	163 026	1 239 908	**1 601 056**
60-64	160 039	147 815	1 056 997	**1 364 851**
≥65	356 201	106 140*	61 234	**523 575**
No. of breast cancers	4232	2711	20 669	**27 612**

*After 1997, most women aged 65 years or older were ineligible for the NBCCEDP because Medicare began to cover all costs for screening mammography. Most of the mammograms for the age group in NBCCEDP were performed from 1996 to 1997.

Table 2. Recommendations for Further Assessment per 100 Screening Mammograms by Age, Setting, and Screening Cycle

	Percentage (95% Confidence Interval)					
	First Screening Mammogram			Subsequent Screening Mammogram		
Age, y	BCSC	NBCCEDP	NHSBSP	BCSC	NBCCEDP	NHSBSP
			Recall*			
50-54	14.4 (13.9-14.9)	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	13.4 (12.8-14.0)	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	12.4 (11.7-13.1)	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	12.1 (11.6-12.5)	8.3 (7.9-8.7)	7.5 (6.8-8.3)	6.9 (6.8-7.0)	5.1 (4.9-5.2)	3.7 (3.5-3.8)
All†	13.1 (12.8-13.4)	11.2 (11.1-11.4)	7.4 (7.0-7.3)	8.0 (7.9-8.1)	6.8 (6.8-6.9)	3.6 (3.6-3.7)
			Noninvasive Diagnostic Workup‡			
50-54	12.1 (11.6-12.6)	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	11.1 (10.6-11.7)	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	9.9 (9.3-10.5)	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	9.2 (8.8-9.6)	5.6 (5.3-6.0)	4.1 (3.6-4.6)	5.9 (5.8-6.0)	3.9 (3.8-4.0)	2.3 (2.2-2.5)
All†	10.7 (10.4-10.9)	8.1 (8.0-8.3)	4.6 (4.5-4.7)	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)	0.97 (0.94-1.0)	1.7 (1.6-1.8)	0.84 (0.82-0.86)
55-59	2.3 (2.0-2.6)	3.1 (3.0-3.3)	2.3 (2.2-2.4)	0.99 (0.94-1.0)	1.4 (1.4-1.5)	0.90 (0.88-0.91)
60-64	2.5 (2.2-2.8)	3.4 (3.2-3.6)	2.5 (2.4-2.6)	1.0 (0.99-1.1)	1.4 (1.4-1.5)	0.97 (0.95-0.99)
≥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†	2.4 (2.3-2.6)	3.1 (3.0-3.2)	2.5 (2.5-2.6)	0.99 (0.97-1.02)	1.4 (1.4-1.5)	0.96 (0.94-0.98)

Abbreviations: BCSC, Breast Cancer Surveillance Consortium (US); NBCCEDP, National Breast and Cervical Cancer Early Detection Program (US); NHSBSP, National Health Service Breast Screening Program (UK).

*Recall includes any recommendation for further workup, including noninvasive imaging (ultrasound, diagnostic imaging, other tests), breast examination, or pathological 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.

‡Noninvasive workup (ultrasound, diagnostic imaging, other tests, or breast examination) but not a recommendation for pathological evaluation.

§Biopsy including any recommendation for pathological 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.

©2003 American Medical Association. All rights reserved.

Downloaded from www.jama.com at UCSF/Library, on April 12, 2007

opsy (age-adjusted data). Conversely, for 100 first screening mammograms, 1.15% in the United States compared with 0.72% in the United Kingdom resulted in a recommendation for open surgical biopsy (age-adjusted data). Most of the difference in open surgical biopsy rates was attributed to procedures among women who did not have breast cancer, with negative open surgical biopsy rates 2 to 3 times as high in the United States vs the United Kingdom. For 100 first screening examinations, 0.82% resulted in negative open surgical biopsy in the United States compared with 0.36%. Positive surgical biopsy rates were more similar between the 2 countries but tended to be higher in the United Kingdom for subsequent examinations.

The cancer detection rates increased with age and were 2 to 3 times as high for first vs subsequent mammograms in both countries (TABLE 4). Despite substantially higher recall rates in the United States, cancer detection rates were similar across settings, particularly for first screening examinations. For 1000 first examinations among women aged 50 to 54 years, 5.8, 5.9, and 6.3 cancers were diagnosed in the BCSC, NBCCEDP, and NHSBSP, respectively. Differences in cancer detection rates between the 2 countries were higher for subsequent examinations, likely reflecting more frequent US screenings.

The estimated number of cancers detected per 1000 women screened during 10 years was also similar between both countries (TABLE 5). If 1000 women aged 50 to 59 years underwent regular mammographic screening during 10 years, approximately 24.5 cancers would be detected in the BCSC, 23.8 in the NBCCEDP, and 19.4 in the NHSBSP. If 1000 women aged 60 to 69 years underwent regular mammographic screening during 10 years, approximately 31.5 cancers would be detected in the BCSC, 26.6 in the NBCCEDP, and 27.9 in the NHSBSP. Although invasive cancer detection rates are more similar between the 2 countries, the in situ cancer rates are higher in the United States. Among women aged 50 to 59 years, approximately 5.8, 7.4, and 3.8 in situ cancers would be detected, respectively. The higher frequency of screening in the United States magnifies the difference in the estimated recall rates between the countries when projected over 10 years. After 10 years of screening 1000 women aged 50 to 59 years, 477 women in the BCSC and 433 in the NBCCEDP vs 175 in the NHSBSP would have been recalled for additional work-up. After 10 years of screening women aged 60 to 69 years, 396 women in the BCSC and 334 in the NBCCEDP vs 133 in the NHSBSP would have been recalled for additional work-up.

For first screening mammograms, there were slightly fewer US invasive cancers diagnosed per 1000 examinations in most size categories (FIGURE). For subsequent examinations, there were lower rates of invasive cancer in all size categories in the United States vs the United Kingdom. The absolute difference in cancer rates between the United States and United Kingdom was highest for invasive tumors 10 to 20 mm.

COMMENT

The recall and negative open surgical biopsy rates associated with screening mammograms were twice as high in US settings than in the United Kingdom; however, cancer detection rates were similar in the 2 countries. In the United Kingdom, half as many women are recalled for diagnostic examinations and half as many women without breast cancer undergo open surgical biopsies as in the United States. These results observed in large numbers of women are similar to recent findings from a series of 60 test cases evaluated by physicians in both countries in whom false-positive rates were higher among US physicians but cancer detection rates were not.[25] The goal of any cancer screening effort is to obtain high can-

Table 3. Recommended Open Surgical Biopsy Rates per 100 Screening Mammograms by Age, Setting, and Screening Cycle*

	Percentage (95% Confidence Interval)			
	First Screening Mammogram		Subsequent Screening Mammogram	
Age, y	BCSC	NHSBSP	BCSC	NHSBSP
	Open Surgical Biopsy			
50-54	1.1 (0.91-1.3)	0.64 (0.62-0.65)	0.30 (0.27-0.33)	0.26 (0.25-0.27)
55-59	1.2 (0.90-1.5)	0.70 (0.65-0.75)	0.31 (0.27-0.35)	0.28 (0.27-0.29)
60-64	0.94 (0.67-1.3)	0.76 (0.69-0.83)	0.36 (0.32-0.41)	0.30 (0.29-0.31)
≥65	1.5 (1.2-1.7)	0.88 (0.65-1.2)	0.42 (0.39-0.45)	0.33 (0.29-0.38)
All†	1.15 (1.1-1.2)	0.72 (0.67-0.77)	0.33 (0.32-0.35)	0.28 (0.27-0.29)
	Positive Open Surgical Biopsy‡			
50-54	0.36 (0.25-0.49)	0.25 (0.24-0.26)	0.08 (0.06-0.10)	0.15 (0.14-0.16)
55-59	0.19 (0.09-0.35)	0.33 (0.29-0.37)	0.11 (0.09-0.14)	0.18 (0.17-0.19)
60-64	0.25 (0.13-0.45)	0.43 (0.38-0.48)	0.12 (0.10-0.15)	0.21 (0.20-0.22)
≥65	0.50 (0.37-0.67)	0.53 (0.36-0.77)	0.18 (0.16-0.20)	0.22 (0.18-0.26)
All†	0.31 (0.27-0.36)	0.36 (0.32-0.39)	0.11 (0.11-0.12)	0.18 (0.17-0.19)
	Negative Open Surgical Biopsy§			
50-54	0.74 (0.59-0.93)	0.39 (0.38-0.40)	0.22 (0.20-0.25)	0.11 (0.10-0.12)
55-59	0.98 (0.73-1.3)	0.37 (0.33-0.41)	0.20 (0.17-0.23)	0.10 (0.09-0.11)
60-64	0.69 (0.46-0.98)	0.34 (0.29-0.38)	0.24 (0.21-0.28)	0.09 (0.09-0.10)
≥65	0.95 (0.77-1.2)	0.36 (0.21-0.56)	0.24 (0.22-0.27)	0.11 (0.09-0.14)
All†	0.82 (0.75-0.89)	0.36 (0.33-0.39)	0.22 (0.21-0.23)	0.10 (0.10-0.11)

Abbreviations: BCSC, Breast Cancer Surveillance Consortium (US); NHSBSP, National Health Service Breast Screening Program (UK).
*For the NHSBSP and the 4 BCSC sites (Colorado, North Carolina, Seattle, Vermont), differentiation of the type of biopsy (percutaneous, including fine needle aspiration or core biopsy, vs open surgical biopsy) was performed. The positive and negative open surgical biopsy rate may not sum due to rounding.
†Adjusted to a standard age distribution.
‡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.

©2003 American Medical Association. All rights reserved.

Downloaded from www.jama.com at UCSF/Library, on April 12, 2007

Table 4. Cancers Detected per 1000 Screening Mammograms by Age, Setting, and Screening Cycle

	Rate per 1000 (95% Confidence Interval)					
	First Screening Mammogram			Subsequent Screening Mammogram		
Age, y	BCSC	NBCCEDP	NHSBSP	BCSC	NBCCEDP	NHSBSP
			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*	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*	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 (0.90-1.7)	1.4 (1.3-1.5)	0.77 (0.6-0.9)	1.1 (0.86-1.3)	0.70 (0.60-0.80)
55-59	0.63 (0.20-1.6)	2.1 (1.4-2.7)	1.4 (1.0-1.8)	0.73 (0.6-0.9)	1.1 (0.83-1.3)	0.90 (0.90-1.0)
60-64	2.4 (1.2-4.1)	3.0 (2.1-3.8)	2.2 (1.6-2.9)	0.96 (0.8-1.2)	1.2 (0.93-1.5)	1.0 (0.90-1.0)
≥65	2.0 (1.3-3.1)	1.7 (0.6-2.8)	1.8 (0.60-4.1)	1.0 (0.9-1.2)	1.6 (1.0-2.1)	1.7 (1.3-2.2)
All*	1.5 (1.2-1.8)	1.9 (1.7-2.2)	1.6 (1.4-1.9)	0.83 (0.77-0.90)	1.2 (1.1-1.3)	0.99 (0.92-1.1)

Abbreviations: BCSC, Breast Cancer Surveillance Consortium (US); NBCCEDP, National Breast and Cervical Cancer Early Detection Program (US); NHSBSP, National Health Service Breast Screening Program (UK).
*Adjusted to a standard age distribution.

Table 5. Estimated Number of Women With at Least 1 Recalled Examination, Cancer Diagnosis, or Biopsy During 10 Years*

	Rate per 1000 (95% Confidence Interval)		
	BCSC	NBCCEDP	NHSBSP
No. of women screened	1000	1000	1000
Time between screening examinations, mean, mo	18	19	36
Aged 50 to 59 years			
Cancer detected	24.5 (19.9-30.7)	23.8 (19.1-28.9)	19.4 (17.5-21.4)
In situ	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	476.6 (441.7-515.1)	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)
Aged 60 to 69 years			
Cancer detected	31.5 (26.3-37.8)	26.6 (20.7-34.2)	27.9 (24.1-32.3)
In situ	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	396.0 (354.9-435.4)	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)

Abbreviations: BCSC, Breast Cancer Surveillance Consortium (US); NBCCEDP, National Breast and Cervical Cancer Early Detection Program (US); NHSBSP, National Health Service Breast Screening Program (UK).
*The time between mammograms was assumed to be the mean interval observed in each setting. Data are estimated intervals in variation around these estimates and were calculated 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 3 of the BCSC sites.

cer 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 differences in recall rates between the 2 countries. Much higher rates of US malpractice lawsuits that focus on missed breast cancer diagnoses[27] provide a strong incentive to increase sensitivity at the expense of specificity, possibly leading US radiologists to recall women when they identify a finding with a low likelihood of cancer.[25,28] In addition, US physicians must read only 480 mammograms annually to fulfill Mammography Quality Standards Act requirements,[15] while radiologists in the United Kingdom are required to read at least 5000 mammograms annually[29] and on average, radiologists interpret 5 to 7 times their US counterparts. Furthermore, although more than 90% of programs in the United Kingdom use double reading, this practice is much less common in the United States. Although the ex-

©2003 American Medical Association. All rights reserved.

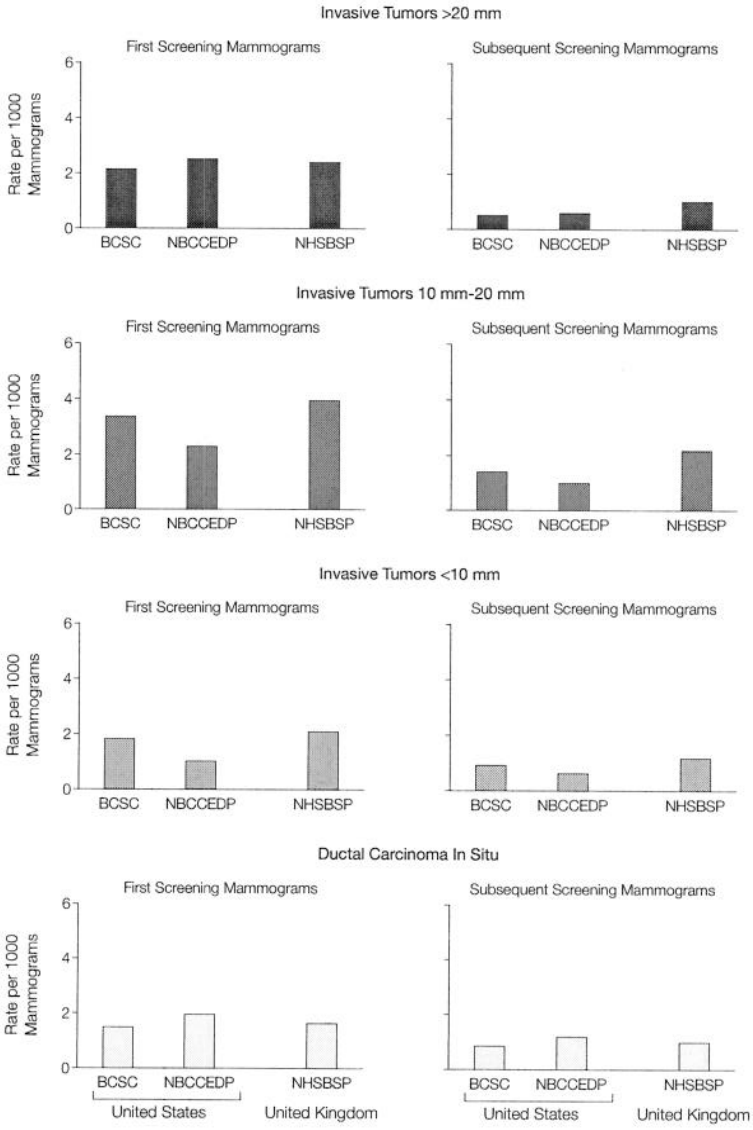
Downloaded from www.jama.com at UCSF/Library, on April 12, 2007

act impact of double reading remains uncertain, some evidence shows that double reading by consensus or arbitration, as used in the United Kingdom, raises detection rates and decreases recall rates.[30,31] Another consideration is the higher centralization of mammogram reading in the United Kingdom, as well as a less mobile population, which might make prior mammograms more readily available for comparison when interpreting results. Although the latter difference might reduce recall rates for subsequent mammograms,[32,33] it does not account for higher recall rates for first screening mammograms.

Most importantly, quality assurance standards for the NHSBSP programs are set nationally and are regularly monitored through a quality assurance network. Ranges of acceptable data 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 targets.[5,29] All screening programs in the United Kingdom receive data that enable a comparison of 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 United States 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 United Kingdom specifically provides instruction and individual feedback regarding recall and cancer detection rates by using a set of test mammography cases called PERFORMS.[35] Although not compulsory, 85% of mammographers from the United Kingdom participated in this test in 2001. Continuing medical education is a requirement for US radiologists but the actual content is not uniformly organized 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 United Kingdom has observed 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 United Kingdom observed a

Figure. Rate per 1000 Screening Examinations of Ductal Carcinoma In Situ and Invasive Breast Cancer by Size for First and Subsequent Screening Mammograms for Each Program, 1998-1999



NHSBSP indicates National Health Service Breast Screening Program; BCSC, Breast Cancer Surveillance Consortium; and NBCCEDP, National Breast and Cervical Cancer Early Detection Program. Results are age-adjusted. Detailed information on the size of the invasive breast cancers is only available in the United Kingdom from 1998 to 1999; therefore, the comparison of tumor size was limited to 1998-1999. In the United Kingdom, tumors that measured 10 mm were included with those that measured 11 to 20 mm; therefore, grouping of 10 to 20 mm for all 3 data sources were used. Because this cutpoint is different than typically used in the United States, the size distributions reported might be slightly different than reported by others in the United States.

©2003 American Medical Association. All rights reserved.

Downloaded from www.jama.com at UCSF/Library, on April 12, 2007

rapid 50% 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 United Kingdom[5] demonstrate that implementation of quality control can be efficient and feedback mechanisms effective. Despite the differences between the 2 countries in the provision and funding of screening mammography, mammography technology is very similar between the 2 countries and similar targets for mammography outcomes, including specific recall and biopsy rates, could be established in the United States. Success in reaching technical targets set by the Mammography Quality Standards Act/Mammography Quality Standards Reauthorization Act demonstrates how a coordinated quality assurance program can work in the United States.[38]

Screening mammography is performed more frequently in the United States than in the United Kingdom. During a 10-year period, women aged 50 years or older will undergo approximately 7 mammograms in the United States vs only 3 in the United Kingdom. More frequent screening likely translates into smaller average cancer size at diagnosis, as evidenced by the slightly lower rates of invasive cancer for 10 mm or more and the higher rates of in situ cancer diagnosed in the United States. Additionally, US screening tends to begin at an earlier age than in the United Kingdom. From our results, it cannot be determined whether these differences affect breast cancer mortality.

We compared the cancer detection rates, which 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 both countries to be similar. Given that the overall age-adjusted breast cancer incidence rates are slightly higher in the United States,[41] one would expect that the United States would have similar or higher cancer detection rates than in the United Kingdom. Thus, it is unlikely that the United Kingdom is missing cancers despite a much lower recall rate. 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 United States.

The main limitation of our study is that we cannot be certain that our definition of screening mammography was the same across all 3 settings. Specifically, we do not know if there was a higher proportion of diagnostic examinations among the US women, which might account for a higher recall rate. However, more diagnostic mammograms should produce a substantially higher cancer rate[43,44] in the 2 US settings, which we did not find. We should also note that our estimation of the total cancers detected during 10 years was based on only 4 years of screening data and the assumptions of the model were simplistic. 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 2 US data sources but this is estimated to be less than 3% of the mammograms described. 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 United Kingdom includes virtually all mammographic screening performed, the US data reflects only a small percentage of mammography performed. Because mammograms from all 50 US states were included and the results from the BCSC and NBCCEDP were broadly similar, these results probably provide the best current evidence of the US performance of mammography screening.

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 2 US data sources were small compared with the differences between the 2 countries and these programs describe different populations, in which breast cancer rates, as well as tumor characteristics, might be different.

Women undergoing screening mammography should consider going to facilities where physicians read a large number of mammography examinations,[28,40] radiologists devote a large percentage of their practice to mammography,[40] and 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 examination with abnormal findings and an open surgical biopsy is recommended, they should discuss all options with a radiologist or surgeon, and consider getting a second opinion.

In the United Kingdom, the NHSBSP has set and reached targets that emphasize high rates of cancer detection and low recall. Recall rates in the United Kingdom are now substantially lower than in the United States 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 to 69 years biennially and reducing recall rates could substantially decrease the cost of mammography, as well as associated anxiety caused by false-positive diagnoses.[49] Efforts to improve US mammographic screening should be targeted to lowering the recall rate without substantially lowering the cancer detection rate.

Author Affiliations: Department of Radiology (Drs Smith-Bindman and Sickles, and Mr Chu), Department of Epidemiology and Biostatistics (Drs Smith-Bidman and Kerlikowske), and Department of Medicine and General Internal Medicine Section, Department of Veterans Affairs (Dr Kerlikowske), University of California, San Francisco; Center for Health Studies, Group Health Cooperative, and Department of Biostatistics, University of Washington, Seattle (Dr Miglioretti); Cancer Screening Evaluation Unit, Institute of Cancer Research, University of London, England (Dr Blanks); Applied Research Program, Na-

©2003 American Medical Association. All rights reserved.

Downloaded from www.jama.com at UCSF/Library, on April 12, 2007

tional Cancer Institute, National Institutes of Health, Bethesda, Md (Dr Ballard-Barbash); Centers for Disease Control and Prevention, Atlanta, Ga (Drs Bobo and Lee); Warwickshire, Solihull, and Coventry Breast Screening Service, Coventry, England (Dr Wallis); and National Health Service Breast Screening Programs, Sheffield, England (Ms Patnick). Dr Bobo is now with the Battelle Centers for Public Health Research and Evaluation, Seattle, Wash.

Author Contributions: *Study concept and design:* Smith-Bindman, Ballard-Barbash, Kerlikowske.
Acquisition of data: Smith-Bindman, Blanks, Lee, Patnick.
Analysis and interpretation of data: Smith-Bindman, Chu, Miglioretti, Sickles, Blanks, Ballard-Barbash, Bobo, Lee, Wallis, Patnick, Kerlikowske.
Drafting of the manuscript: Smith-Bindman, Chu, Sickles, Ballard-Barbash, Bobo, Kerlikowske.
Critical revision of the manuscript for important intellectual content: Smith-Bindman, Chu, Miglioretti, Sickles, Blanks, Ballard-Barbash, Bobo, Lee, Wallis, Patnick, Kerlikowske.
Statistical expertise: Smith-Bindman, Chu, Miglioretti, Blanks, Bobo.
Obtained funding: Smith-Bindman, Kerlikowske.
Administrative, technical, or material support: Smith-Bindman, Chu, Miglioretti, Bobo, Lee, Patnick.
Study supervision: Smith-Bindman, Ballard-Barbash.
Funding/Support: This work was supported in part by grants from the National Cancer Institute (CA86032), Breast Cancer Surveillance Consortium cooperative agreement U01CA63740, the National Cancer Institute–funded Statistical Coordinating Center (U01CA86076), and the Department of Defense (DAMD179919112).
Acknowledgment: We thank Travis Seawards for his invaluable assistance in preparing the tables and manuscript; Kim Lovy, for her analysis of the United Kingdom data; Sandy Thames, for her help obtaining and interpreting the data from the Centers for Disease Control and Prevention; and Deborah Grady, MD, MPH, Virginia Ernster, PhD, and William Barlow, PhD, for their helpful comments on the manuscript.

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. UK National Health Service Breast Screening Web site. The NHS Breast Screening Programme. Available at: http://www.cancerscreening.nhs.uk/breastscreen. Accessibility verified September 10, 2003.
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 Am J Roentgenol.* 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, for International Breast Cancer Screening Network (IBSN) and the European Network of Pilot Projects for Breast Cancer Screening. Breast cancer screening programmes in 22 countries: current policies, administration and guidelines. *Int J Epidemiol.* 1998;27:735-742.
7. Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. *AJR Am J Roentgenol.* 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; 2:36-47.
9. May D, Lee N, Nadel M, et al. The National Breast and Cervical Cancer Early Detection Program: report on the first 4 years of mammography provided to medically underserved women. *AJR Am J Roentgenol.* 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, England: Her Majesty's Stationery Office; 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. *Ann Intern Med.* 2003;138:168-173.
12. Breast Cancer Surveillance Consortium Web site. Available at: http://www.breastscreening.cancer.gov/elements.html#guestionnaires. Accessibility verified September 10, 2003.
13. American College of Radiology. *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, et al. Detection of DCIS in women undergoing screening mammography. *J Natl Cancer Inst.* 2002;94:1151-1159.
15. Pub L No. 102-539.
16. Kerlikowske K, Grady D, Barclay J, et al. 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, et al. Medical audit of a rapid-throughput mammography screening practice: methodology and results of 27,114 examinations. *Radiology.* 1990;175:323-327.
18. Leitch AM, Dodd GD, Costanza M, et al. American Cancer Society guidelines for the early detection of breast cancer: update 1997. *CA Cancer J Clin.* 1997; 47:150-153.
19. Feig SA, D'Orsi CJ, Hendrick RE, et al. American College of Radiology guidelines for breast cancer screening. *AJR Am J Roentgenol.* 1998;171:29-33.
20. US Preventive Services Task Force. Screening for breast cancer: recommendations and rationale. Available at: http://www.ahrq.gov/clinic/3rduspstf/breastcancer/brcanrr.htm#consideration. Accessibility verified September 10, 2003.
21. US 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, et al. 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, et al. Positive predictive value of screening mammography by age and family history of breast cancer. *JAMA.* 1993;270: 2444-2450.
24. UK Department of Health. *Statistical Bulletin, Breast Screening Programme, England: 1999-2000.* England: National Statistics; 2001. Bulletin 2001/10.
25. 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.
26. Moss M. Spotting breast cancer: doctors are weak link. *New York Times.* June 27, 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, et al. 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, England: National Health Service Breast Screening Program; 1997. Publication 15.
30. Kopans D. Double reading. *Radiol Clin North Am.* 2000;38:719-724.
31. Williams LJ, Hartswood M, Prescott RJ. Methodological issues in mammography double reading studies. *J Med Screen.* 1998;5:202-206.
32. Christiansen C, Wang F, Barton M, et al. Predicting the cumulative risk of false-positive mammograms. *J Natl Cancer Inst.* 2000;92:1657-1666.
33. Burnside ES, Sickles EA, Sohlich RE, Dee KE. Differential value of comparison with previous examinations in diagnostic versus screening mammography. *AJR Am J Roentgenol.* 2002;179:1173-1177.
34. Bassett L, Hendrick R, Bassford T, et al. *Quality Determinants of Mammography: Clinical Practice Guideline #13.* Rockville, Md: Agency for Health Care Policy and Research; 1994. AHCPR Publication 95-0632.
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, ed. *Medical Imaging 1999: Image Perception and Performance.* San Diego, Calif: Spie, Bellingham, Washington; 1999:157-168.
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 1998-March 1999.* Sheffield, England: National Health Service Breast Screening Program; 2000.
38. Suleiman O, Spelic D, McCrohan J, et al. Mammography in the 1990's: the United States and Canada. *Radiology.* 1999;210:345-351.
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;14:449-461.
40. Sickles E, Wolverton D, Dee K. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology.* 2002;224:861-869.
41. International Agency for Research on Cancer (IARC). Available at: http://www.iarc.fr/. Accessibility verified September 10, 2003.
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;177:543-549.
43. Sohlich R, Sickles E, Burnside E, Dee K. Interpreting data from audits when screening and diagnostic mammography outcomes are combined. *AJR Am J Roentgenol.* 2002;178:681-686.
44. Dee K, Sickles E. Medical audit of diagnostic mammography examinations: comparison with screening outcomes obtained concurrently. *AJR Am J Roentgenol.* 2001;176:729-733.
45. Elmore JG, Barton MB, Moceri VM, et al. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med.* 1998;338: 1089-1096.
46. Beam CA, Sullivan DC, Layde PM. Effect of human variability on independent double reading in screening mammography. *Acad Radiol.* 1996;3:891-897.
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;127:955-965.
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;2:145-152.
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.

©2003 American Medical Association. All rights reserved.

Downloaded from www.jama.com at UCSF/Library, on April 12, 2007

mester at delivery), severity of pre-eclampsia, gestational age at delivery, and birth weight (Table 1) were comparable with those from previous studies from Western Europe.[2,3,5] In light of the lack of ethnicity-related differences in ADMA concentrations in our sample, ethnicity does not appear to explain our results.

Thus, our results support the hypothesis that pre-eclampsia in low- and high-risk populations may have distinct underlying causes.

Renke Maas, MD
maas@uke.uni-hamburg.de
Rainer H. Böger, MD
Edzard Schwedhelm, PhD
AG Klinische Pharmakologie
Institut für Experimentelle und Klinische Pharmakologie
Universitätsklinikum Hamburg Eppendorf
Hamburg, Germany

Juan P. Casas, MD
Patricio López-Jaramillo, MD, PhD
Instituto Colombiano de Investigaciones Biomédicas
Bucaramanga
Colombia

Norma Serrano, MD, MSc
Luis A. Díaz, MD
Universidad Autónoma de Bucaramanga
Bucaramanga

Funding/Support: Financial support to the VILANO group and the Pregeno project was provided by COLCIENCIAS (Colombia).

Acknowledgment: We thank Mariola Kastner and Anna Steenpaß for excellent technical assistance, and the midwives and the pregnant women who participated in this study.

1. Roberts JM, Cooper DW. Pathogenesis and genetics of pre-eclampsia. *Lancet*. 2001;357:53-56.
2. Fickling SA, Williams D, Vallance P, Nussey SS, Whitley GS. Plasma concentrations of endogenous inhibitor of nitric oxide synthesis in normal pregnancy and pre-eclampsia. *Lancet*. 1993;342:242-243.
3. Holden DP, Fickling SA, Whitley GS, Nussey SS. Plasma concentrations of asymmetric dimethylarginine, a natural inhibitor of nitric oxide synthase, in normal pregnancy and preeclampsia. *Am J Obstet Gynecol*. 1998;178:551-556.
4. Savvidou MD, Hingorani AD, Tsikas D, Frölich JC, Vallance P, Nicolaides KH. Endothelial dysfunction and raised plasma concentrations of asymmetric dimethylarginine in pregnant women who subsequently develop pre-eclampsia. *Lancet*. 2003;361:1511-1517.
5. Ellis J, Wennerholm UB, Bengtsson A, et al. Levels of dimethylarginines and cytokines in mild and severe preeclampsia. *Acta Obstet Gynecol Scand*. 2001;80:602-608.
6. Herrera JA, Salmeron B, Hurtado H. Prenatal biopsychosocial risk assessment and low birth weight. *Soc Sci Med*. 1997;44:1107-1114.

CORRECTION

Incorrect Data in Table: In the Original Contribution entitled "Comparison of Screening Mammography in the United States and the United Kingdom" published in the October 22/29, 2003, issue of THE JOURNAL (2003;290:2129-2137), there were incorrect data in Table 5 due to rounding errors. On page 2134, for cancer detected for women aged 50 to 59 years, the rate per 1000 women of 24.5 should have been 24.8 for BCSC and 19.4 should have been 19.2 for NHSBSP, and for cancer detected in women aged 60 to 69 years, the rate per 1000 women of 31.5 should have been 31.6 for BCSC and 26.6 should have been 27.2 for NBCCEDP.

©2004 American Medical Association. All rights reserved.

Downloaded from www.jama.com at UCSF/Library, on April 12, 2007

Physician Predictors of Mammographic Accuracy

Rebecca Smith-Bindman, Philip Chu, Diana L. Miglioretti, Chris Quale, Robert D. Rosenberg, Gary Cutter, Berta Geller, Peter Bacchetti, Edward A. Sickles, Karla Kerlikowske

Background: **The association between physician experience and the accuracy of screening mammography in community practice is not well studied. We identified characteristics of U.S. physicians associated with the accuracy of screening mammography.** ***Methods:*** **Data were obtained from the Breast Cancer Surveillance Consortium and the American Medical Association Master File. Unadjusted mammography sensitivity and specificity were calculated according to physician characteristics. We modeled mammography sensitivity and specificity by multivariable logistic regression as a function of patient and physician characteristics. All statistical tests were two-sided.** ***Results:*** **We studied 209 physicians who interpreted 1 220 046 screening mammograms from January 1, 1995, through December 31, 2000, of which 7143 (5.9 per 1000 mammograms) were associated with breast cancer within 12 months of screening. Each physician interpreted a mean of 6011 screening mammograms (95% confidence interval [CI] = 4998 to 6677), including a mean of 34 (95% CI = 28 to 40) from women diagnosed with breast cancer. The mean sensitivity was 77% (range = 29%–97%), and the mean false-positive rate was 10% (range = 1%–29%). After adjustment for the patient characteristics of those whose mammograms they interpreted, physician characteristics were strongly associated with specificity. Higher specificity was associated with at least 25 years (versus less than 10 years) since receipt of a medical degree (for physicians practicing for 25–29 years, odds ratio [OR] = 1.54, 95% CI = 1.14 to 2.08; P = .006), interpretation of 2500–4000 (versus 481–750) screening mammograms annually (OR = 1.30, 95% CI = 1.06 to 1.59; P = .011) and a high focus on screening mammography compared with diagnostic mammography (OR = 1.59, 95% CI = 1.37 to 1.82; P<.001). Higher overall accuracy was associated with more experience and with a higher focus on screening mammography. Compared with physicians who interpret 481–750 mammograms annually and had a low screening focus, physicians who interpret 2500–4000 mammograms annually and had a high screening focus had approximately 50% fewer false-positive examinations and detected a few less cancers.** ***Conclusion:*** **Raising the annual volume requirements in the Mammography Quality Standards Act might improve the overall quality of screening mammography in the United States. [J Natl Cancer Inst 2005;97:358–67]**

Screening mammography is a nonspecific test for breast cancer, because only 5%–10% of screening mammograms that are interpreted as abnormal harbor cancer *(1–5)*. Although patient characteristics such as age and breast density contribute to variations in reported mammographic accuracy *(1,6,7)*, it is not clear how physician characteristics affect variability in accuracy.

A growing body of evidence has shown that physicians with greater experience in performing procedures, such as cardiac angioplasty *(8)*, have a higher proportion of patients with good outcomes *(9)*. Physician training in mammographic interpretation has been associated with improved accuracy *(10,11)*. The few studies that have evaluated the relationship between annual volume of mammographic interpretation and accuracy, however, have obtained conflicting results. Some studies have reported that volume is of prime importance *(12,13)*, whereas others have reported that accuracy is associated with the interplay of many interrelated factors involving physician experience but that volume itself is not important *(14,15)*. However, all of these studies *(12–15)* used practice sets of mammograms that were greatly enriched with mammograms showing cancer; some of these practice sets contained up to 100 times more cancer-associated mammograms than generally encountered in actual practice, which raises concerns about context bias *(16,17)*. Two studies evaluated the association between mammographic volume and accuracy with the prospective interpretation of clinical mammograms by a small number of physicians *(18,19)* and found that physicians who read higher volumes of mammograms tended to have improved accuracy. No large study has evaluated the association between physicians' volume and accuracy by use of prospectively collected clinical data in the United States on a broad sample of physicians.

In the United States, the Mammography Quality Standards Act of 1992 requires physicians to interpret at least 960 mammographic examinations within a 2-year period to qualify to interpret mammograms *(20)*. This minimum is 10-fold lower than the number required by the United Kingdom National Health Service Breast Screening Program *(21)* and reflects a minimum volume of approximately 10 mammograms per week. Although it seems reasonable to assume that increasing experience will improve the accuracy of mammographic interpretation, the values chosen by the Mammography Quality Standards Act and the National Health Service Breast Screening Program were arbitrary

Affiliations of authors: Departments of Radiology (RS-B, PC, CQ, EAS) and Epidemiology and Biostatistics (RS-B, PB, KK), University of California, San Francisco, CA; Center for Health Studies, Group Health Cooperative and Department of Biostatistics, University of Washington, Seattle, WA (DLM); Department of Radiology, University of New Mexico, Albuquerque, NM (RDR); Center for Research Design and Statistical Methods, University of Nevada School of Medicine, Applied Research Facility, Reno, NV (GC); Health Promotion Research, University of Vermont, College of Medicine, Burlington, VT (BG); General Internal Medicine Section, Department of Veterans Affairs, University of California, San Francisco, CA (KK).

Correspondence to: Rebecca Smith-Bindman, MD, Department of Radiology, University of California, San Francisco, 1600 Divisadero St., San Francisco, CA 94115 (e-mail: Rebecca.Smith-Bindman@Radiology.UCSF.Edu).

See "Notes" following "References."

DOI: 10.1093/jnci/dji060

Journal of the National Cancer Institute, Vol. 97, No. 5, © Oxford University Press 2005, all rights reserved.

minima derived primarily from perceptions about the supply of physicians able to interpret mammograms rather than from actual data to ensure adequate practice and skill *(22)*. The purpose of this study was to evaluate physician predictors associated with accuracy of screening mammographic interpretation in community practice in the United States.

Patients and Methods

Data Sources

We obtained data on mammographic interpretations, volume and cancer outcomes from mammography registries that participate in the Breast Cancer Surveillance Consortium *(1,23,24)*, a National Cancer Institute–funded consortium that collects patient demographic and clinical information *(25)*, mammographic interpretation, and cancer diagnoses from participating facilities in seven states. Four registries—Colorado (Colorado Mammography Project), New Mexico (New Mexico Mammography Project), San Francisco (San Francisco Mammography Registry), and Vermont (Vermont Breast Cancer Surveillance System)—contributed data to this study. Details of data collection have been reported previously *(1,26–30)*. The Breast Cancer Surveillance Consortium links data within registries from patient surveys and radiologist reports and ascertains cancer outcomes through linkage with state tumor registries (Colorado and Vermont), Surveillance Epidemiology and End Results (SEER[1]) tumor registries (San Francisco and New Mexico), and pathology databases (Vermont and New Mexico).

Physician characteristics (age and years since receipt of medical degree) were obtained from the American Medical Association Physician Profile Service *(31)*. Linkage with the Breast Cancer Surveillance Consortium data was done in a way that maintained physician confidentiality. Institutional Review Boards of all collaborating institutions approved the study.

Subjects

The study subjects were physicians who interpreted screening mammograms between January 1, 1995, and December 31, 2000. Overall, 95% of physicians who practice at facilities that participate in the Breast Cancer Surveillance Consortium were included in the analysis. We excluded screening examinations that occurred after December 31, 2000, to ensure at least 12 months follow-up for a cancer diagnosis after a normal or abnormal screening result and an additional 18 months for the cancer to be reported to the tumor registries, which would provide a cancer ascertainment that was at least 94.3% complete *(26)*. We assumed that all physicians interpreted an average of at least 480 mammograms per year, the minimum number required by Mammography Quality Standards Act guidelines, although a particular mammography registry may not capture all interpretations. Consequently, we excluded 45 physicians who appeared to interpret less than an average of 480 mammograms annually or during each year of the study period, because the volume of mammographic interpretations estimated for these physicians is likely to be inaccurate. The mean annual volume of the 45 excluded physicians was 388 mammographic interpretations (95% confidence interval [CI] = 372 to 405 mammographic interpretations). For any physician, we also excluded any calendar year during which that physician interpreted less than 300 mammograms. For example, a physician who read 1200, 1100, 200, and 1300 mammograms in each year of the 4-year study would be included, but his or her accuracy and annual volume would not be assessed during the third year.

Among the 209 physicians, the mean age (± standard deviation) was 52.2 ± 9.6 years, the mean number of years since receipt of a medical degree was 24.5 ± 10.6 years, and 46 were female (Table 1).

Mammographic Volume and Screening Focus

We calculated each physician's mean annual volume of mammographic interpretations (including both screening and diagnostic examinations) over the study period and then stratified annual volume into groups that had been used by others *(13,18)*, and we roughly balanced the number of physicians in each group when possible. The mean annual volume of mammographic interpretations was 1572, and the mean ranged from 1397 to 1928 across the four registries ($P = .01$). The median annual volume was 1054, and the median ranged from 835 to 1682 across the four registries ($P = .01$). Of the 209 physicians, 63 (30.1%) interpreted 481–750 mammograms annually, for a total of 123 789 (10.2%) of all 1 220 046 screening mammograms in this study. An additional 32 physicians (15.3%) interpreted 751–1000 mammograms annually, for a total of 91 801 (7.5%) of all 1 220 046 screening mammograms. Thus, 95 (45.4%) of the physicians interpreted fewer than 1001 mammograms annually, and these physicians interpreted 17.7% of all screening mammograms.

We assessed each physician's relative focus on screening as opposed to diagnostic mammography as their ratio of screening to diagnostic mammograms interpreted. The median ratio of screening to diagnostic mammographic examinations was 5.6 (interquartile range = 4.2–7.6), and this ratio was comparable across the four registries. We dichotomized this ratio at 5 (<5

Table 1. Characteristics of physicians included in this study

Characteristic	No. (%)
Sex	
Male	163
Female	46 (22.0)
Physician age	
<40 y	22 (11.3)
40–49 y	60 (30.8)
50–59 y	73 (37.4)
60–69 y	33 (16.9)
≥70 y	7 (3.6)
Time since receipt of medical degree	
<10 y	16 (07.7)
10–14 y	26 (12.4)
15–19 y	37 (17.7)
20–24 y	27 (12.9)
25–29 y	24 (11.5)
30–34 y	43 (20.6)
>34 y	35 (17.2)
Average annual volume of mammogram interpretation	
481–750 mammograms	63 (30.1)
751–1000 mammograms	32 (15.3)
1001–1500 mammograms	41 (19.6)
1501–2500 mammograms	43 (20.6)
2501–4000 mammograms	16 (7.7)
>4000 mammograms	14 (6.7)
Ratio of screening to diagnostic mammograms	
<5	81 (0.3)
≥5	128 (0.6)

vs. ≥5) as a round cut point that approximately balanced the numbers of physicians in these two groups.

Screening Mammography Accuracy

We calculated annual volume and screening focus from all of a physician's interpretations but restricted the analysis of mammography accuracy to screening examinations. We considered mammograms to be diagnostic whenever the woman reported a breast symptom [consistent with the American College of Radiology Breast Imaging Reporting and Data Systems (BI-RADS) *(32)*] or the mammogram occurred within 9 months of a previous screening examination. Women could have more than one screening examination included as long as the interval between examinations was more than 9 months.

A screening mammogram was classified as positive *(32)* if the initial assessment was incomplete or suspicious for cancer (BI-RADS interpretations 0, 4, or 5; n = 92439 or 7.6% of total screening mammograms) or if the initial assessment was "probably benign" (BI-RADS interpretation 3) but had a recommendation for immediate further assessment (n = 27753 or 2.3% of total screening mammograms). The remaining mammograms were classified as negative. Mammograms without a BI-RADS assessment were excluded from the analyses (0.10% of total screening mammograms). Women were considered to have breast cancer if reports from a breast pathology database, SEER program, or state tumor registry showed invasive carcinoma or ductal carcinoma in situ within 12 months of the index mammogram.

If breast cancer was diagnosed within 12 months of a positive screening mammogram, the mammogram was considered a true positive. If breast cancer was diagnosed within 12 months of a negative screening mammogram, the mammogram was considered a false negative. If no breast cancer was diagnosed within 12 months of a negative screening mammogram, the mammogram was considered a true negative. If no breast cancer was diagnosed within 12 months of a positive screening mammogram, the mammogram was considered a false positive.

To adjust each physician's accuracy according to the characteristics of his or her patients, we included patient age, physician-reported assessment of breast density, and a classification of mammographic examination as a first or a subsequent examination in our multivariable models. Breast density was classified as almost entirely fat, scattered fibroglandular densities, heterogeneously dense, or extremely dense. A mammogram was considered a patient's "first" mammogram if there was no registry record of a prior mammogram within 4 years and if the patient reported no prior mammogram within 4 years. Remaining mammograms were considered subsequent.

Statistical Analysis

We calculated the overall sensitivity and specificity of screening mammography for each physician. Whenever the value in any cell was equal to zero, we added 0.5 to the value in all cells to obtain a less extreme value. Unadjusted mammographic sensitivity and specificity were calculated according to patient characteristics (age, breast density, and whether examination was a first or a subsequent) and physician characteristics (age, years since receipt of medical degree, average annual volume of mammogram interpretations, and ratio of screening to diagnostic mammographic interpretations). We plotted the sensitivity against the false-positive rate of screening mammography, with each physician contributing a single point to this graph. We then graphed the sensitivity and false-positive rate of screening mammograms stratified by physician characteristics, with each mammogram weighed equally.

We modeled sensitivity and specificity as a function of patient and physician characteristics by use of multivariable logistic regression. Because of the colinearity of physician age and time since receipt of medical degree, only the latter was included in the multivariable analysis. To determine whether patient and physician characteristics influence the threshold at which a physician operates (which results in a tradeoff between sensitivity and specificity) or the accuracy of mammographic interpretation (additional probability of a positive mammogram if a woman has cancer), we jointly modeled the false-positive rate (1 minus the specificity) and true-positive rate (sensitivity) in a single receiver operator characteristic (ROC)–type logistic regression model. This model included main effects for each covariate and cancer status plus interactions of each covariate with cancer status *(33)*. Specifically,

$$\text{logit}[p(y_i = 1 \mid x_i, d_i)] = x_i\boldsymbol{\beta} + x_i d_i \boldsymbol{\delta},$$

where y_i is the mammography outcome (1 if positive, 0 if negative) for the ith woman, $\boldsymbol{xi}$ is a vector of her covariate values including an intercept term, and d_i is an indicator of whether or not she had cancer diagnosed during the 1-year follow-up period. By use of this notation, the false-positive rate for the covariate combination × is defined as $p(y = 1|\boldsymbol{x}, d = 0)$, which is equal to the inverse logit of $\boldsymbol{x\beta}$. Sensitivity is $p(y = 1|\boldsymbol{x}, d = 1)$, which is equal to the inverse logit of $\boldsymbol{x}(\boldsymbol{\beta} + \boldsymbol{\delta})$. Thus, the β coefficients measure the influence of × on the overall probability of a recall (i.e., threshold effect), and $\boldsymbol{\delta}$ measures the additional influence of × on the probability of a recall given that the woman has cancer (i.e., accuracy effect). If $\boldsymbol{\delta} = 0$, then the covariate × influences the false-positive rate and sensitivity equally. This model allowed us to evaluate differences in interpretive performance that reflect a threshold effect (i.e., a shift along an ROC curve; in Fig. 1, movement from point A to point B) versus an accuracy effect (i.e., differences that reflect performance on a different ROC curve; in Fig. 1, movement from point A to point C). We report multivariable results for specificity, sensitivity, and overall accuracy. Odds ratios (ORs) for sensitivity and specificity reflect how well physicians performed with respect to a given covariate along an ROC curve (if the accuracy effect is not statistically significant), whereas odds ratios for accuracy reflect a shift associated with a given covariate to a new ROC curve. For example, given an overall ROC curve for physicians, a statistically significant positive accuracy effect means a given covariate is associated with a shift to a different ROC curve that reflects better performance. An improvement in accuracy can reflect a statistically significant increase in the specificity without a corresponding statistically significant reduction in the sensitivity, a statistically significant increase in the sensitivity without a statistically significant decrease in the specificity, or an improvement in both sensitivity and specificity. If the accuracy effect is not statistically significantly different from 1, changes in specificity or sensitivity associated with a covariate reflect a shift along an ROC curve as opposed to a shift to a different ROC curve (Fig. 1). The models were fit by way of generalized estimating equations *(34)* with an independent working covariance matrix by

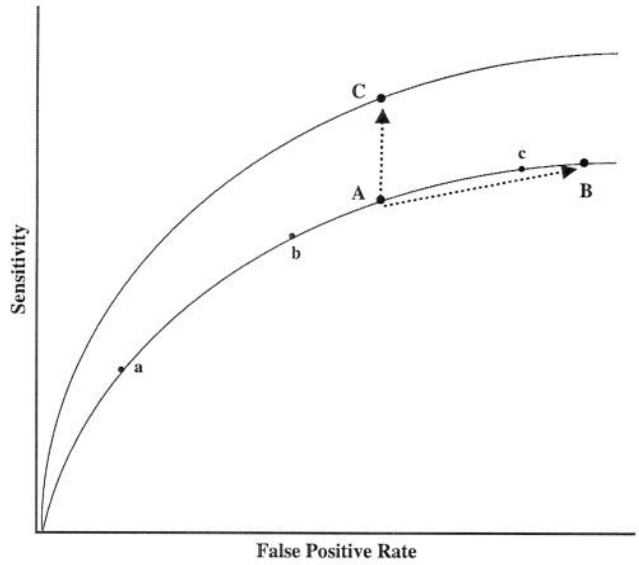


Fig. 1. Mammography accuracy and the interpretation threshold. Differences in physician performance that reflect an improvement in accuracy are shown by the shift from point A to C. Differences in physician performance that reflect a shift in the threshold used to interpret an examination as abnormal are shown by a shift from point A to B.

use of the GENMOD procedure in the SAS package (version 8.2; SAS Institute, Cary NC) of programs to account for the correlation among multiple mammograms interpreted by the same physician.

To demonstrate the real-world implications of differences in accuracy, we used the estimated sensitivity, specificity, and positive predictive values for all possible combinations of annual volume and screening focus to calculate expected numbers of cancers detected and false-positive diagnoses per 10 000 women screened annually, standardized to a single population of women with the covariate distribution and the same number of cancers (5.9 cancers per 1000 mammograms) as observed in this cohort. All statistical tests were two-sided.

Results

The study subjects were 209 physicians who interpreted 1 220 046 screening mammograms between January 1, 1995, and December 31, 2000, including 7143 (5.9 per 1000 mammograms) diagnosed as breast cancer within 12 months of the screening mammogram. Each physician interpreted a mean of 6011 screening mammograms (95% CI = 4998 to 6677) of which a mean of 34 (95% CI = 28 to 40) were from women diagnosed with breast cancer within 12 months of the index mammogram.

Sensitivity and Specificity of Mammography by Patient Characteristics

The sensitivity and specificity of mammographic interpretation varied substantially and statistically significantly by patient characteristics (Table 2). For example, for subsequent screening mammograms, as patient age increased from younger than 40 years to older than 70 years, the false-positive rate decreased from 10.5% (95% CI = 10.1 to 10.9) to 6.5% (95% CI = 6.4 to 6.6) and the sensitivity increased from 52.7% (95% CI = 39.5 to 65.9) to 79.7% (95% CI = 77.6 to 81.9). The false-positive rate was lower, and sensitivity was higher when breast density was predominantly fat or contained scattered fibroglandular densities. Lower false-positive rates were observed for subsequent examinations than for first examinations, whereas higher sensitivities were observed for first screening examinations.

Physician Variability in Mammography Sensitivity and False-Positive Rates

Physicians exhibited wide variations in mammography sensitivity and specificity. The mean sensitivity was 77% (range = 29%–97%, 95% CI = 76% to 79%), and the mean false-positive rate was 10% (range = 1%–29%, 95% CI = 9% to 10%). The

Table 2. Accuracy of screening mammography by patient characteristics

	First screening mammogram					Subsequent screening mammograms				
	No.	Sensitivity, % (95% CI)*	False-positive rate, % (95% CI)	Likelihood ratio		No.	Sensitivity, % (95% CI)	False-positive rate, % (95% CI)	Likelihood ratio	
				Positive	Negative				Positive	Negative
Patient age†										
<40 y	51 494	84.6 (79.2 to 90.1)	15.0 (14.7 to 15.3)	5.64	0.18	18 454	52.7 (39.5 to 65.9)	10.5 (10.1 to 10.9)	5.0	0.53
40–49 y	131 272	81.6 (78.4 to 84.8)	13.5 (13.3 to 13.7)	6.05	0.21	246 198	68.6 (65.4 to 71.8)	9.2 (9.1 to 9.4)	7.4	0.35
50–59 y	69 150	82.5 (79.2 to 85.8)	12.5 (12.3 to 12.8)	6.60	0.20	278 559	75.2 (72.9 to 77.4)	8.4 (8.3 to 8.5)	8.9	0.27
60–69 y	43 162	85.6 (82.1 to 89.1)	11.3 (11.0 to 11.6)	7.58	0.16	178 278	77.1 (74.7 to 79.4)	7.6 (7.5 to 7.7)	10.1	0.25
≥70 y	41 038	87.9 (85.4 to 90.5)	9.7 (9.4 to 10.0)	9.07	0.13	162 441	79.7 (77.6 to 81.9)	6.5 (6.4 to 6.6)	12.3	0.22
Density‡										
Almost entirely fat	16 615	94.0 (89.0 to 99.1)	6.5 (6.1 to 6.9)	14.50	0.06	47 516	88.8 (83.3 to 94.3)	3.6 (3.4 to 3.7)	25.0	0.12
Scattered fibroglandular densities	73 156	90.1 (87.4 to 92.8)	12.1 (11.8 to 12.3)	7.46	0.11	243 996	82.3 (80.1 to 84.5)	7.5 (7.4 to 7.6)	11.0	0.19
Heterogeneously dense	56 936	82.0 (77.8 to 86.1)	12.5 (12.2 to 12.8)	6.55	0.22	180 201	72.4 (69.8 to 75.1)	8.8 (8.7 to 9.0)	8.2	0.30
Dense	19 708	77.8 (70.8 to 84.8)	14.7 (14.2 to 15.2)	5.29	0.26	45 643	65.9 (60.4 to 71.5)	10.5 (10.2 to 10.8)	6.3	0.38
Unknown	169 701	83.1 (81.0 to 85.2)	13.6 (8.5 to 13.7)	6.12	0.20	366 574	74 (72.2 to 75.9)	8.6 (8.5 to 8.7)	8.6	0.28

*CI = confidence interval.

†The point estimates changed little when calculated on the basis of a standardized distribution of breast density; therefore, the crude results are provided.

‡The point estimates changed little when calculated on the basis of a standardized distribution of patient age; therefore, the crude results are provided.

mean sensitivity for 95% of the physicians was between 48% and 95%, and the mean false-positive rate for 95% of the physicians was between 2% and 22%. Physicians with the highest false-positive rates tended to have the highest sensitivity, whereas physicians with the lowest false-positive rates tended to have the lowest sensitivity (Fig. 2). Thus, some of the difference among physician false-positive rates reflects their threshold for calling examinations abnormal (reflected as a tradeoff between sensitivity and specificity). However, some of the variation in sensitivity and specificity (and thus overall accuracy) was not the result of differences in threshold because at each false-positive rate, there was substantial variation in sensitivity between physicians. For example, at a false-positive rate of approximately 10%, the sensitivity ranged from 33% to 96%.

Sensitivity and Specificity of Mammography by Physician Characteristics

To identify physician characteristics that could explain the variation in physician accuracy, we first calculated physician sensitivity and specificity without adjusting for patient mixture. We found variations in the false-positive rates that paralleled physician experience (Fig. 3). In general, the false-positive rate declined (i.e., specificity improved) with increasing physician age, with increasing time since receipt of medical degree, and with increasing annual volume. For example, among subsequent screening mammograms (Fig. 3, B), the false-positive rate was 10.3% among physicians younger than age 40 years but only 6.8% among physicians aged 60–69 years. Additionally, physicians who had a higher focus on screening mammography than on diagnostic mammography had a lower false-positive rate (among subsequent examinations, 6.7% vs. 10.2%). Differences in sensitivity by physician experience were smaller and the confidence intervals largely overlapped, suggesting that the differences were not statistically significant.

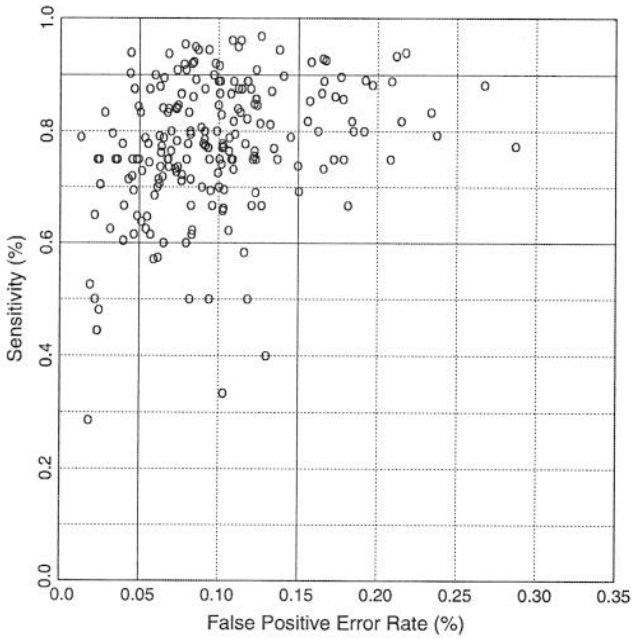


Fig. 2. Sensitivity versus the false-positive rate of screening mammography interpretation. Each physician is represented by a single point.

Combined Effects of Patient and Physician Characteristics on the Sensitivity and Specificity of Mammography

From the multivariable logistic regression analysis, several patient characteristics were associated with specificity (Table 3). A statistically significant increase in specificity was associated with an increase in patient age, with subsequent examinations, and with a breast density that was almost entirely fat. The following physician characteristics were also associated with a statistically significant increase in specificity: at least 25 years (versus less than 10 years) since receipt of medical degree (for physicians 25–29 years, OR = 1.54, 95% CI = 1.14 to 2.08), interpretation of 2500–4000 (versus 481–750) mammograms annually (OR = 1.30, 95% CI = 1.06 to 1.59), and a higher focus on screening mammography than on diagnostic mammography (OR = 1.59, 95% CI = 1.37 to 1.82). Interpretation of 1500–2500 mammograms was associated with a non–statistically significant improvement in specificity (OR = 1.16, 95% CI = 0.97 to 1.39).

Several patient characteristics were strongly associated with sensitivity. Increased sensitivity was associated with increased patient age, with first mammographic examinations, and with a breast density that was almost entirely fat or contained scattered fibroglandular densities. Physician characteristics were less consistently associated with sensitivity. A higher focus on screening mammography than on diagnostic mammography was associated with a lower sensitivity (OR = 0.82, 95% CI = 0.69 to 0.98), but sensitivity was not statistically significantly associated with a physician's annual volume or time since receipt of medical degree.

Overall accuracy is presented in Table 3. A statistically significant increase in overall accuracy was associated with a patient age older than 50 years and with breast density other than extremely dense. A statistically significant increase in overall accuracy was associated with 25–35 years since receipt of medical degree (e.g., for 25–29 years since receipt of their medical degree, OR for accuracy = 1.54, 95% CI = 1.05 to 2.26; P = .025). This result primarily reflects improved specificity (OR = 1.54, 95% CI = 1.14 to 2.08; P = .006) without a statistically significant change in sensitivity (OR = 1.0, 95% CI = 0.72 to 1.40; Table 3). A statistically significant increase in accuracy was also associated with a higher focus on screening mammography than on diagnostic mammography (OR = 1.29, 95% CI = 1.08 to 1.55), reflecting a statistically significant increase in specificity (OR = 1.59, 95% CI = 1.37 to 1.82) with a smaller reduction in sensitivity (OR = 0.82, 95% CI = 0.69 to 0.98). There was no statistically significant difference in accuracy as a function of physicians' annual volume (none of the groups was different than the lowest volume category), suggesting that the differences in specificity by annual volume largely reflect differences among physicians in their threshold for calling a mammogram abnormal. Interpretation of 751–1000 mammograms annually was associated with improved accuracy (OR = 1.33, 95% CI = 0.97 to 1.83), as characterized by small increases in both sensitivity (OR = 1.17, 95% CI = 0.87 to 1.56) and specificity (OR = 1.14, 95% CI = 0.93 to 1.41). However, this level of mammogram interpretation was not statistically significant (P = .08).

Association of Physician Experience with False-Positive Rates and Cancer Detection Rates

Physicians who had a higher focus on screening mammography than on diagnostic mammography or annual volume of

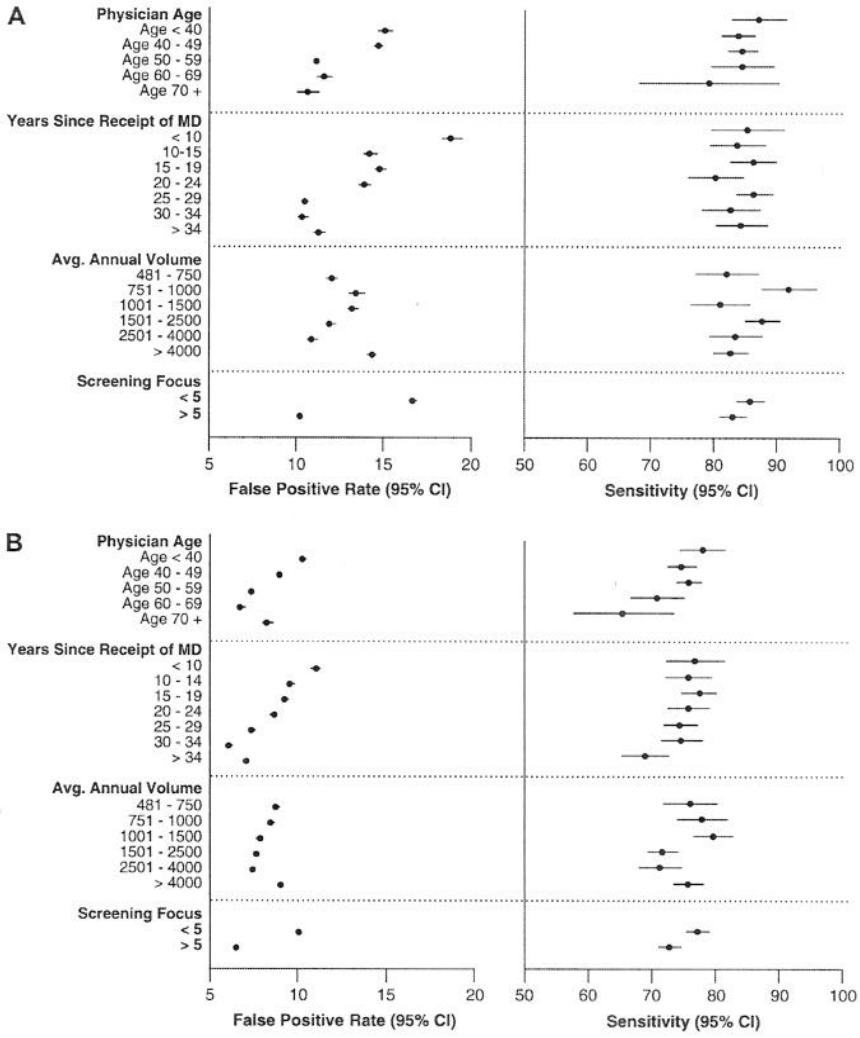


Fig. 3. False-positive rate and sensitivity (and 95% confidence intervals [CIs]) of screening mammography by physician characteristics for first (**A**) and subsequent (**B**) screening examinations. Error bars = 95% CIs. Some error bars are not visible because they do not extend beyond the symbol.

2500–4000 mammograms compared with 480–750 mammograms had lower false-positive rates. For physicians with a higher screening focus, this result reflects improved accuracy (defined as improved performance along a more accurate ROC curve). For physicians with a higher volume, this result reflects a shift along a ROC curve to operate in an area that emphasizes improved specificity. The difference in how these physicians perform will substantially affect the patients whose mammograms they interpret. Compared with physicians who interpret the minimum number of mammograms annually (i.e., 481–750 mammograms) and had a low screening focus (ratio less than 5), physicians who interpret 2500–4000 mammograms annually and had a high screening focus (ratio greater than or equal to 5) had approximately 50% fewer false-positive examinations (674 versus 1279 false-positive examinations per 10 000 screening examinations) and detected only a few less cancers (44 versus 47 per 10 000 screening examinations) (Table 4). Thus, a physician who interprets 3000 mammograms annually and has a high focus on screening mammography would have approximately 182 fewer false-positive examinations and would miss approximately one cancer per year, compared with a low-volume physician who does not focus to the same degree on screening mammography. A physician who interprets 1500–2500 mammograms annually and has a high focus on screening mammography would have approximately 40% fewer false-positive examinations and miss approximately one cancer per 5000 screening examinations, compared with the low-volume physician who does not focus to the same degree on screening mammography. These differences in sensitivity and specificity are reflected by the positive predictive value of mammography, which is nearly twice as high as in the high-volume, high-screening-focus category as in the low-volume, low-screening-focus category (6.1% vs. 3.6%).

Discussion

We found substantial physician variation in mammographic sensitivity and specificity that was not explained by the characteristics of patients whose mammograms they interpreted. The most dramatic difference was in the false-positive rate, which varied from 1% to 29%. In general, the most experienced physicians had the lowest false-positive rates. Physicians who had been practicing the longest, who interpreted 2500–4000 mammograms annually, and who emphasized screening, as opposed to diagnostic, mammography had lower false-positive rates than their less-experienced counterparts. For physicians who had practiced the longest and who had a high focus on screening mammography, overall accuracy was improved as well, meaning that they had higher specificity without an equal loss in sensitivity. For physicians

Table 3. Influence of patient and physician characteristics on the odds of a correct mammogram interpretation in women with and without breast cancer simultaneously adjusting for threshold*

	Specificity		Sensitivity		Accuracy†	
	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*
Patient age						
<40 y	0.90 (0.83 to 0.98)	<.012	0.94 (0.61 to 1.4)	.770	0.85 (0.56 to 1.28)	.428
40–49 y	1.0 (referent)		1.0 (referent)		1.0 (referent)	
50–59 y	**1.08** (1.04 to 1.10)	<.001	**1.25** (1.06 to 1.47)	.007	**1.34** (1.14 to 1.58)	<.001
60–69 y	**1.16** (1.11 to 1.20)	<.001	**1.36** (1.12 to 1.64)	.002	**1.57** (1.29 to 1.91)	<.001
≥70 y	**1.32** (1.25 to 1.41)	<.001	**1.51** (1.26 to 1.81)	<.001	**2.00** (1.66 to 2.40)	<.001
Screening						
First	1.0 (referent)		1.0 (referent)		1.0 (referent)	
Subsequent	**1.59** (1.49 to 1.67)	<.001	**0.52** (0.45 to 0.61)	<.001	**0.82** (0.70 to 0.98)	.024
Density						
Almost entirely fat	**2.38** (1.67 to 3.33)	<.001	**3.98** (2.20 to 7.17)	<.001	**9.37** (5.07 to 17.32)	<.001
Scattered fibroglandular densities	1.19 (0.91 to 1.54)	.200	**2.19** (1.64 to 2.93)	<.001	**2.60** (1.98 to 3.45)	<.001
Heterogeneously dense	1.05 (0.81 to 1.37)	.704	1.34 (1.00 to 1.79)	0.046	**1.41** (1.05 to 1.90)	0.024
Extremely dense	1.0 (referent)		1.0 (referent)		1.0 (referent)	
Unknown	0.91 (0.68 to 1.20)	.500	**1.48** (1.07 to 2.04)	.017	1.34 (0.99 to 1.82)	0.058
Time since receipt of medical degree						
<10 y	1.0 (referent)		1.0 (referent)		1.0 (referent)	
10–14 y	1.16 (0.88 to 1.54)	.282	0.98 (0.68 to 1.43)	.921	1.14 (0.73 to 1.78)	.564
15–19 y	1.22 (0.92 to 1.64)	.172	1.07 (0.79 to 1.46)	.654	1.31 (0.92 to 1.87)	.135
20–24 y	1.18 (0.88 to 1.59)	.276	0.96 (0.70 to 1.33)	.817	1.14 (0.78 to 1.64)	.501
25–29 y	**1.54** (1.14 to 2.08)	.006	1.00 (0.72 to 1.40)	.999	**1.54** (1.05 to 2.26)	.025
30–34 y	**1.67** (1.25 to 2.22)	<.001	0.86 (0.63 to 1.19)	.367	**1.44** (0.99 to 2.12)	.060
>34 y	**1.59** (1.12 to 2.22)	.008	0.76 (0.55 to 1.04)	.084	1.20 (0.82 to 1.73)	.347
Average annual volume mammogram interpretation						
481–750 mammograms	1.0 (referent)		1.0 (referent)		1.0 (referent)	
751–1000 mammograms	1.14 (0.93 to 1.41)	.216	1.17 (0.87 to 1.56)	.292	1.33 (0.97 to 1.83)	.080
1001–1500 mammograms	1.05 (0.85 to 1.30)	.657	1.07 (0.80 to 1.44)	.643	1.13 (0.87 to 1.46)	.373
1501–2500 mammograms	1.16 (0.97 to 1.39)	.092	0.91 (0.72 to 1.15)	.449	1.06 (0.86 to 1.32)	.571
2501–4000 mammograms	**1.30** (1.06 to 1.59)	.011	0.83 (0.63 to 1.10)	.197	1.08 (0.82 to 1.42)	.586
>4000 mammograms	1.03 (0.85 to 1.25)	.789	0.96 (0.74 to 1.23)	.719	0.98 (0.77 to 1.25)	.878
Ratio of screening to diagnostic mammographic interpretation						
<5	1.0 (referent)		1.0 (referent)		1.0 (referent)	
>5	**1.59** (1.37 to 1.82)	<.001	**0.82** (0.69 to 0.98)	.026	**1.29** (1.08 to 1.55)	.005

*P values correspond with the odds ratio (OR) to the left. CI = confidence interval. Statistically significant ORs (P<.05) are shown in boldface type.

†Improved sensitivity at given specificity or improved specificity at given sensitivity.

who interpreted at a high volume (2500–4000 mammograms annually), the difference in performance reflected a shift in the threshold used by these physicians to interpret an examination as abnormal (thus, a shift along an ROC curve). The differences in sensitivity were of much smaller magnitude than the differences in the false-positive rate; consequently, the higher-volume physicians did not miss many cancers even with the higher threshold they used to interpret an examination as abnormal (approximately one cancer per year).

Our results have important implications for the practice of screening mammography. We estimated that, compared with physicians who interpreted the minimum number allowed by

Table 4. Estimated differences in patient outcomes stratified by physician differences in screening mammography*

Annual volume: No. mammograms interpreted	Focus on screening	Sensitivity, %	Specificity, %	False-positive rate, %	No. cancers detected	No. false-positive diagnoses	PPV
480–750	Low	80.8	87.1	12.9	47	1279	3.6
	High	77.7	91.4	8.6	45	855	5.0
750–1000	Low	83.0	88.5	11.5	49	1143	4.1
	High	80.1	92.4	7.6	47	759	5.8
1000–1500	Low	81.8	87.7	12.3	48	1226	3.8
	High	78.8	91.8	8.2	46	818	5.3
1500–2500	Low	79.4	88.7	11.3	46	1121	4.0
	High	76.1	92.5	7.5	45	744	5.6
2500–4000	Low	77.9	89.7	10.3	46	1020	4.3
	High	74.4	93.2	6.8	44	674	6.1
>4000	Low	80.1	87.4	12.6	47	1250	3.6
	High	76.9	91.6	8.4	45	834	5.1

*Estimates assume that 10000 women underwent screening mammography, that the multivariable distribution of patient characteristics, and that the total number of cancers (5.9 per 1000 mammograms) was the same as it is in this cohort. PPV = positive predictive value.

Mammography Quality Standards Act (i.e., 480–750 mammograms per year) and who have a lower screening focus, physicians who interpret 2500–4000 mammograms annually and have a higher screening focus have 50% fewer false-positive diagnoses (168 vs. 320 per 2500 examinations) and miss approximately one cancer per 2500 mammograms interpreted. We found that physicians with a higher screening focus have substantially improved specificity, slightly lower sensitivity, and overall improved accuracy. Our results indicate that physicians who focus on screening are better at screening than those who do not. One possible explanation is that physicians who have a larger proportion of diagnostic examinations (i.e., a low screening focus) may expect higher underlying rates of cancer, which might lead them to recall a larger percentage of patients.

There is considerable debate over how to analyze data describing the accuracy of diagnostic testing. Although ROC analyses have been a mainstay of diagnostic imaging research, there are several limitations of this method for evaluating the accuracy of mammography. ROC curve analysis cannot be used to understand the actual sensitivity and specificity in clinical practice *(35)*, and some ROC analyses, such as those that rely on the area under the curve, assume that every location along an ROC curve is equivalent. For example, if physician a has a sensitivity of 20% and a false-positive rate of 1%, physician b has a sensitivity of 85% and a false-positive rate of 5%, and physician c has a sensitivity of 90% and a false-positive rate of 30% (Fig. 1), all physicians can be said to perform along a single ROC curve, with each physician using a different threshold to interpret mammograms as abnormal. Although the performance of all three physicians can be plotted on the same ROC curve, it is not the case that each point along the curve reflects equally desirable performance. Yet area under the ROC curve analysis would not detect differences between these physicians. Specificity will tend to impact many more individuals than sensitivity. Thus, for physician c, the slightly higher sensitivity needs to be weighed against the substantially higher false-positive rate, and the performances of physicians b and c should not be considered comparable. Lastly, in some instances, a clinically relevant improvement in test accuracy (such as an improvement in sensitivity with only a small change in specificity) may not be regarded as an improvement via a ROC curve analysis, if the curve appears relatively steep in that region so that both points fall along the same curve *(35)*. Thus, we used the calculated sensitivity and specificity of each physician as the important outcome, because they are clinically relevant and easily understood. We used ROC curve analysis to determine whether the differences we detected were caused by threshold differences between physicians. We have identified physician characteristics that are associated with accuracy (time since receipt of medical degree and a high focus on screening mammography), as well as physician characteristics that are associated with a shift along an ROC curve (high annual volume).

Our results are consistent with those of previous studies *(12,13)* that used practice sets and found that more experienced physicians have lower false-positive rates. Our findings are in contrast with those of Beam et al. *(15)* who used a practice set and found that the most recently trained physicians perform better and that annual volume is not an important predictor of accuracy. In that study, physicians' performance on the practice set differed dramatically from what we found in our study using actual clinical mammograms. The mean sensitivity of mammography was 90% in the Beam study (versus 77% with actual clinical mammograms in our study), and the mean false-positive rate was 38% (versus 10% with actual clinical mammograms in this study). Thus, mammogram interpretation in routine clinical practice appears to differ substantially from that the testing situation described in the Beam study *(15)* in which the high proportion of cancers probably lowers the threshold for interpreting examinations as abnormal *(1,16,17)*. Additionally, the Beam study's nonstandard analysis method (each mammogram, via its BI-RADS score, contributed several estimates to each physician's accuracy) could also account for the differing results. Lastly, given the ROC method used in the Beam study, the authors could not differentiate physicians who performed on the same ROC curve—i.e., who differed in characteristics that influenced the threshold but not the accuracy.

Our results support the three studies of mammographic accuracy and volume that used prospectively interpreted clinical data. Sickles et al. *(19)* demonstrated that three physicians with special training in mammography had lower false-positive rates and higher cancer detection rates than seven general physicians who each interpreted only sufficient numbers of mammograms to satisfy federal regulations. Kan et al. *(18)* demonstrated that the physicians in British Columbia, each of whom interpreted 2000–4000 mammograms annually, had lower false-positive rates than physicians who interpreted less than 2000 mammograms annually or more than 4000 annually. Théberge et al. (36) demonstrated that radiologists who read more than 1500 mammograms annually had higher breast cancer detection rates while maintaining lower false-positive rates. Our finding of improved specificity among more experience physicians agree with those of Barlow et al. *(37)*. Whereas we found that experienced physicians were also more accurate, they found that experienced physicians tended to increase the threshold they used to consider a mammogram abnormal without improved accuracy. Our results also differed with respect to annual volume. Paralleling the other measures of experience, we found that increased volume (up to 4000 mammograms per year) is associated with improved specificity, whereas Barlow et al. found that increased volume is associated with worse specificity but improved sensitivity. There are several differences in our research methods that may account for these differences. First, Barlow et al. used physician's self-reported annual volume, rather than actual volume, and physicians may have incorrectly estimated their annual volume. The physicians in the study of Barlow et al. reported reading many more mammograms than we found; 25% of physicians read fewer than 1000 mammograms annually in Barlow's study compared with 45% in our study. Similarly, whereas 37% of physicians in the Barlow study reported having read more than 2000 mammograms annually, we found only half as many physicians (21%) read at such high volumes. Although we may have underestimated annual volume for physicians who interpret mammograms at facilities that do not participate in the Breast Cancer Surveillance Consortium, we believe that this would have had only limited impact on overall estimates of annual volume. Three of the four registries that we included (Vermont, San Francisco, and New Mexico) have almost complete population-based capture of mammograms, and thus we almost certainly captured the majority of mammograms for those physicians in the study. Second, Barlow et al. used broad categories to characterize physician annual volume, combining all physicians with annual volumes of more than 2000 into a single category. We found, as have others *(18)*, that specificity improves as volume increases up to 4000

mammograms annually but that physicians with volumes of more than 4000 have worse specificity. Combining all physicians with volumes of more than 2000 mammograms annually could have masked trends. Additionally, volume was assessed in only a single year in the Barlow study, whereas we averaged physician volume over 4 years, to account for variability across the years. Lastly, Barlow et al. used ROC methodology similar to that used by Beam et al. *(15)*, in which the full range of BI-RADS assessments are analyzed by use of an ordinal regression model rather than by dichotomizing the interpretation as normal or abnormal as occurs in clinical practice. Surprisingly, by use of this ROC methodology, Barlow et al. found that patient age does not impact the accuracy of mammography, which contrasts with our work and the work of many others *(1)*. These unexpected results raise questions about the ROC results of that study.

Our study demonstrated that annual mammographic volume, time since receipt of medical degree, and a focus on screening mammography are important contributors to mammographic accuracy. However, these factors did not explain all of the variation in physician performance. Many other factors potentially contribute to mammographic accuracy, such as whether physicians regularly assess their outcomes (learn from their mistakes), which types of ongoing medical education they complete, and perhaps whether they have concerns about medical malpractice.

We recommend that there be explicit discussion of what the goals of mammography should be. Should physicians maximize sensitivity at the expense of having very high false-positive rates or should they maximize sensitivity while achieving a lower, but reasonable, false-positive rate? Some of the large variation that we found among physicians may reflect differences in their individual expectations about ideal mammography performance (with some physicians choosing to emphasize sensitivity at the expense of very high false-positive rates). If the goal is to maximize sensitivity while achieving a reasonable false-positive rate, one action could be to raise the minimum number of mammograms physicians must interpret annually. An argument against raising the minimum is that this approach would decrease the supply of physicians who can interpret mammograms. Our data, however, suggest that the impact would be small if the minimum level is raised moderately. For example, if the minimum level is raised to 750 mammograms annually, although 30% fewer physicians would interpret mammograms, only 10% more screening mammograms would have to be interpreted by the remaining higher-volume physicians. Although an annual volume of 2500 mammograms seems ideal from a performance perspective if minimizing the false-positive rate were a goal, this change would need to occur slowly to prevent a shortage of physicians who interpret mammograms. A compromise of 1500 mammograms might be a practical solution because it would probably lead to a substantial reduction in the false-positive rate (40% in our estimate) yet would not create as much of a burden on the remaining higher-volume physicians.

A strength of our study is that the data were collected from actual clinical practice in four geographic areas across the United States and that 95% of physicians in those areas who practice at facilities that participate in the Breast Cancer Surveillance Consortium were included in this analysis. A limitation of our study is that we do not know whether greater experience, higher annual volume, and a greater focus on screening mammography improve interpretations or whether the better physicians simply choose to interpret more examinations. That is, it is not possible to disentangle what is cause and what is effect. Nonetheless, physicians who are interpreting more screening mammograms are doing a better job. Another limitation is sample size; although our sample size was large, it was not large enough to look separately at ductal carcinoma in situ and invasive cancer.

Although some variation in physician performance is inevitable, the degree of variation that we found, particularly for the false-positive rates, is large. Consequently, finding and implementing interventions to minimize this variation should be a priority. The false-positive rate in the United States is higher than that in other countries *(38)*, and it is twice as high as the rate in the United Kingdom *(5)*, although cancer detection rates are similar in the two countries. One of the major factors producing these differences in rates between the United States and the United Kingdom could be the annual volume of mammograms interpreted by physicians. The median annual number of mammograms that physicians interpreted in our sample (1053 mammograms) contrasts starkly with the median annual number of mammograms that physicians interpret in the United Kingdom (7000 mammograms) *(21)*. In the United States, the minimum value required by the Mammography Quality Standards Act is very low, approximately two mammograms per clinical workday, and the mean is fewer than five mammograms per clinical workday. Most factors that influence the sensitivity of mammography are not easily modified, e.g., a woman's age, mammographic breast density, and a physician's years of experience. Physician volume and screening focus can be altered, particularly because the Mammography Quality Standards Act is actively involved in the monitoring of physician volume. Raising the annual volume requirements in the Mammography Quality Standards Act might improve the overall quality of screening mammography in the United States.

References

(1) Carney P, Miglioretti D, Yankaskas B, Kerlikowske K, Rosenberg R, Rutter C, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the performance of screening mammography. Ann Intern Med 2003;138:168–73.

(2) 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.

(3) May D, Lee N, Nadel M, Henson RM, Miller DS. The National Breast and Cervical Cancer Early Detection Program: report on the first 4 years of mammography provided to medically underserved women. AJR Am J Roentgenol 1998;170:97–104.

(4) Sickles E, Ominsky S, Sollitto R, Galvin HB, Monticciolo DL. Medical audit of a rapid-throughput mammography screening practice: methodology and results of 27,114 examinations. Radiology 1990;175:323–7.

(5) Smith-Bindman R, Chu PW, Miglioretti DL, Sickles EA, Blanks R, Ballard-Barbash R, et al. Comparison of screening mammography in the United States and the United Kingdom. JAMA 2003;290:2129–37.

(6) Elmore JG, Wells CK, Howard DH, Feinstein AR. The impact of clinical history on mammographic interpretations. JAMA 1997;277:49–52.

(7) Kerlikowske K, Grady D, Barclay J, Sickles ES, Ernster V. Likelihood ratios for modern screening mammography. Risk of breast cancer based on age and mammographic interpretation. JAMA 1996;276:39–43.

(8) Hannan EL, Racz M, Ryan TJ, McCallister BD, Johnson LW, Arani DT, et al. Coronary angioplasty volume-outcome relationships for hospitals and cardiologists. JAMA 1997;277:892–8.

(9) Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. Ann Intern Med 2002;137:511–20.

(10) Linver MN, Paster SB, Rosenberg RD, Key CR, Stidley CA, King WV. Improvement in mammography interpretation skills in a community radiology practice after dedicated teaching courses: 2-year medical audit of 38,633 cases [erratum ain Radiology 1992;184:878]. Radiology 1992;184: 39–43.
(11) Elmore JG, Miglioretti DL, Reisch LM, Barton MB, Kreuter W, Christiansen CL, et al. Screening mammograms by community radiologists: variability in false-positive rates. J Natl Cancer Inst 2002;94:1373–80.
(12) Elmore JG, Wells CK, Howard DH. Does diagnostic accuracy in mammography depend on radiologists' experience? J Women's Health 1998;7: 443–9.
(13) Esserman L, Cowley H, Eberle C, Kirkpartrick A, Change S, Berbaum K, et al. Improving the accuracy of mammography: volume and outcome relationships. J Natl Cancer Inst 2002; 94.
(14) Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists. Findings from a national sample. Arch Intern Med 1996;156:209–13.
(15) Beam C, Conant E, Sickles E. Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. J Natl Cancer Inst 2003;95:282–90.
(16) Egglin T, Feinstein A. Context bias. A problem in diagnostic radiology. JAMA 1996;276:1752–5.
(17) Elmore JG, Miglioretti DL, Carney PA. Does practice make perfect when interpreting mammography? Part II. J Natl Cancer Inst 2003;95:250–2.
(18) 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;215:563–7.
(19) Sickles E, Wolverton D, Dee K. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. Radiology 2002;224:861–9.
(20) The Mammography Quality Standards Act of 1992, Pub. L. No. 102-539.
(21) Department of Health, U.K., Statistical Bulletin, Breast Screening Programme, England: 1999–2000. National Statistics; March 2000.
(22) National Mammography Quality Assurance Advisory Committee. Summary minutes, Washington, DC, January 23–25, 1995.
(23) Ballard-Barbash R, Taplin SH, Yankaskas BC, Ernster VL, Rosenberg RD, Carney PA, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. AJR Am J Roentgenol 1997;169:1001–8.
(24) Barlow W, Lehman C, Zheng Y, Ballard-Barbash R, Yankaskas BC, Cutter GR, et al. Performance of diagnostic mammography for women with signs or symptoms of breast cancer. J Natl Cancer Inst 2002;94:1151–9.
(25) Breast Cancer Surveillance Consortium Web Site. http://www.breastscreening.cancer.gov/elements.html#questionnaires. [Last accessed: February 2, 2005.]
(26) Ernster VL, Ballard-Barbash R, Barlow WE, Zheng Y, Weaver DL, Cutter G, et al. Detection of ductal carcinoma in situ in women undergoing screening mammography. J Natl Cancer Inst 2002;94:1546–54.
(27) Kerlikowske K, Carney PA, Geller B, Mandelson MT, Taplin SH, Malvin K, et al. Performance of screening mammography among women with and without a first-degree relative with breast cancer. Ann Intern Med 2000;133:855–63.
(28) Kerlikowske K, Miglioretti DL, Ballard-Barbash R, Weaver DL, Buist DS, Barlow WE, et al. Prognostic characteristics of breast cancer among postmenopausal hormone users in a screened population. J Clin Oncol 2003;21:4314–21.
(29) Miglioretti DL, Rutter CM, Geller BM, Cutter G, Barlow WE, Rosenberg R, et al. Effect of breast augmentation on the accuracy of mammography and cancer characteristics. JAMA 2004;291:442–50.
(30) National Cancer Institute Monograph: BCSC report: evaluating screening performance in practice. Available at: http://breastscreening.cancer.gov/espp.pdf. [Last accessed: February 2, 2005.]
(31) American Medical Association CoSA. http://www.ama-assn.org/ama/pub/category/12850.html. [Last accessed: February 2, 2005.]
(32) American College of Radiology. Breast imaging reporting and data system (BI-RADS). 3rd ed. Reston (VA): The College; 1998.
(33) Miglioretti DL, Heagerty PJ. Marginal modeling of multilevel binary data with time-varying covariates. Biostatistics 2004;5:381–98.
(34) Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13–22.
(35) Pepe MS, Urban N, Rutter C, Longton G. Design of a study to improve accuracy in reading mammograms. J Clin Epidemiol 1997;50:1327–38.
(36) Théberge I, Hébert-Croteau N, Langlois A, Major D, Brisson J. Volume of screening mammography and performance in the Quebec population-based Breast Cancer Screening Program. CMAJ 2005;172:195–199.
(37) Barlow WE, Chi C, Carney PA, Taplin SH, D'Orsi C, Cutter G, et al. Accuracy of screening mammography interpretation by characteristics of radiologists. J Natl Cancer Inst 2004:24;1840–50.
(38) Elmore JG, Yakano CY, Koepsell TD, et al. International variation in screening mammography Interpretations in community-based programs. J Natl Cancer Inst 2003;95:1384–93.

Notes

[1]*Editor's note:* SEER is a set of geographically defined, population-based, central cancer registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Registry data are submitted electronically without personal identifiers to the NCI on a biannual basis, and the NCI makes the data available to the public for scientific research.

This work was supported in part by the National Cancer Institute (CA86032 and Breast Cancer Surveillance Consortium cooperative agreements U01CA63740, U01CA86076, U01CA63736, U01CA70013 and U01CA69976) and The Department of Defense (DAMD179919112 and DAMD170010193).

Manuscript received August 6, 2004; revised November 4, 2004; accepted January 11, 2005.

ORIGINAL PAPER

Comparing the performance of mammography screening in the USA and the UK

Rebecca Smith-Bindman, Rachel Ballard-Barbash, Diana L Miglioretti, Julietta Patnick, Karla Kerlikowske

J Med Screen 2005;**12**:50–54

To compare the performance of screening mammography in the USA and the UK, a consecutive sample of screening mammograms was obtained in women aged 50 and older from 1996 to 1999 who participated in the Breast Cancer Surveillance Consortium in the USA ($n = 978,591$) and the National Health Service Breast Cancer Screening Program in the UK ($n = 3.94$ million), including 6943 diagnosed with breast cancer within 12 months of screening. Recall rates were defined as the percentage of screening mammograms with a recommendation for further evaluation including diagnostic mammography, ultrasound, clinical examination or biopsy, and cancer detection rates including invasive cancer and ductal carcinoma *in situ* diagnosed within 12 months of a screening mammogram. All results were stratified by whether examinations were first or subsequent and adjusted to a standard age distribution. Among women who underwent a first screening mammogram, 13.3% of women in the USA versus 7.2% of women in the UK were recalled for further evaluation (relative risk for recall 1.9; 95% CI 1.8–1.9). For subsequent examinations recall rates were approximately 50% lower, but remained twice as high in the USA as in the UK. A similar percentage of women underwent biopsy in each setting, but rates of percutaneous biopsy were lower and rates of open surgical biopsy were higher in the USA. Women undergo screening approximately every 18 months in the USA and every 36 months in the UK. Based on a 20-year period of screening, the estimated percentage of women who would be recalled for additional testing was nearly threefold higher in the USA. The number of cancers detected was also higher in the USA (55 versus 43), and most of the increase was in the detection of small invasive and *in situ* cancers. The numbers of large cancers detected (>2 cm) were very similar between the two countries. Recall rates are approximately two to three times higher in the USA than in the UK. Importantly, despite less frequent screening in the USA, there are no substantial differences in the rates of detection of large cancers. Efforts to improve mammographic screening in the USA should target lowering the recall rate without reducing the cancer detection rate.

See end of article for authors' affiliations

Correspondence to: Rebecca Smith-Bindman, Department of Radiology, University of California San Francisco, UCSF/Mt. Zion Campus, 1600 Divisadero Street, San Francisco, Box 1667, CA 94115, USA; Rebecca.Smith-Bindman @Radiology.UCSF.edu

Accepted for publication 10 November 2004

INTRODUCTION

The provision of screening mammography differs greatly between the USA and the UK. In the USA, screening is provided in diverse settings where women self-refer or are referred by their medical provider.[1] In the UK, a single organized screening programme run by the National Health Service (NHS) provides virtually all mammographic screening for women aged 50 years or older, and women of eligible age are invited through population-based databases to undergo screening.[2,3] As a result of differences in the organization of mammography between the two countries, there are differences in the ages of women screened, the intervals between mammographic examinations, the proportion of women recalled for additional imaging examinations, such as diagnostic mammography, and the methods used to further evaluate examinations considered suspicious for cancer.[4–7] Differences in the performance of mammography between the USA and other countries have been reported.[8] However, it is not clear if differences in performance are a reflection of differences in the age distribution and frequency of screening between the two populations or are due to differences in the practice of mammography. We have compared mammography between the USA and UK and the results have recently been reported in the *Journal of the American Medical Association*.[7] The results are summarized with additional analyses herein.

METHODS

Data sources

Data on screening mammography were obtained from the Breast Cancer Surveillance Consortium (BCSC) in the USA,[9] and from the National Health Service Breast Screening Program (NHSBSP) in the UK.[3,10] The BCSC is a National Cancer Institute (NCI)-funded consortium of mammography registries in San Francisco (California); Colorado; New Hampshire; New Mexico; North Carolina; Western Washington; and Vermont.[9,11] The registries that participate in the BCSC are not organized breast screening programmes (with the exception of the Western Washington registry based on the programme at the Group Health Cooperative), but have been funded to collect data on mammography performance from diverse community-based settings across the USA. The consortium of registries probably reflects the current practice of mammography in the USA.[12–14] The government-funded NHSBSP provides free breast cancer

screening in the UK for women 50 years or older.[3,10] Women aged 50–64 years are invited by letter – with an information leaflet – to attend breast screening every three years through a system that relies on centralized computer databases. By 1995, the NHSBSP had achieved national coverage.

Results of all screening mammograms in women aged 50 years or older conducted within each setting between 1 January 1996 and 31 December 1999 were included in this report. The study was approved by the University of California, San Francisco Institutional Review Board, and each of the BCSC registries has approval from their respective IRBs to collect data for research purposes.

Mammogram interpretation

A mammogram was classified as positive (recall) if it was interpreted as abnormal (BI-RADS 0, 3, 4, 5)[15] and a recommendation for prompt diagnostic imaging, clinical evaluation, or biopsy was given. Women were considered to have breast cancer detected if active-case follow-up or reports from a pathology database or tumour registry showed invasive carcinoma or ductal carcinoma *in situ* (DCIS) within 12 months of a positive screening mammogram. Cancers that occurred within 12 months of a normal interpretation (false-negative examinations) were not included in this analysis because they are not routinely collected within the NHSBSP. As recall and cancer detection rates vary by age and whether women have undergone previous mammography,[16–18] all analyses were subdivided into first or subsequent screening examination and five-year age groups.

Analysis

Recall, non-invasive work-up (ultrasound, diagnostic mammography or other tests, or clinical breast examination) and biopsy rates were calculated per 100 screening mammographic examinations, stratified by whether the exam was first or subsequent and adjusted to a standard age distribution. The age distributions of the two data sources were weighed equally to create the standard age distribution. Each mammogram was counted once when calculating the non-invasive work-up rate and biopsy rate, even if more than one test or biopsy was recommended. The percutaneous biopsy rate was calculated as the number of fine-needle aspirations or core biopsies, and the open surgical biopsy rate was calculated as the number of open surgical biopsies per 100 mammographic examinations. The open surgical biopsy rate was subdivided into those that resulted in a diagnosis of cancer (positive open surgical biopsy rate) and those that did not result in a diagnosis of cancer (negative open surgical biopsy rate). The cancer detection rate was calculated as the number of breast cancers detected per 1000 examinations. The Cochran–Mantel–Haenszel method was used to calculate the relative risk (RR) and 95% confidence interval (CI) of each outcome (recall and cancer detection) in the USA compared with the UK.

Since mammography screening is performed more frequently in the USA, one would expect fewer cancers to be diagnosed per subsequent screening examination in the USA. To compare cancer detection rates for a similar screening period, we used a simulation model to estimate the numbers of cancers detected and women recalled per 1000 women undergoing screening mammography over a 20-year period, assuming a screening interval of 18 months for the USA and 36 months for the UK. We simulated one million women beginning screening at age 50 and followed these women for 20 years of screening at the intervals specified above. We used our estimated age-specific recall rates and cancer detection rates based on four years of data, and we assumed that the likelihood of recall was independent from one exam to the next and a woman could have cancer detected only once over the 18 years of screening. The method of this modelling was slightly different from that we have used previously. Specifically, we only included invasive cancers where the size was known.[7] SAS version 8.2 (SAS Institute Inc, Cary, NC, USA) was used for all statistical analyses.

RESULTS

This analysis included nearly five million mammograms: 978,591 from the BCSC and 3.94 million from the NHSBSP, which led to the diagnosis of 6943 cases of breast cancer among women aged 50 and older.[7] Recall rates were approximately twice as high in the USA for first as well as subsequent examinations. Among the first screening mammograms, 13.3% of women in the USA versus 7.2% of women in the UK were recalled for further evaluation, including diagnostic mammography, ultrasound, clinical examination or biopsy (RR 1.9, 95% CI 1.8–1.9; Table 1). Recall rates were lower for subsequent examinations but remained twice as high in the USA: 8.0% versus 3.6% in the USA and UK, respectively (RR 2.2, 95% CI 2.2–2.3). Biopsy rates were nearly identical between the two countries. Approximately 2.5% of first and 1% of subsequent screening examinations were followed by a recommendation for biopsy. Thus, the higher recall rate in the USA was primarily because of a higher rate of diagnostic imaging (11.3% versus 4.5% for first examinations [RR 2.5, 95% CI 2.4–2.6] and 6.9% versus 2.6% for subsequent examinations [RR 2.7, 95% CI 2.7–2.7] in the USA and UK, respectively). Although the total biopsy rates were similar in the two countries, biopsies were more likely to be open surgical biopsies in the USA (1.15% versus 0.72% for first and 0.33% versus 0.29% for subsequent exams; Table 2). Most of the difference in open surgical biopsy rates was among women who did not have breast cancer; negative open surgical biopsies were approximately twice as high in the USA (first examinations RR 2.17, 95% CI 1.94–2.42; subsequent examinations RR 2.17, 95% CI 2.03–2.32).

The cancer detection rates increased with age (data not shown) and were two to three times higher for the first – compared with subsequent – mammograms in both countries (Table 3). Despite substantially higher recall rates in the

Table 1 Recall, non-invasive diagnostic testing and biopsy rates (and 95% CI) per 100 first and subsequent screening mammograms by setting and screening cycle

	First screening mammogram			Subsequent screening mammogram		
	USA (95% CI)	UK (95% CI)	RR% (95% CI)	USA (95% CI)	UK (95% CI)	RR% (95% CI)
Recall	13.3 (13.1–13.6)	7.2 (7.0–7.3)	1.9 (1.8–1.9)	8.0 (7.9–8.1)	3.6 (3.6–3.7)	2.2 (2.2–2.3)
Non-invasive diagnostic tests	11.3 (11.1–11.6)	4.5 (4.4–4.6)	2.5 (2.4–2.6)	6.9 (6.9–7.0)	2.6 (2.6–2.6)	2.7 (2.7–2.7)
Biopsy	2.4 (2.3–2.6)	2.5 (2.5 2.6)	0.96 (0.91–1.02)	0.99 (0.97–1.0)	0.96 (0.94–0.98)	1.1 (1.0–1.1)

Relative risk of recall, non-invasive diagnostic testing, and biopsy rates (and 95% CI) are shown comparing the USA with the UK.[7] Adapted from Table 2, Smith-Bindman *et al.*, 2003.[7]

Table 2 Recommended open surgical biopsy rates per 100 screening mammograms, by setting and screening cycle[7]

	First screening mammogram			Subsequent screening mammogram		
	USA (95% CI)	UK (95% CI)	RR	USA (95% CI)	UK (95% CI)	RR% (95% CI)
Open surgical biopsy	1.15 (1.1–1.2)	0.72 (0.67–0.77)	1.62 (1.48–1.77)	0.33 (0.32–0.35)	0.28 (0.27–0.29)	1.16 (1.11–1.22)
Positive open surgical biopsy*	0.31 (0.27–0.36)	0.36 (0.32–0.39)	0.94 (0.80–1.11)	0.11 (0.11–0.12)	0.18 (0.17–0.19)	0.60 (0.55–0.65)
Negative open surgical biopsy†	0.82 (0.75–0.89)	0.36 (0.33–0.39)	2.17 (1.94–2.42)	0.22 (0.21–0.23)	0.10 (0.10–0.11)	2.17 (2.03–2.32)

The positive and negative open surgical biopsy rate may not sum due to rounding. The specific method of biopsy could be determined for four of the seven BCSC sites and these data were used to determine the method of biopsy in the USA.
*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.
Adapted from Table 3, Smith-Bindman *et al.*, 2003.[7]

Table 3 Cancers detected per 1000 screening mammograms, by setting and screening cycle[7]

	First screening mammogram		Subsequent screening mammogram	
	USA (95% CI)	UK (95% CI)	USA (95% CI)	UK (95% CI)
Cancer	8.6 (7.9–9.4)	10.1 (9.4–10.7)	3.6 (3.5–3.7)	5.4 (5.2–5.5)
Invasive	7.2 (6.5–7.8)	8.4 (7.8–9.0)	2.8 (2.7–2.9)	4.3 (4.2–4.5)
In situ	1.5 (1.2–1.8)	1.6 (1.4–1.9)	0.83 (0.77–0.90)	0.99 (0.92–1.1)

Adapted from Table 4, Smith-Bindman *et al.*, 2003.[7]

Table 4 Estimated number of women with at least one recalled examination, cancer diagnosis, or biopsy over 20 years of screening in each country[7]

	Number per 1000 women screened from 20 years	
	USA	UK
Mean number of months between exams	18	36
Mean number of examinations/woman	13.7	7.3
Cancers detected	55.1	43.0
In situ	12.3	8.3
Invasive	42.8	34.8
Women recalled	694.3	267.1
Noninvasive diagnostic tests	553	183
Biopsy	142	85
Open surgical biopsy*	54	25

*Open surgical biopsies are a subset of all biopsies.
Adapted from Table 5, Smith-Bindman *et al.*, 2003.[7]

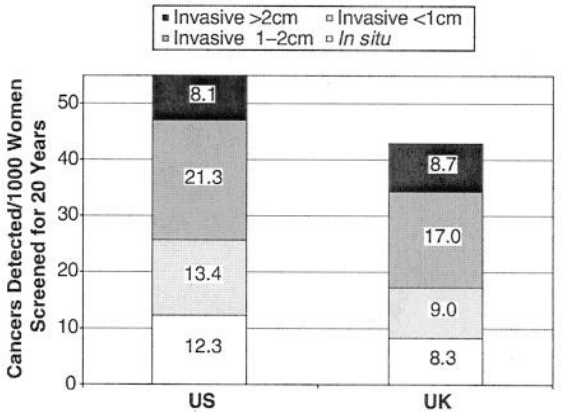


Figure 1 Cancers detected among 1000 women screened for 20 years in each country. Detailed information on the size of the invasive breast cancers is only available in the UK from 1998–1999, limiting the comparison of tumour size to that period. Additionally, invasive cancers without tumour size were excluded

USA, cancer detection rates were similar across settings for first screening examinations. For 1000 first screening examinations (8.6%, 95% CI 7.9–9.4 and 10.1%, 95% CI 9.4–10.7), cancers were diagnosed in the BCSC and NHSBSP, respectively. Cancer detection rates were higher in the UK for subsequent examinations, probably reflecting less frequent screening.

To determine if women in the UK are more likely to be diagnosed with large cancers if routinely screened in each country, we took into consideration both the types of cancers detected at each screening examination as well as the frequency of screening over time (Table 4). Figure 1 shows the cancer rates, by size, for 1000 women who underwent regular mammographic screening over 20 years in each country, including 13.7 screening examinations in the USA and 7.3 screening examinations in the UK. Overall, more cancers are detected in the USA; approximately 55.1 cancers would be detected in the USA and 43.0 in the UK. Most of the additional cancers detected in the USA were *in situ* and small invasive cancers. The number of large cancers detected was very similar between the two countries (8.1 versus 8.7) per 1000 women who underwent screening over 20 years. The higher frequency of screening in the USA magnifies the difference in the estimated recall rates between the countries when projected over 20 years. After 20 years of screening 1000 women, nearly three times as many women in the USA are recalled for additional work-up (694 versus 267, Table 4).

DISCUSSION

In the UK, half as many women are recalled for diagnostic tests following screening mammograms and half as many women without breast cancer undergo open surgical biopsies compared with the USA. However, cancer detection rates are similar. The goal of any cancer screening effort is to obtain high cancer detection rates while avoiding unnecessary diagnostic evaluation following false-positive results that are costly and associated with psychological morbidity.

We hypothesize several possible explanations for the differences in recall rates between the two countries. First, in the USA there are much higher rates of malpractice lawsuits that might lead American radiologists to recall women when they identify a suspicious finding, even though they may have a low likelihood of cancer.[19,20] While this is often said to be a possible explanation of the high recall rate in the USA, there

has been no study linking the malpractice environment to physician performance. Second, compared with their UK counterparts American physicians are low-volume readers, and on average read approximately 10–20% of the number of mammograms read annually by UK physicians. Increased volume has been associated with lower recall.[19–21] Third, although over 90% of programmes in the UK use double reading, this practice is much less common in the USA and some evidence suggests that double reading by consensus or arbitration as used in the UK raises detection rates and decreases recall rates.[22,23] Lastly, and perhaps most importantly, quality assurance standards for the NHSBSP programmes 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 programme operates at local and national levels to monitor and achieve these target values.[5,24] Furthermore, there is an organized programme of professional development in the UK that provides instruction and individual feedback regarding recall and cancer detection rates, using a set of test mammography cases called PERFORMS.[25] No comparable programme exists in the USA, but one is being developed by the American College of Radiology.

Screening mammography is performed more frequently in the USA than the UK. More frequent screening probably translates into smaller average cancer size at diagnosis. However, despite screening every three years instead of every one to two years, the rates of large tumours were fairly similar in the USA and UK (8.1 versus 8.7 per 1000 women screened for 20 years). There were substantially more *in situ* and small invasive cancers detected in the USA, yet this was not associated with as large a reduction in the number of advanced cancers, raising the possibility that *in situ* and early-stage cancers may be overdiagnosed in the USA. This study is unable to determine what impact the higher rate of diagnosis of *in situ* and small invasive cancers will have on mortality.

The main limitation of our study is that we cannot be certain that our definition of screening mammography was the same across both settings. Specifically, we do not know if misclassification of diagnostic examinations as screening examinations occurred to a greater extent in American women, and might account for a higher recall rate. A higher proportion of diagnostic mammograms among American women would, however, produce a substantially higher cancer rate in the USA setting,[26,27] which we did not find. We should also note that our estimation of the total cancers detected over 20 years was based on only four years of screening data, and the assumptions of the model were simplistic. However, our estimated recall rates are similar to the results found by others.[28,29] By pooling data within each programme, we have ignored variations by region, physician and other variables in each programme.[30,31] Lastly, although the data from the UK include virtually all mammographic screening performed in that country, the USA data reflect only a small percentage of mammography performed in the USA. However, prior analyses of these data have found that women participating in the BCSC are similar to women in the USA in general.

We compared cancer detection rates between the two countries, a widely used measure of mammography performance, as it closely approximates the total cancer rates.[31,32] Age-adjusted cancer rates have been noted to be significantly higher in the USA,[33] thus one would expect that in the USA the number of cancers detected would be higher. The number of cancers detected at each screening examination is higher in the UK (Table 3), whereas the number of cancers detected over two decades of screening is higher in the USA (Table 4). There were more cancers detected in the USA over the two decades of screening because women are exposed to more frequent screening examinations.

What can be done? The USA has demonstrated dramatic improvement in the technical components of mammography through regulations and oversight provided by Mammography Quality Standards Act (MQSA). Similar quality improvement could be targeted to the interpretative component of mammography. Specific interpretive goals and benchmarks of screening mammography could be set and developed within the radiological professional societies with oversight by MQSA. It seems that, unless specific benchmarks of physician performance are set and physician evaluation embraced, a major opportunity will be lost to further improve this aspect of the performance of screening mammography in the USA.

In the UK, the NHSBSP has set and reached targets that emphasize high rates of cancer detection and low recall. Recall rates in the UK are now substantially lower than in the USA without large differences in cancer detection. We believe this success stems primarily from a centralized programme of continuous quality improvement. Screening women aged 50–69 years biennially and reducing recall rates could substantially reduce the cost of mammography, as well as associated anxiety caused by false-positive diagnoses.[34,35] Efforts to improve USA mammographic screening should be targeted to lowering the recall rate without substantially lowering the cancer detection rate. The current MQSA standards could be revised to include guidelines that emphasize high standards on interpretation.

ACKNOWLEDGEMENTS

This work was supported by the Department of Defense (DAMD179919112) and the National Cancer Institute (CA86032 and cooperative agreements UO1CA63731, UO1CA63736, UO1CA63740, UO1CA69976, UO1CA70013, UO1CA70040, UO1CA86076 and UO1CA86082) as part of the NCI's Breast Cancer Surveillance Consortium. Some of these analyses have been previously reported in Smith-Bindman *et al.*, Comparison of Screening Mammography in the United States and United Kingdom. *JAMA* 2003;**290**: 2129–37 (Copyright American Medical Association © 2003. All rights reserved).

Authors' affiliations
Rebecca Smith-Bindman, *Associate Professor,* Department of Radiology, University of California, San Francisco, USA; Department of Epidemiology and Biostatistics, University of California, San Francisco
Rachel Ballard-Barbash, *Associate Director,* Applied Research Program, National Cancer Institute, NIH, Bethesda, MD, USA
Diana L.Miglioretti, *Assistant Scientific Investigator,* Center for Health Studies, Group Health Cooperative and *Assistant Professor,* Department of Biostatistics, University of Washington
Julietta Patnick, *Director,* National Health Service Breast Screening Programs, Sheffield, England
Karla Kerlikowske, *Professor,* Department of Epidemiology and Biostatistics, University of California, San Francisco; Department of Medicine, University of California, San Francisco; General Internal Medicine Section, Department of Veterans Affairs, University of California, San Francisco

REFERENCES

1 Houn F, Brown ML. Current practice of screening mammography in the United States: data from the National Survey of Mammography Facilities. *Radiology* 1994;**190**:209–15

2 Blanks R, Moss S, Patnick J. Results from the United Kingdom NHS breast screening programme 1994–1999. *J Med Screen* 2000;**7**:195–8
3 National Health Service Breast Screening Program Website. 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**:1373–7
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**:11–4
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 Epidemiol* 1998;**27**:735–42
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**:2129–37
8 Elmore JG, Yakano CY, Koepsell TD, *et al.* International variation in screening mammography. Interpretations in community-based programs. *J Natl Cancer Inst* 2003;**95**:1384–93
9 Ballard-Barbash R, Taplin SH, Yankaskas BC, *et al.* Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. *Am J Roentgenol* 1997;**169**:1001–8
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. *Ann Int Med* 2003;**138**:168–173
12 Barlow W, Lehman C, Zheng Y, *et al.* Performance of diagnostic mammography for women with signs or symptoms of breast cancer. *J Natl Cancer Inst* 2002;**94**:1151–9
13 Carney PA, Miglioretti DL, Yankaskas BC, *et al.* Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Intern Med* 2003;**138**:168–75
14 Miglioretti DL, Rutter CM, Geller BM, *et al.* Effect of breast augmentation on the accuracy of mammography and cancer characteristics. *JAMA* 2004;**291**:442–50
15 American College of R. *Breast Imaging Reporting and Data System (BI-RADS)* 3rd edn. Reston, VA: American College of Radiology, 1998
16 Kerlikowske K, Grady D, Barclay J, *et al.* 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, *et al.* Medical audit of a rapid-throughput mammography screening practice: methodology and results of 27,114 examinations. *Radiology* 1990;**175**:323–7
18 May D, Lee N, Nadel M, *et al.* The National Breast and Cervical Cancer Early Detection Program: report on the first 4 years of mammography provided to medically underserved women. *Am J Roentgenol* 1998;**170**:97–104
19 Esserman L, Cowley H, Eberle C, *et al.* Improving the accuracy of mammography: volume and outcome relationships. *J Natl Cancer Inst* 2002;**94**:369–75
20 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**:563–7
21 Sickles E, Wolverton D, Dee K. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology* 2002;**224**:861–9
22 Kopans D. Double reading. *Radiol Clin North Am* 2000;**38**:719–24
23 Williams L, Hartswood M, Prescott R. Methodological issues in mammography double reading studies. *J Med Screen* 1998;**5**:202–6
24 NHS Breast Screening Radiologists Quality Assurance Committee. *Quality Assurance Guidelines for Radiologists.* Sheffield, UK: NHSBSP, 1997, (publication no. 15, revised May 1997)
25 Gale A, Cowley H. Breast cancer screening: comparison of radiologists' performance in a self-assessment scheme and in actual breast screening In: Krupinski EA, ed. *Medical Imaging 1999: Image Perception and Performance. Proc SPIE1 # 999.* Vol. 3663 157–68
26 Sohlich R, Sickles E, Burnside E, Dee K. Interpreting data from audits when screening and diagnostic mammography outcomes are combined. *AJR* 2002;**178**:681–6
27 Dee K, Sickles E. Medical audit of diagnostic mammography examinations: comparison with screening outcomes obtained concurrently. *AJR* 2001;**176**:729–33
28 Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations (see comments). *N Engl J Med* 1998;**338**:1089–96
29 Christiansen C, Wang F, Barton M, *et al.* Predicting the cumulative risk of false-positive mammograms. *J Natl Cancer Inst* 2000;**92**:1657–66
30 Beam CA, Sullivan DC, Layde PM. Effect of human variability on independent double reading in screening mammography. *Acad Radiol* 1996;**3**:891–7
31 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**:24–8
32 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
33 http://www.iarc.fr/
34 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
35 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

Differences in the Quality of Breast Cancer Care Among Vulnerable Populations

David A. Haggstrom, M.D.[1]
Chris Quale, Ph.D.[2]
Rebecca Smith-Bindman, M.D.[2,3]

[1] Cancer Prevention Fellowship Program, Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, Bethesda, Maryland.

[2] Department of Radiology, University of California, San Francisco, California.

[3] Department of Epidemiology and Biostatistics, University of California, San Francisco, California.

This research was supported by the University of California, San Francisco, Division of General Internal Medicine, General Internal Medicine Clinical Research Fellowship; Health Resources and Services Administration Faculty Development Grant (1D14 HP001 78–01); California Breast Cancer Research Program (9PB-0205); Department of Defense (DAMD17–99–1-9112); National Institutes of Health (NIH) CA86032.

Address for Reprints: David A. Haggstrom, Executive Plaza North, Room 4009C, 6130 Executive Blvd., Bethesda, MD, 20892–7344; Fax: (301) 435–3710; E-mail: haggstrd@mail.nih.gov

Received April 26, 2005; revision received June 1, 2005; accepted June 28, 2005.

BACKGROUND. It is unknown whether differences in the quality of breast cancer care among women from racial and ethnic minority groups, the elderly, and rural areas have changed over time across the continuum of care.

METHODS. The linked Surveillance, Epidemiology, and End Results-Medicare database identified 22,701 women ages 66–79 years diagnosed with early stage breast cancer from 1992–1999. Multiple breast cancer processes of care were measured, including breast-conserving surgery, radiation therapy, documentation of estrogen receptor status, surveillance mammography, and a combined measure of "adequate care".

RESULTS. African-American and Hispanic women were significantly less likely to receive adequate care than White women in unadjusted comparisons (54.7% and 58.0% vs. 68.4% for African-American and Hispanic vs. White women) and adjusted comparisons (adjusted odds ratio [AOR] 0.67; 95% confidence interval [95% CI] 0.59–0.76, and AOR 0.77; 95% CI 0.66–0.90 for African-American and Hispanic women, respectively). The proportion of Asian/Pacific Islander women receiving adequate care was similar to White women. When considering only women diagnosed with breast cancer from 1997–1999, African-American women remained less likely than White women to receive adequate care (AOR 0.63; 95% CI 0.50–0.79). Women ages 75–79 years were less likely to receive adequate care compared with women ages 66–69 years (AOR 0.74; 95% CI 0.69–0.80), and women from rural (vs. metropolitan) areas were less likely to receive adequate care (AOR 0.81; 95% CI 0.73–0.89).

CONCLUSIONS. The quality of breast cancer care is lower among vulnerable populations across the continuum of care, and many of these differences have not improved in more recent years. ***Cancer* 2005;104:2347–58.**
© 2005 American Cancer Society.

KEYWORDS: breast cancer, therapy, African Americans, Hispanic Americans, Asian Americans, physician's practice patterns, women's health.

Substantial racial and ethnic differences in breast cancer outcomes and the use of healthcare services have been observed. African-American women have a lower incidence of breast cancer, but death rates among African-American women are higher than White women, and the mortality gap has increased.[1] African-American women have been less likely to receive radiation therapy (XRT) compared with White women,[2,3] a treatment difference that may partly explain these differences in breast cancer mortality. Hispanic and Asian/Pacific Islander women in the United States both have been less likely to receive breast-conserving surgery (BCS) surgery,[4,5] a treatment difference that may adversely impact their quality of life (QOL). Other vulnerable populations are also more likely to receive inadequate care

© 2005 American Cancer Society
DOI 10.1002/cncr.21443
Published online 6 October 2005 in Wiley InterScience (www.interscience.wiley.com).

for breast cancer, including the elderly, women who live in rural areas, and women of lower socioeconomic status.[6–8]

To understand the reasons for differences in cancer outcomes, it is important to assess the quality of breast cancer care. When assessing the quality of breast cancer care, process measurements have the advantage of occurring more frequently than mortality and of being potentially modifiable within the context of the healthcare delivery system.[9] The Institute of Medicine (IOM) report, Ensuring Quality Cancer Care,[10] stressed the importance of delivering high-quality services across the full span of the cancer care continuum. We sought to evaluate differences in breast cancer treatment and surveillance as a way to characterize differences in the quality of breast cancer care across as broad a spectrum of healthcare services as possible.

Several evidence-based guidelines regarding appropriate treatment processes for early stage breast cancer have been published that enable evaluation of the quality of breast cancer care. Breast-conserving treatment (including both BCS and XRT) was recommended by the 1990 National Institutes of Health Consensus Panel as preferable to mastectomy because of equivalent survival and improved QOL.[11] BCS without XRT is considered inadequate because BCS is associated with a higher cancer recurrence rate without radiation. Because assessment of estrogen receptor (ER) status is necessary to determine whether chemotherapy is appropriate, the adequate documentation of ER status can also be used to measure quality. Finally, the quality of care provided in the survivorship phase of cancer is important, as many cancer patients live for years after initial treatment. Surveillance mammography on an annual basis after diagnosis with breast cancer is recommended as a clinical guideline.[12]

By using part of the framework implemented in the National Healthcare Disparities Report, we set out to describe differences in the use of breast cancer care by vulnerable "priority" populations.[13] These populations include women from racial and ethnic minorities, the elderly, rural populations, and low-income groups. We planned to expand and update previous knowledge regarding differences in the quality of breast cancer care among these populations using the most recent information available. We used population-based data from the Surveillance, Epidemiology, and End Results (SEER)-Medicare program that includes breast cancers diagnosed from 1992 to 1999 and treated through 2002.

MATERIALS AND METHODS

Data Source

The SEER-Medicare database used for this analysis was created as a collaborative effort of the National Cancer Institute, SEER program, and the Centers for Medicare and Medicaid Services (CMS) to create a population-based source of information for cancer-related research. The SEER cancer registries identified all incident breast cancer diagnoses in selected areas from 1992 to 1999 and provided diagnostic and treatment information within 4 months after the date of diagnosis. The Medicare files provided claims information about breast cancer treatment (including primary breast cancer treatment that may not have been captured in the SEER data) and surveillance after diagnosis through 2002. The process for linking these data has been described by Warren et al.[14] The SEER program includes population-based tumor registries in several geographic areas and covers approximately 14% of the U.S. population.

Study Population

Women ages 66 to 79 years who were diagnosed with Stage I or II breast cancer from 1992 to 1999 were included in our study. We included women beginning at age 66 years so that there would be at least 1 year of Medicare claims before diagnosis by which to assess comorbidity. We excluded women older than age 79 years because of more uncertainty regarding appropriate care among these women, and we also excluded women with tumors > 5 cm because neoadjuvant chemotherapy, as opposed to surgery, is commonly the first course of treatment for these women.

Women were excluded who were diagnosed with another cancer before, or within 2 years after, the incident breast cancer to ensure all treatment was intended for the identified breast cancer. Medicare does not receive individual billing claims for patients enrolled in HMO plans, and healthcare coverage may be provided as part of either Medicare Part A or B. Therefore, women were excluded who had HMO coverage or were not covered by both Parts A and B of Medicare 1 year before, or 2 years after, diagnosis with breast cancer. Finally, we excluded those women who died during the period when they were eligible for treatment.

Predictors

Race or ethnicity was defined as non-Hispanic White, African-American, non-Hispanic; Hispanic, Asian/Pacific Islander, Native American, or other/unknown using the recoded SEER race variable.[15] Hispanic women were identified by SEER through an algorithm that

TABLE 1
Operational Definitions of Breast Cancer Processes of Care

Breast cancer process of care	SEER	Medicare
Breast-conserving surgery	segmental/subtotal mastectomy, lumpectomy, quandrantectomy, tylectomy, wedge resection, nipple resection, excisional biopsy, or partial mastectomy, not otherwise specified	ICD-9-CM procedure codes: 85.20-85.23 CPT codes: 19120, 19125, 19126, 19160, 19162
Mastectomy	subcutaneous, total (simple), modified radical, radical, or extended radical mastectomy	ICD-9-CM procedure codes: 85.41, 854.2, 854.3, 85.44, 85.45-85.48 CPT codes: 19240, 19220, 19180, 19182
Radiation therapy	Beam radiation, radioactive implants, radioisotopes, or other radiation	ICD-9-CM procedure codes: 92.21-92.29 Revenue center codes: 0330, 0333 CPT codes: 77401-77499, 77750-77799
Estrogen receptor status	Data from SEER	
Surveillance mammography		CPT codes: 70690-76092

ICD-9-CM: Internal Classification of Disease, 9th revision, Clinical Modification; CPT: Current Procedural Terminology.

uses Spanish surnames. Other categories of race/ethnicity were determined by SEER from medical records and registration information. The Asian/Pacific Islander category includes Chinese, Japanese, Filipino, and Hawaiian women.

Age was calculated from the patient's birth date using SEER and then grouped into 5-year categories. Rural residence was obtained from SEER and dichotomized by designating the assigned area as rural (county either distant or adjacent to a metropolitan area, and including counties with a population of $\leq$ 20,000) or metropolitan (county in a metropolitan area, and including areas with a population of $\geq$ 250,000). Socioeconomic status was obtained from SEER by using median income in the patient's census tract; if census tract information was missing, then median income in the patient's zip code was used.[15]

Individual year of diagnosis, SEER region, tumor size, and American Joint Committee on Cancer tumor stage were obtained from SEER and included as categoric predictors of cancer treatment. Comorbidity was measured with a Charlson-based comorbidity index derived from Medicare inpatient and outpatient claims. This index has previously been shown to be predictive of 2-year mortality and the receipt of less aggressive therapy, specifically BCS without XRT.[16]

Processes of Care

Surgery receipt was measured using information from both SEER and Medicare.[2,17–19] Surgery was considered to occur if it was noted in either SEER or Medicare (Table 1). In cases of disagreement concerning type of surgery, the most invasive surgery (mastectomy) noted in either SEER or Medicare was used. *Radiation therapy* receipt was measured using information from both SEER and Medicare[20,21] (Table 1). If XRT was noted to occur in either SEER or Medicare, then it was considered to have occurred. Treatment with both surgery and XRT were measured within a 6-month interval after breast cancer diagnosis.

Estrogen receptor status documentation was measured using SEER data alone, as this information is not available in Medicare. ER status was categorized as 1) positive, 2) negative, 3) borderline, 4) not done, 5) ordered but no result recorded in chart, and 6) unknown or not documented in patient records. These categories were collapsed and dichotomized into adequate (Categories 1–3) and inadequate (Categories 4–6) ER status documentation.

Mammography surveillance was measured by using information from Medicare alone. For each patient, mammograms that occurred within the first 6 months after initial surgical treatment were considered part of the treatment period, and mammograms that occurred 7–18 months after initial surgical treatment were considered to have occurred for surveillance.[22] All mammograms performed within the first 3–5 years after a breast cancer diagnosis are labeled as diagnostic, and therefore, we included all billing codes that may have been used to identify a mammogram (Table 1).

Adequate care was defined as either the receipt of XRT among women receiving BCS, or the receipt of mastectomy. In addition, adequate care was defined as adequate documentation of ER status and the receipt of surveillance mammography among all women. Therefore, a woman was defined as receiving adequate care if she received all of the following: either mastectomy or BCS with XRT, adequate documentation of ER status, and surveillance mammography.

Analysis

We calculated the proportion of women receiving BCS among women with Stage I or II breast cancer. XRT is recommended for women undergoing breast-conserving treatment, and thus, we calculated the proportion of women receiving XRT only among women who had received BCS. Both the proportion of women with adequate documentation of ER status and surveillance mammography were calculated among all women with Stage I or II breast cancer who underwent surgery. The proportion of women with adequate care was calculated among all women with Stage I or II breast cancer. Changes over time in each breast cancer process of care among different racial and ethnic groups were charted with moving 3-year averages. Changes over time by year in the overall proportion of women receiving each process of care were evaluated for statistical significance using a nonparametric test for trend across ordered groups.[23]

We evaluated the associations among race/ethnicity, age, and rural status, and the receipt of each breast cancer process of care (BCS, XRT, adequate documentation of ER status, mammography surveillance) and adequate care using chi-squared tests. By using all of the predictors and covariates previously described, we created multivariate logistic regression models for the outcomes of receipt of BCS, XRT, and adequate care. Similar logistic regression models were developed for the outcomes of adequate documentation of ER status and surveillance mammography, but these models excluded clinical predictors of tumor size and stage, as women should receive these care processes regardless of tumor size or stage.

All multivariate logistic regression models were repeated, including median area level income as a predictor to determine whether socioeconomic status affects racial/ethnic, age, or rural differences in breast cancer care. Income was modeled as a continuous log term. Potential interactions between race/ethnicity and age, as well as race/ethnicity and income, were tested; there were no significant interactions in either case, and therefore, no interaction terms were included in the final models. Finally, all logistic regression models were repeated and limited to breast cancer diagnoses from 1997 to 1999 for the purpose of determining if any observed patterns of care persisted at the end of the decade.

RESULTS

There were 22,701 women ages 66–79 years diagnosed with early stage breast cancer from 1992 to 1999, including, 1137 African-American, 727 Hispanic, and 772 Asian/Pacific Islander women (Table 2). The majority of women (83.1%) lived in metropolitan areas, and most were diagnosed with Stage I tumors (65.8%) and tumors <2 cm in size (68.9%).

TABLE 2
Patient Characteristics of Women with Early-Stage Breast Cancer

	No. of women	Percentage
Race		
White	19804	87.2
African-American	1137	5.0
Hispanic	727	3.2
Asian/Pacific Islander	772	3.4
Native American	39	0.2
Other/unknown	222	1.0
Age in yrs		
66–69	6395	28.2
70–74	8862	39.0
75–79	7444	32.8
Geography		
Rural	3830	16.9
Metropolitan	18871	83.1
SEER site		
Atlanta	1309	5.8
Connecticut	3038	13.4
Detroit	3693	16.3
Hawaii	614	2.7
Iowa	3321	14.6
Los Angeles	3498	15.4
New Mexico	896	4.0
San Francisco	1599	7.0
San Jose/Monterey	1050	4.6
Seattle	2588	11.4
Utah	1095	4.8
AJCC tumor stage		
Stage I	14926	65.8
Stage II	7775	34.2
Tumor size		
< 2 cm	15634	68.9
2–5 cm	6866	30.2
unknown	201	0.9
Comorbidity index		
0	18150	80.0
1	3271	14.4
2	671	3.0
≥ 3	177	0.8
unknown	432	1.9
Median income		
mean (SD)	$54,685 ($25,203)	
Year of diagnosis		
1992–1999 cases/yr	range 2,722–3,092/year	

AJCC: American Joint Committee on Cancer.

Almost half (48.8%) of women received BCS, and among women who received BCS, 88.1% received XRT (Table 3). Among women who did not receive BCS, 45 (0.4%) had no surgery identified. Greater than 80% of women had adequate documentation of ER status or surveillance mammography after breast cancer sur-

TABLE 3
Proportion of Women Receiving Breast Cancer Processes of Care by Race/Ethnicity, Age, and Rural Status

	Breast-conserving treatment									
	Breast-conserving surgery		Radiation therapy		Adequate ER documentation		Surveillance mammography		Adequate care	
No. of women eligible	22,701		11,086		22,656		22,656		22,701	
	%	*P* value	%	*P* value	%	*P* value	%	*P* value	%	*P* value
All women	48.8		88.1		82.2		85.3		67.3	
Race/ethnicity										
White (*reference group*)	49.1		88.9		82.7		86.0		68.4	
African-American	49.3	0.93	79.9	< 0.001	75.8	< 0.001	78.7	< 0.001	54.7	< 0.001
Hispanic	46.1	0.11	84.1	0.008	76.0	< 0.001	79.3	< 0.001	58.0	< 0.001
Asian/Pacific Islander	43.0	0.001	89.6	0.73	86.5	0.005	84.2	0.15	71.4	0.08
Age in yrs										
66–69 (*reference group*)	49.9		91.9		82.2		87.2		69.7	
70–74	49.2	0.43	90.4	0.03	81.9	0.62	86.3	0.12	68.5	0.12
75–79	47.5	0.005	82.0	< 0.001	82.5	0.67	82.5	< 0.001	63.7	< 0.001
Geography										
Metropolitan	51.6		88.4		81.5		85.1		66.4	
Rural	35.1	< 0.001	85.9	0.009	85.5	< 0.001	85.9	0.22	71.3	< 0.001

ER: estrogen receptor.
All proportions are adjusted to a standard age distribution. P values are from Pearson chi-square test. P value < 0.05 means that the group is different from the referent.

gery. By using a composite measure to aggregate individual processes of care, approximately two-thirds (67.3%) of women received adequate care.

Race and Ethnicity

Breast-Conserving Surgery

Asian/Pacific Islander women were significantly less likely to receive BCS compared with White women in both unadjusted (43.0% vs. 49.1%, Table 3) and adjusted comparisons (OR 0.52; 95% CI 0.43–0.63; Table 4). Hispanic women were not significantly less likely to receive BCS than White women in unadjusted comparison (46.1% vs. 49.1%, $P = 0.11$), but, in adjusted comparison (OR 0.85; 95% CI 0.72–0.99), this difference became statistically significant ($P = 0.046$). African-American women were no less likely to receive BCS than White women.

Radiation Therapy

Among women who received BCS, African-American and Hispanic women were significantly less likely to receive XRT compared with White women in unadjusted (79.9% and 84.1% vs. 88.9% for African-American and Hispanic vs. White women) and adjusted comparisons (OR 0.55; 95% CI 0.44–0.70, and OR 0.68; 95% CI 0.50–0.94 for African-American and Hispanic women, respectively). Asian/Pacific Islander women were no less likely to receive XRT than White women.

Estrogen Receptor Status

Hispanic women were significantly less likely to have adequate documentation of ER status compared with White women in unadjusted (76.0% vs. 82.7%) and adjusted comparisons (OR 0.82; 95% CI 0.69–0.99). Similarly, African-American women were less likely to have adequate documentation of ER status than White women in unadjusted (75.8% vs. 82.7%) and adjusted comparisons (OR 0.87; 95% CI 0.75–1.00), although for the adjusted comparison, this difference bordered on statistical significance ($P = 0.057$). Asian/Pacific Islander women were more likely to have adequate documentation of ER status than White women in unadjusted comparison (86.5% vs. 82.7%), but this difference was no longer present in adjusted comparison (OR 0.87; 95% CI 0.67–1.11).

Surveillance Mammography

African-American and Hispanic women were significantly less likely to receive surveillance mammography than White women in unadjusted (78.7% and 79.3% vs. 86.0% for African-American and Hispanic vs. White women) and adjusted comparisons (OR 0.58; 95% CI 0.50–0.68, and OR 0.73; 95% CI 0.60–0.88 for African-American and Hispanic women, respectively). Asian/Pacific Islander women were no less likely than White women to receive surveillance mammography.

TABLE 4
Likelihood of Receiving Breast Cancer Processes of Care among Women Diagnosed with Breast Cancer from 1992 to 1999

	Breast-conserving surgery		Radiation therapy		Adequate ER documentation		Surveillance mammography		Adequate care	
	AOR[a]	95% CI	AOR[a]	95% CI	AOR[b]	95% CI	AOR[b]	95% CI	AOR[a]	95% CI
Race/ethnicity										
White		referent		referent		referent		referent		referent
African-American	1.03	(0.91–1.18)	0.55	(0.44–0.70)[e]	0.87	(0.75–1.00)	0.58	(0.50–0.68)[e]	0.67	(0.59–0.76)[e]
Hispanic	0.85	(0.72–0.99)[c]	0.68	(0.50–0.94)[c]	0.82	(0.69–0.99)[c]	0.73	(0.60–0.88)[d]	0.77	(0.66–0.90)[d]
Asian/Pacific Islander	0.52	(0.43–0.63)[e]	0.76	(0.48–1.21)	0.87	(0.67–1.11)	1.08	(0.83–1.39)	1.06	(0.86–1.29)
Age in yrs										
66–69		referent		referent		referent		referent		referent
70–74	0.92	(0.86–0.99)[c]	0.82	(0.70–0.97)[c]	0.97	(0.89–1.06)	0.91	(0.82–0.99)[c]	0.94	(0.87–1.01)
75–79	0.85	(0.79–0.92)[e]	0.40	(0.34–0.47)[e]	1.00	(0.92–1.10)	0.67	(0.61-0.74)[e]	0.74	(0.69–0.80)[e]
Geography										
Metropolitan		referent		referent		referent		referent		referent
Rural	0.72	(0.65–0.79)[e]	0.75	(0.60–0.92)[d]	0.67	(0.59–0.77)[e]	0.93	(0.82–1.06)	0.81	(0.73–0.89)[e]

ER: estrogen receptor; AOR: adjusted odds ratio; 95% CI: 95% confidence interval.
[a] Regression model adjusted for year of diagnosis, SEER region, Charlson-based comorbidity index, tumor size, and stage.
[b] Regression model adjusted for year of diagnosis, SEER region, and Charlson-based comorbidity index.
[c] Significant at $P < 0.05$.
[d] Significant at $P < 0.01$.
[e] Significant at $p < 0.001$

Adequate Care

African-American and Hispanic women were significantly less likely to receive adequate care than White women in unadjusted (54.7% and 58.0% vs. 68.4% for African-American and Hispanic vs. White women) and adjusted comparisons (OR 0.67; 95% CI 0.59–0.76, and OR 0.77; 95% CI 0.66–0.90) for African-American and Hispanic women, respectively). Asian/Pacific Islander women were no less likely than White women to receive adequate care.

Age

The receipt of most breast cancer processes of care declined with increasing patient age. Comparing women ages 75–79 years to women ages 66–69 years, there was a modest difference in receipt of BCS (OR 0.85; 95% CI 0.79–0.92, Table 4). Women aged 75–79 years were substantially less likely to receive XRT than women aged 66–69 years (OR 0.40; 95% CI 0.34–0.47). There was no significant difference in adequate documentation of ER status, but there was a significant difference in surveillance mammography between women ages 75–79 years and those ages 66–69 years (OR 0.67; 95% CI 0.61–0.74). Overall, the receipt of adequate care was lower among older (75–79 yrs) than younger (66–69 yrs) women (OR 0.74; 95% CI 0.69–0.80).

Rural

Women from rural areas were significantly less likely than women from metropolitan areas to receive BCS (OR 0.72; 95% CI 0.65–0.79, Table 4), XRT (OR 0.75; 95% CI 0.60–0.92), and adequate documentation of ER status (OR 0.67; 95% CI 0.59–0.77). Rural women were no less likely than metropolitan women to receive surveillance mammography. Overall, rural women were less likely to receive adequate care than metropolitan women (OR 0.81; 95% CI 0.73–0.89).

Income

Including median area level income in the multivariate regression models did not significantly alter results among African-American women (results not shown in table). African-American women remained significantly less likely than White women to receive XRT (OR 0.62; 95% CI 0.48–0.79), surveillance mammography (OR 0.65; 95% CI 0.55–0.77), and adequate care (OR 0.72; 95% CI 0.63–0.82). Among Hispanic women, adjusting for median area level income attenuated the results. Hispanic women were no longer significantly less likely than White women to receive BCS (OR 0.93; 95% CI 0.79–1.10), XRT (OR 0.74; 95% CI 0.53–1.03), or adequate documentation of ER status (OR 0.84; 95% CI 0.70–1.01). Yet Hispanic women remained significantly less likely to receive surveillance mammography (OR 0.77; 95% CI 0.63–0.93) and overall adequate care than White women (OR 0.81; 95% CI 0.69–0.95). Lower use of breast cancer processes of care among older women and rural women also persisted after adjusting for income.

When we considered median area level income

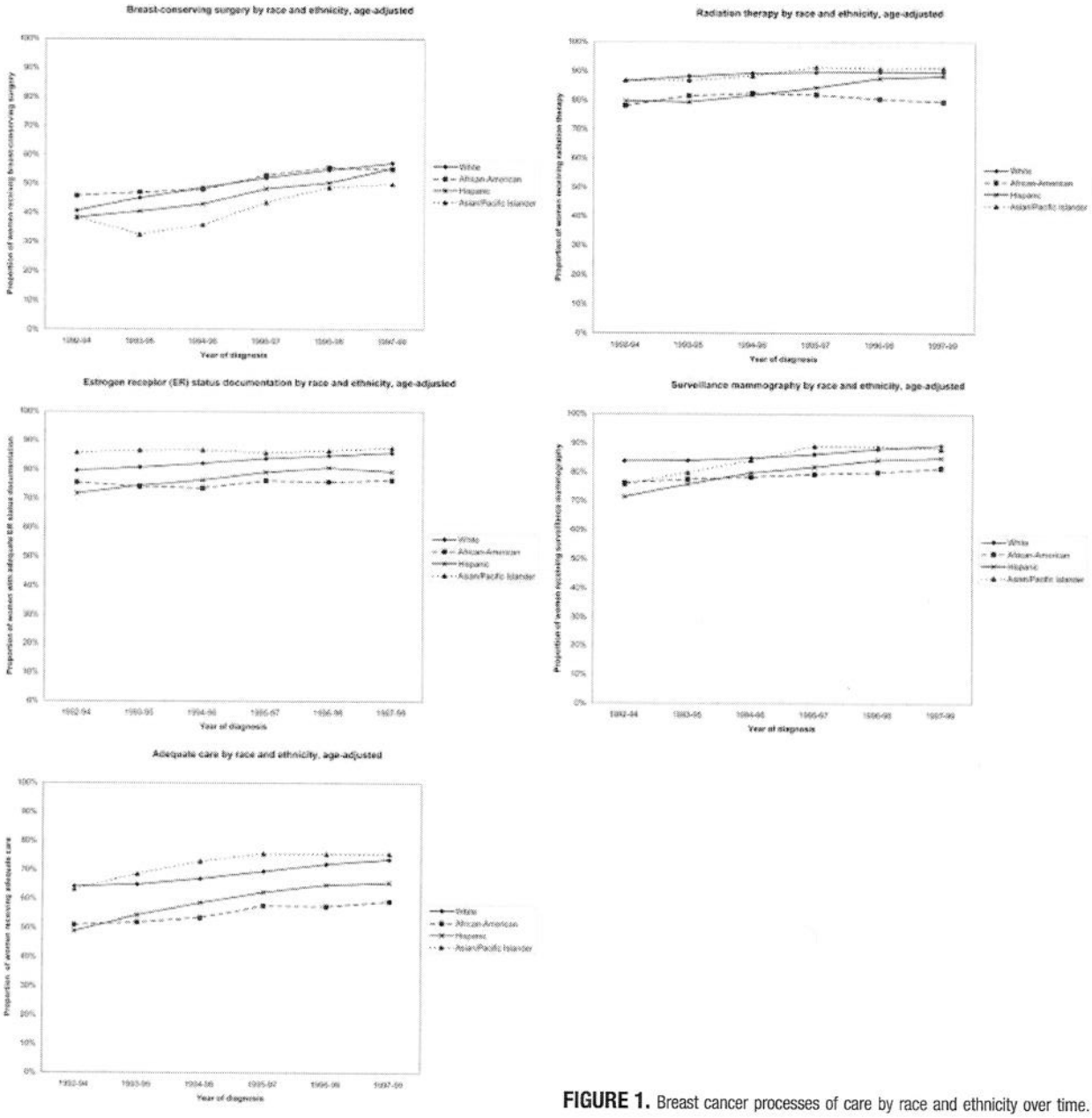


FIGURE 1. Breast cancer processes of care by race and ethnicity over time.

directly, we found that women living in areas of higher socioeconomic status were more likely to receive most processes of care. For every 25% increase in income, women had 7% greater odds of receiving BCS, 5% greater odds of receiving XRT, 4% greater odds of surveillance mammography, and 3% greater odds of receiving adequate care. There was no greater likelihood of adequate documentation of ER status with increasing income.

Time Trends

The proportion of women who received every breast cancer process of care, as well as adequate care, increased over time between the periods of 1992–1994 and 1997–1999 (Fig. 1). The receipt of BCS increased from 40.8% to 56.5%. The receipt of XRT (among women who received BCS) increased from 86.1% to 89.0%, the adequate documentation of ER status in-

TABLE 5
Likelihood of Receiving Breast Cancer Processes of Care among Women Diagnosed with Breast Cancer from 1997 to 1999

	Breast-conserving surgery		Radiation therapy		Adequate ER documentation		Surveillance mammography		Adequate care	
No. of women eligible	8319		4678		8300		8300		8319	
	AOR[a]	95% CI	AOR[a]	95% CI	AOR[b]	95% CI	AOR[b]	95% CI	AOR[a]	95% CI
Race/ethnicity										
White		referent		referent		referent		referent		referent
African-American	1.08	(0.86–1.35)	0.48	(0.33–0.70)[e]	0.70	(0.54–0.91)[d]	0.58	(0.43–0.78)[e]	0.63	(0.50–0.79)[e]
Hispanic	1.04	(0.80–1.36)	0.94	(0.55–1.60)	0.87	(0.63–1.20)	0.84	(0.59–1.20)	0.90	(0.69–1.18)
Asian/Pacific Islander	0.61	(0.45–0.83)[d]	0.77	(0.38–1.53)	0.92	(0.62–1.38)	1.28	(0.83–1.99)	1.16	(0.84–1.60)
Age in yrs										
66–69		referent		referent		referent		referent		referent
70–74	0.90	(0.80–1.01)	0.94	(0.73–1.22)	1.05	(0.90–1.23)	0.87	(0.73–1.05)	1.00	(0.88–1.13)
75–79	0.85	(0.76–0.96)[c]	0.48	(0.38–0.62)[e]	1.03	(0.88–1.21)	0.65	(0.54–0.78)[e]	0.76	(0.67–0.86)[e]
Geography										
Metropolitan		referent		referent		referent		referent		referent
Rural	0.74	(0.64–0.87)[e]	0.69	(0.50–0.95)[c]	0.71	(0.56–0.89)[d]	1.06	(0.84–1.34)	0.89	(0.75–1.05)
Median income										
25% increase	1.06	(1.03–1.09)[e]	1.03	(0.98–1.08)	1.02	(0.99–1.06)	1.05	(1.01–1.09)[d]	1.04	(1.01–1.07)[d]

ER: estrogen receptor; AOR: adjusted odds ratio; 95% CI: 95% confidence interval.

[a] Regression model adjusted for median area level income, year of diagnosis, SEER region, Charlson-based comorbidity index, tumor size, and stage.

[b] Regression model adjusted for median area level income, year of diagnosis, SEER region, and Charlson-based comorbidity index.

[c] Significant at $P < 0.05$.

[d] Significant at $P < 0.01$.

[e] Significant at $P < 0.001$

creased from 79.3% to 85.1%, and surveillance mammography increased from 82.8% to 88.4%. The proportion of women who received adequate care increased from 63.0% to 72.1%. The change over time by year was significant for every breast cancer process of care measured ($P < 0.001$).

Recent Breast Cancer Diagnoses

Many important differences among vulnerable populations persisted at the end of the decade. All differences between African-American and White women in breast cancer care remained significant when considering only breast cancers diagnosed from 1997 to 1999 (Table 5). In addition, differences in the adequate documentation of ER status between African-American and White women became statistically significant in adjusted comparison only when considering more recent cancer diagnoses (OR 0.70; 95% CI 0.54–0.91). However, there were no longer significant differences between Hispanic and White women in adequate care, or any individual process of care, when considering only more recent breast cancer diagnoses. As before, there was no significant difference in adequate care between Asian/Pacific Islander and White women diagnosed with breast cancer from 1997 to 1999. Among women from rural areas, lower use of individual processes of care was unchanged, although overall, rural women were no longer significantly less likely to receive adequate care. Age-related patterns of care remained largely the same when only more recent years were considered.

DISCUSSION

Breast cancer care across the continuum of cancer treatment and surveillance outlined here has improved over time. Yet despite these improvements, only 67.3% of women received adequate care. There also remain significant gaps in the quality of care among vulnerable populations. African-American women were significantly less likely to receive adequate care compared with White women, and this gap persisted or worsened among women diagnosed with breast cancer later in time (1997–99). Gaps in quality of care among the elderly and women from rural areas and areas of lower socioeconomic status also persisted at the end of the decade.

Racial and Ethnic Differences

Asian/Pacific Islander women were less likely to receive BCS, but they were no less likely to receive any other breast cancer process of care. Lower rates of BCS

have been observed previously among Asian/Pacific Islander women in California,[4,5] but the Asian/Pacific Islander population of our study is notably different because it includes women from other regions, including Hawaii. Hispanic women were also less likely to receive BCS; a similar pattern of lower use was observed previously in California.[4]

African-American and Hispanic women were less likely to receive XRT, although this difference did not persist among Hispanic women at the end of the decade. These racial and ethnic differences in XRT are consistent with a previous study that used data from SEER registries alone.[24] Differences in receipt of XRT among women with early stage breast cancer is important because XRT appears to have a mortality benefit,[25] although adequate numbers of older women have not been studied.[26] A recent study found that African-American women had higher mortality after breast cancer than White women after adjustment for prognostic factors,[27] and we would suggest that racial differences in treatment may contribute to these mortality differences.

African-American and Hispanic women were less likely to receive both adequate documentation of ER status and surveillance mammography. To our knowledge, previous studies have not addressed racial or ethnic differences in the documentation of ER status. A racial difference in mammography after surgery for early stage breast cancer was previously observed in 1991.[28]

When we considered only women diagnosed with breast cancer from 1997 to 1999, African-American women remained less likely than White women to receive adequate care and became significantly less likely to have adequate documentation of ER status. When we considered more recent diagnoses among Hispanic women, however, differences in care were no longer present, which suggests that breast cancer care may have improved among this population during the course of the decade.

Age Differences

Age-related patterns of care can be difficult to interpret in terms of appropriateness because of the low proportion of women older than age 70 years who enroll in clinical trials. Nonetheless, XRT is currently considered to be a necessary component of breast-conserving treatment,[11] and we found that XRT decreased with increasing age. Declining use of XRT with advancing age has been observed previously and associated with patient preferences.[29] Potential exists for both undertreatment among healthy older women, as well as overtreatment among the frail elderly.

Rural Differences

Women who live in rural areas were less likely to receive BCS and XRT. These findings are consistent with a previous study[7] and likely reflect multiple factors, including barriers to travel in rural areas that make access to XRT more difficult. In contrast, rural differences in surveillance mammography were not significant. Mammography examinations may not be influenced by the same geographic factors as more intensive therapies.

Socioeconomic Differences

Previous research has suggested that racial differences in breast cancer treatment may be explained by socioeconomic status.[8,30] In our study, racial differences in breast cancer care did not change substantively when adjusted for median area level income. These findings likely differ from others because our study included multiple population-based cancer registries[8] and measured a different spectrum of breast cancer processes of care.[30] Hispanic women were no longer significantly less likely to receive many processes of care when we adjusted for income, which suggests that differences in socioeconomic status or social class[31] may play a larger role in treatment differences among this population.

Lower median area level income was independently associated with lower use of most breast cancer processes of care. Although culture is commonly associated with race or ethnicity, there may also be a "culture" of poverty—denoting shared beliefs and behaviors—that influences patients' choices of specific treatments.[32] Because median income was measured at the level of the census tract or zip code, these findings suggest that the socioeconomic status of the community in which a person lives is related to the quality of healthcare delivered to individuals in that community.

Explanations for Differences

Potential explanations for differences in patterns of care among vulnerable populations include poor healthcare access, regional variations, or a higher disease burden among these populations. Poor healthcare access, as measured by insurance coverage, was not clearly evident here given that all women were covered by Medicare Parts A and B. Differences in cancer care between similarly insured African-American and White patients have been observed previously in both breast cancer[3] and lung cancer.[33] Another dimension of access to consider in future research that uses Medicare claims is any

unmeasured variation in supplemental insurance. In addition, the uninsured represent another vulnerable population that should be considered in future studies. Whereas significant regional variation in breast cancer care may be present, our final models accounted for SEER region. Concerning disease burden, adjustments for comorbidity, as well as tumor size and cancer stage when appropriate, were incorporated into the overall results.

There are other possible explanations why the vulnerable populations described here may receive inadequate care. At the level of the patient–physician relationship, either partner may influence what care is received. Among physicians or institutions, there may be bias in provision of clinical services based upon a patient's racial identity,[34] relative lack of affluence, or even rural background. Access to clinical resources or less competence among physicians who serve vulnerable populations may also explain gaps in quality of care.[35] Conversely, different patients may have different preferences for care, and the perceptions and decisions of individuals may be influenced by cultural beliefs.[36] Family support mechanisms may further contribute to differences in care among vulnerable populations. Minority women, as well as women who live in rural areas and areas of lower socioeconomic status, may also seek care at more poorly performing medical institutions within a given region.

The process of documenting ER status has unique properties when considering explanations for any measured difference between groups. If the decision for surgery has been made, then assessment of ER status should routinely follow. It seems unlikely that patient preferences play any role in this regular process of care. Furthermore, if a patient has reached the point of surgery, access to the healthcare system (with or without insurance) no longer serves as a viable explanation for any gap in quality. Thus, the possible set of explanations for racial, ethnic, and rural differences in adequate documentation of ER status can be narrowed to either bias in the provision of services or gaps in quality of care at medical institutions predominantly serving these patients. Of course, this study measured adequate *documentation* of ER status, not necessarily *performance* of the test, but nonetheless, systematic differences in documentation of ER status reflect differential record-keeping at a minimum and, quite possibly, differential care.

Limitations

The primary limitation of our study design to answer the questions posed here, is its potential for misclassification of different processes of care. SEER-Medicare data has been found to be reasonably accurate for determination of surgery,[17,18] XRT,[20] and surveillance mammography.[22] SEER is commonly the gold standard for ascertainment of ER status in population-based studies,[37] and our definition of adequate documentation of ER status has face validity. Misclassification of race or ethnicity should also be considered. SEER data has previously been found to be accurate for the identification of Whites and African-Americans, and Hispanic and Asian/Pacific Islander ethnicity variables likely provide acceptable specificity.[15] Another limitation is that there may have been insufficient power to detect differences between groups when no difference was found, particularly for analyses limited to breast cancers diagnosed from 1997 to 1999. Therefore, the relative improvement in the receipt of all processes of care among Hispanic women may have been due to insufficient sample sizes. Nonetheless, there was sufficient power to detect multiple differences in breast cancer care in more recent years, and in all instances where a difference was present, the pattern was consistently one of lower use among vulnerable populations. Finally, there are other processes of care besides the ones measured here that contribute to optimal breast cancer care, including axillary lymph lymph node dissection, chemotherapy, chemoprevention (tamoxifen), and other prognostic biomarkers (progesterone receptor and Her-2). If there is underuse of any of these care processes, then the findings described here may underestimate gaps in the quality of breast cancer care.

Conclusions

These results highlight the importance of a cancer data system to track healthcare quality and disparities across as broad a spectrum of the continuum of cancer care as possible. Another report by the Institute of Medicine suggested that current cancer data systems like SEER-Medicare while important are not ideal,[38] and we would agree. In our view, at least two characteristics should be considered in the future when designing the ideal system. First, the ideal cancer data system should not only collect data that makes comparisons among vulnerable groups possible, but the data system should also routinely collect information that allows testing and monitoring suspected causes of these differences. Second, an ideal cancer data system would rapidly identify gaps in quality so that their recognition would make an immediate difference in the care of patients involved. Our findings of differences in breast cancer

care among vulnerable populations demonstrate patterns largely consistent with previous literature. Importantly, this study provides information indicating that the problem of gaps in the quality of care among vulnerable populations, despite being identified as a target for intervention previously, persist and require bold, decisive action. Surveillance needs to be accompanied by interventions targeted both at minorities[39] and other vulnerable populations, and evidence-based interventions should be disseminated widely.

REFERENCES

1. Weir HK. Thun MJ, Hankey BF, et al. Annual report to the nation on the status of cancer, 1975–2000, featuring the uses of surveillance data for cancer prevention and control. *J Natl Cancer Inst.* 2003;95:1276–1299.
2. Ballard-Barbash R, Potosky AL, Harlan LC, et al. Factors associated with surgical and radiation therapy for early stage breast cancer in older women. *J Natl Cancer Inst.* 1996;88:716–726.
3. Mandelblatt JS, Kerner JF, Hadley J, et al. Variations in breast carcinoma treatment in older Medicare beneficiaries: is it black or white. *Cancer.* 2002;95:1401–1414.
4. Morris CR, Cohen R, Schlag R, et al. Increasing trends in the use of breast-conserving surgery in California. *Am J Public Health.* 2000;90:281–284.
5. Prehn AW, Topol B, Stewart S, et al. Differences in treatment patterns for localized breast cancer among Asian/Pacific Islander women. *Cancer.* 2002;95:2268–2275.
6. Kerner JF, Mandelblatt JS, Silliman RA, et al. Outcomes and preferences for treatment in older women nationwide study. Screening mammography and breast cancer treatment patterns in older women. *Breast Cancer Res Treat.* 2001;69:81–91.
7. Tropman SE, Ricketts TC, Paskett E, et al. Rural breast cancer treatment: evidence from the Reaching Communities for Cancer Care (REACH) project. *Breast Cancer Res Treat.* 1999; 56:59–66.
8. Bradley CJ, Given CW, Roberts C. Race, socioeconomic status, and breast cancer treatment and survival. *J Natl Cancer Inst.* 2002;94:490–496.
9. Eddy DM. Performance measurement: problems and solutions. *Health Aff (Millwood).* 1997;17:7–25.
10. Hewitt M, Simone J, editors. Ensuring quality cancer care. Washington, DC: National Academy Press, 1999.
11. Treatment of early stage breast cancer. NIH Consensus Statement Online 1990 Jun 18–21 [online] 8:1–19. Available at URL: http://consensus.nih.gov/cons/081/081_statement.htm [accessed February 10, 2005].
12. Smith T, Davidson N, Schapira D, et al. American Society of Clinical Oncology 1998 update of recommended breast cancer surveillance guidelines. *J Clin Oncol.* 1999;17: 1080–1082.
13. National healthcare disparities report: summary. February 2004. Agency for Healthcare Research and Quality, Rockville, MD. Available at URL: http://www.ahrq.gov/qual/nhdr03/nhdrsum03.htm [accessed February 10, 2005].
14. Warren JL, Klabunde CN, Schrag D, et al. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care.* 2002;40(8 Suppl):3–18.
15. Bach PB, Guadagnoli E, Schrag D, et al. Patient demographic and socioeconomic characteristics in the SEER-Medicare database. *Med Care.* 2002;40(8 Suppl):19–25.
16. Klabunde CN, Potosky AL, Legler JM, et al. Development of a comorbidity index using physician claims data. *Clin Epidemiology.* 2000;53:1258–1267.
17. Cooper GS, Virnig BA, Klabunde CN, et al. Use of SEER-Medicare data for measuring cancer surgery. *Med Care.* 2002;40(8 Suppl):43–47.
18. Du X, Freeman JL Warren JL, et al. Accuracy and completeness of Medicare claims data for surgical treatment of breast cancer. *Med Care.* 2000;38:719–727.
19. Cooper GS, Zhong Y, Stange KC, et al. Agreement of Medicare claims and tumor registry data for assessment of cancer-related treatment. *Med Care.* 2000;38:411–421.
20. Virnig BA, Warren JL, Cooper GS, et al. Studying radiation therapy using SEER-Medicare-linked data. *Med Care.* 2002; 40(8 Suppl):49–54.
21. Du X, Freeman JL, Goodwin JS. Information on radiation treatment in patients with breast cancer: the advantages of the linked Medicare and SEER data. *Clin Epidemiol.* 1999; 52:463–470.
22. Nattinger AB, Schapira MM, Warren JL, et al. Methodological issues in the use of administrative claims data to study surveillance after cancer treatment. *Med Care.* 2002;40(8 Suppl):69–74.
23. Cuzick J. A Wilcoxon-type test for trend. *Stat Med.* 1985;4: 87–90.
24. Li CI, Malone KE, Daling JR. Differences in breast cancer stage, treatment, and survival by race and ethnicity. *Arch Intern Med.* 2003;163:49–56.
25. Vinh-Hung V, Verschraegen C. Breast-conserving surgery with or without radiotherapy: pooled-analysis for risks of ipsilateral breast tumor recurrence and mortality. *J Natl Cancer Inst.* 2004;96:115–121.
26. Kunkler I, Williams L, Prescott R, King C. Re: Breast-conserving surgery with or without radiotherapy: pooled-analysis for risks of ipsilateral breast tumor recurrence and mortality. *J Natl Cancer Inst.* 2004;96:1255.
27. Chlebowski RT, Chen Z, Anderson GL, et al. Ethnicity and breast cancer: factors influencing differences in incidence and outcome. *J Natl Cancer Inst.* 2005;97:439–448.
28. Schapira MM, McAuliffe TL, Nattinger AB. Underutilization of mammography in older breast cancer survivors. *Med Care.* 2000;38:281–289.
29. Mandelblatt JS, Hadley J, Kerner JF, et al. Patterns of breast carcinoma treatment in older women: patient preference and clinical and physical influences. *Cancer.* 2000;89:561–573.
30. Gilligan MA, Kneusel MS, Hoffman RG, et al. Persistent differences in sociodemographic determinants of breast conserving treatment despite overall increased adoption. *Med Care.* 2002;40:181–189.
31. Isaacs SL, Schroeder SA. Class—the ignored determinant of the nation's health. *N Engl J Med.* 2004;351:1137–1142.
32. Freeman HP. Poverty, culture, and social injustice: determinants of cancer disparities. *CA Cancer J Clin.* 2004;54: 72–77.
33. Bach PB, Cramer LD, Warren JL, et al. Racial differences in the treatment of early stage lung cancer. *N Engl J Med.* 1999;341:1198–1205.
34. Smedley B, Stith A, Nelson AR (eds). Unequal treatment: confronting racial and ethnic disparities in health care. Washington, DC: National Academy Press, 2003.

35. Bach PB, Pham HH, Schrag D, et al. Primary care physicians who treat blacks and whites. *N Engl J Med.* 2004;351:575–584.
36. Margolis ML, Christie JD, Silvestri GA. Racial differences pertaining to a belief about lung cancer surgery: results of a multicenter survey. *Ann Intern Med.* 2003;139:558–563.
37. Li CI, Daling JR, Malone KE. Incidence of invasive breast cancer by hormone receptor status from 1992 to 1998. *J Clin Oncol.* 2003;21:28–34.
38. Hewitt M, Simone J, editors. Enhancing data systems to improve the quality of cancer care. Washington, DC: National Academy Press, 2000.
39. Beach MC, Cooper LA, Robinson KA, et al. Strategies for Improving Minority Healthcare Quality. January 2004. Agency for Healthcare Research and Quality, Rockville, MD. Available at URL: http://www.ahrq.gov/clinic/epcsums/minqusum.htm [accessed May 27, 2005].

Screening Mammography in the American Elderly

Christopher R. Kagay, MD, Christopher Quale, PhD, Rebecca Smith-Bindman, MD

Background: Substantial differences exist in estimates of the proportion of elderly women who undergo screening mammography and the impact of race and ethnicity on mammography usage.

Methods: A representative 5% sample of elderly women living in 11 Surveillance, Epidemiology, and End Results areas from 1991 to 2001 was constructed using Medicare data. Biennial rates of screening mammography (at least one mammogram within each 2-year period) were calculated for overlapping 2-year periods, adjusting to a 2000–2001 age and race distribution. Multivariate repeated-measures logistic regression was used to examine predictors of screening usage.

Results: The sample included 146,669 women. Between 1991 and 2001 the age- and race-adjusted proportion of women aged 65 years and older who underwent at least biennial screening mammography increased from 35.8% to 47.9%. Mammography screening increased for all racial and ethnic groups, but remained significantly higher for non-Hispanic white women as compared with all other groups. The biennial screening rate in 2000–2001 was 50.6% for non-Hispanic white, 40.5% for African-American, 34.7% for Asian-American, 36.3% for Hispanic, and 12.5% for Native-American women. After controlling for age, site, physician access, comorbidities, education, and income, African Americans (odds ratio [OR]=0.80, 95% confidence interval [CI]=0.78–0.83), Asian Americans (OR=0.53, CI=0.51–0.55), Hispanics (OR=0.70, CI=0.67–0.74), and Native Americans (OR=0.37, CI=0.29–0.46) were still all less likely than non-Hispanic white women to undergo screening.

Conclusions: Elderly women undergo significantly less mammography screening than is suggested by self-reported surveys. All groups of non-white women undergo less screening than do white women. The magnitude of the difference in screening rates comparing Asian-American and Hispanic women with white women is especially large; however, other studies have questioned the sensitivity of Medicare data for identifying people of Asian and Hispanic ethnicity. For African-American women, the magnitude of the gap is smaller, but it is of concern that the gap in screening as compared with white women has grown over time.
(Am J Prev Med 2006;31(2):142–149) © 2006 American Journal of Preventive Medicine

Introduction

Breast cancer is the fifth leading killer of American women,[1] and is the second leading cause of cancer deaths, behind lung cancer.[2] Over the past decade, mortality rates from breast cancer among all women have declined substantially.[3] This decrease in mortality is attributable both to improvements in breast cancer therapy and to increases in the rate of screening mammography leading to earlier cancer detection.[3] While there is a lack of firm evidence for the benefits of mammography past age 69,[4–7] current treatment guidelines[6] recommend continued screening for elderly women in the absence of substantial comorbid medical conditions. However, substantial differences exist in the published literature describing the fundamental questions of what fraction of elderly American women actually receive regular screening mammography,[8–15] and whether mammography screening rates differ in the elderly by race and ethnicity.[16–20]

The most widely cited estimates of screening mammography usage come from self-reported data collected by the Behavioral Risk Factor Surveillance System (BRFSS)[9] and the National Health Interview Study (NHIS).[8] These data suggest relatively high levels of screening among the American elderly, with as many as 70% to 80% of women aged 65 to 69 years receiving at least biennial (once every 2 years) screening. Despite the wide reliance on these estimates, numerous analyses[21–27] have cast doubt on the reliability of self-reported estimates of mammography screening, suggesting that studies based on self-reported data may overstate screening rates.

The availability of a large, geographically diverse data set of Medicare claims offers the opportunity to exam-

From the Department of Radiology, Massachusetts General Hospital, Harvard Medical School (Kagay), Boston, Massachusetts; and Radiology Department, University of California-San Francisco (Quale, Smith-Bindman), San Francisco, California

Address correspondence and reprint requests to: Rebecca Smith-Bindman, MD, 1600 Divisadero St, Rm C-250, San Francisco CA 94115-1667. E-mail: Rebecca.smith-bindman@radiology.ucsf.edu. Alternate E-mail: ckagay@partners.org.

© 2006 American Journal of Preventive Medicine • Published by Elsevier Inc.
0749-3797/06/$–see front matter
doi:10.1016/j.amepre.2006.03.029

Table 1. Study population characteristics

	1991–1992		2000–2001	
	n	%	*n*	%
Aged ≥65 and alive throughout 2-year period	98,615	100.0	97,413	100.0
HMO enrollment during 2-year period	(4,279)[a]	−4.3	(15,263)[a]	−15.7
Not enrolled in Part B during period	(4,258)	−4.3	(1,451)	−1.5
Breast cancer survivor	(323)	−0.3	(1,673)	−1.7
Study population	89,755	91.0	79,026	81.1
White, non-Hispanic	76,038	84.7	63,649	80.5
African American, non-Hispanic	6,406	7.1	5,603	7.1
Asian American, non-Hispanic	2,725	3.0	4,671	5.9
Hispanic	1,909	2.1	2,442	3.1
Native American, non-Hispanic	151	0.2	185	0.2
Unknown/other	2,526	2.8	2,476	3.1
Mean age (SD)	76.2	(7.2)	77.4	(7.3)

[a]Numbers in parentheses refer to women excluded from the study for the given indication.
HMO, health maintenance organization; SD, standard deviation.

ine, with an objective and sensitive[28,29] data source, how frequently elderly women undergo screening mammography, and whether racial and ethnic disparities in breast cancer screening persist.

Methods

Data Source

A representative 5% sample of Medicare-eligible women aged ≥65 living in 11 Surveillance, Epidemiology, and End Results (SEER) areas representing 13.9% of the population of the United States (including 12.1% of the African-American population, 30% to 60% of the Asian-American population, 25.0% of the Hispanic population, and 12.5% of the white population)[30] was constructed. Three data sources were used to build this sample: (1) 1991–2001 Medicare denominator and claims information for women with no cancer diagnosis through 1999, (2) 1991–2001 Medicare denominator and claims information for women with any cancer diagnosis through 1999 except for breast cancer or non-Hodgkins lymphoma (NHL), and (3) 1991–2001 Medicare denominator and claims information for women with any breast cancer or NHL diagnosis. It was necessary to combine the three sets of component files because they had been obtained separately. The combination of the files yielded a 5% population-based sample of all women in the 11 SEER areas. Data analysis was conducted in 2003 through 2005 using SAS software, version 8.2 (SAS Institute Inc., Cary NC, 2003).

Inclusion

Women were included who were alive throughout a given calendar year (or 2-year period for analyses focused on 2-year screening), who were aged ≥65 years at the beginning of the period of analysis, who were enrolled in Part B of Medicare for the entire year(s), and who were not enrolled in a risk-based HMO in the year(s) of analysis. Women enrolled in HMOs were excluded because Medicare does not receive claims for their services. For women ultimately diagnosed with breast cancer, information on mammography usage was included only for the period up until 90 days before breast cancer diagnosis because of the difficulty differentiating screening and diagnostic mammography immediately before breast cancer diagnosis.[13] 146,669 women contributed at least one year of data to the study. Selected characteristics of the study population and the number of women excluded by cause are shown in Table 1.

Identification of Mammography

Screening mammography was considered to be coded with current procedural terminology (CPT) 76091 (bilateral diagnostic) or 76092 (bilateral screening).[13] Because of the relatively recent adoption of a mammography screening CPT code by Medicare and the still-frequent use by providers of diagnostic codes for reimbursement of screening mammography,[12,13,31] both codes are included in this study. Two recent analyses[28,29] have compared Medicare claims data with mammography registry data. Both found that Medicare claims are a relatively sensitive measure of mammography screening in the elderly population when compared with mammography registry data; Medicare claims were found to be 85% sensitive[29] on a mammogram-per-mammogram basis and 94% sensitive[28] when capturing any use of screening mammography within a 2-year period. The higher capture rate is relevant to the 2-year analysis used in this paper.

Analysis

Descriptive statistics for biennial use of screening mammography by age, race, and ethnicity were calculated during overlapping 2-year periods from 1991–1992 through 2001–2002. Age-adjusted, race- and ethnicity-adjusted, and age- and race and ethnicity–adjusted rates were calculated based on a 2000–2001 age/race and ethnicity distribution of the overall study population.

Multivariate repeated-measures logistic regression was used to identify predictors of mammography screening. A woman-year data set was constructed using eligibility criteria as above, in which an individual woman could contribute up to 11 years of data depending on her eligibility status. Covariates included year of analysis; SEER site; age; visits within the year of analysis to primary care providers, obstetricians/gynecologists, and ER physicians; inpatient hospitalization in year; comorbid conditions in the year of analysis; and ZIP code–

Table 2. Biennial rates of screening mammography by age and race/ethnicity

	1991–1992	1992–1993	1993–1994	1994–1995	1995–1996	1996–1997	1997–1998	1998–1999	1999–2000	2000–2001
n	89,755	89,821	88,932	87,792	85,925	83,309	80,637	78,911	78,704	79,026
Biennial mammography (unadjusted)	0.38	0.38	0.39	0.40	0.41	0.42	0.44	0.47	0.48	0.48
Biennial mammography (age and race/ethnicity adjusted)	0.36	0.36	0.37	0.38	0.40	0.41	0.43	0.46	0.47	0.48
Biennial mammography by race/ethnicity (age adjusted)										
White	0.38	0.38	0.39	0.40	0.42	0.43	0.46	0.49	0.50	0.51
African American	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.39	0.40	0.41
Asian American	0.25	0.26	0.27	0.28	0.29	0.29	0.30	0.33	0.34	0.35
Hispanic	0.26	0.26	0.26	0.26	0.28	0.29	0.31	0.34	0.37	0.36
Native American	0.17	0.16	0.11	0.13	0.13	0.14	0.16	0.16	0.16	0.12
Biennial mammography by age (race/ethnicity adjusted)										
65–69	0.49	0.49	0.49	0.51	0.53	0.54	0.56	0.59	0.61	0.61
70–74	0.44	0.45	0.46	0.48	0.49	0.50	0.53	0.56	0.57	0.57
75–79	0.36	0.37	0.37	0.38	0.40	0.42	0.45	0.48	0.50	0.50
80–84	0.24	0.25	0.26	0.27	0.28	0.29	0.32	0.35	0.37	0.37
85–89	0.14	0.14	0.14	0.15	0.16	0.16	0.17	0.20	0.21	0.23
≥90	0.05	0.05	0.05	0.05	0.05	0.06	0.07	0.07	0.09	0.09

level socioeconomic status (SES) information. Thirty groups of comorbid conditions were created following the method of Elixhauser et al.,[32] which has been validated[33] as implemented using publicly available SAS code.[34] Aggregate, ZIP code–level measures of SES were used due to the lack of individually available measures of income and education. Following other researchers' use of Medicare data, ZIP code–level median household income,[35,36] and ZIP code–level percentage of non–high school graduates[36] were used to partially control for variations in SES. Multivariate models were run both including and excluding SES and physician/hospital-use variables.

Results

The age- and race-adjusted proportion of women aged ≥65 who underwent at least biennial screening mammography (one mammogram within a 2-year period) increased from 35.8% in 1991–1992 to 47.9% in 2000–2001 (Table 2). Mammography screening increased for all racial and ethnic groups, but remained significantly higher for non-Hispanic white women as compared with all other groups. In 1991–1992, screening rates for non-Hispanic whites were 21%, 49%, and 45% higher than those for African Americans, Asians, and Hispanics, respectively. In 2000–2001, these differentials were 25%, 46%, and 39%. In 2000–2001, screening rates were highest (61.2%) for women aged 65 to 69 and declined with increasing age.

Biennial mammography screening rates by race and ethnicity and year for women aged 65 to 69 are shown in Figure 1. Screening rates for this youngest subgroup of the elderly population were 61.2% overall, 53.5% for African-American women, 46.0% for Asian women, 47.5% for Hispanic women, and 64.5% for white women. Mammography screening rates increased on average 3.1% for African-American women, 3.3% for Asian women, 3.0% for Hispanic women, and 2.4% per year for white women. By contrast, for elderly women aged ≥75, rates increased 2.5% for African-American women, 3.0% for Asian women, 3.7% for Hispanic women, and 4.0% per year on average for white women.

In multivariate models adjusting for year, site, age, comorbidities, physician visits, and SES, African Americans (odds ratio [OR]=0.80, 95% confidence interval [CI]=0.78–0.83), Asian Americans (OR=0.53, CI=0.51–0.55); Hispanics (OR=0.70, CI=0.67–0.74), and Native Americans (OR=0.37, CI=0.29–0.46) were all significantly less likely to undergo screening as compared with non-Hispanic whites (Table 3). Increasing age was associated with less-frequent use of mammography. Women who saw primary care providers in a given year were more likely to undergo mammography (OR=1.73, CI=1.71–

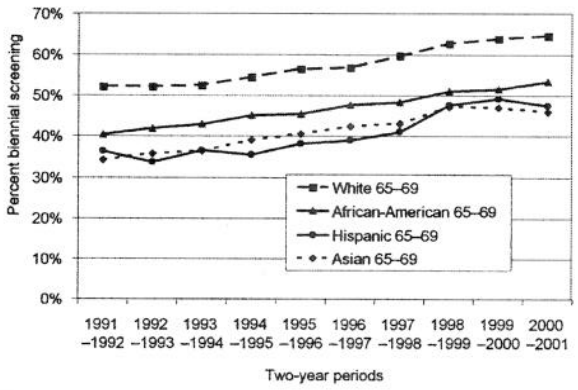


Figure 1. Biennial screening mammography rates by race, among women aged 65 to 69. Biennial mammography rates for women aged 65 to 69, enrolled in Medicare, and living in 11 Surveillance, Epidemiology, and End Results areas, 1991–1992 through 2000–2001.

Table 3. Predictors of screening mammography for each year of study

	Base regression[a]		Base regression plus physician/hospital usage		Base regression, physician/hospital usage, and SES	
	OR	95% CI	OR	95% CI	OR	95% CI
Race/ethnicity						
White		Referent		Referent		Referent
African American	0.65	(0.63–0.67)	0.71	(0.69–0.74)	0.80	(0.78–0.83)
Asian American	0.51	(0.49–0.53)	0.49	(0.47–0.51)	0.53	(0.51–0.55)
Hispanic	0.58	(0.55–0.61)	0.61	(0.58–0.64)	0.70	(0.67–0.74)
Native American	0.26	(0.21–0.33)	0.33	(0.26–0.41)	0.37	(0.29–0.46)
Age group (years)						
65–69		Referent		Referent		Referent
70–74	0.86	(0.85–0.87)	0.87	(0.85–0.88)	0.87	(0.86–0.88)
75–79	0.65	(0.63–0.66)	0.66	(0.65–0.67)	0.66	(0.65–0.68)
80–84	0.40	(0.39–0.41)	0.42	(0.41–0.43)	0.42	(0.41–0.43)
85–89	0.20	(0.19–0.21)	0.21	(0.21–0.22)	0.21	(0.21–0.22)
≥90	0.07	(0.06–0.07)	0.08	(0.07–0.08)	0.08	(0.07–0.08)
Healthcare usage						
No visits[b]		Referent		Referent		Referent
Visit to PCP in year	—	—	1.73	(1.70–1.76)	1.73	(1.71–1.76)
Visit to OB/Gyn in year	—	—	3.18	(3.13–3.24)	3.18	(3.13–3.24)
Visit to ER in year	—	—	0.96	(0.94–0.97)	0.95	(0.94–0.97)
Hospitalized in year	—	—	0.75	(0.74–0.76)	0.75	(0.74–0.77)
ZIP code level, % non–high school grads						
≤10 (most education)		Referent		Referent		Referent
10–14.9	—	—	—	—	0.89	(0.87–0.91)
15–19.9	—	—	—	—	0.82	(0.80–0.84)
20–24.9	—	—	—	—	0.78	(0.76–0.81)
25–49.9	—	—	—	—	0.70	(0.67–0.72)
>50 (least education)	—	—	—	—	0.57	(0.54–0.61)

Notes: ORs and 95% CIs from a repeated-measures logistical regression based on 146,669 women, each contributing between 1 and 11 years of data. In all three regressions, SEER site, 30 comorbid disease categories,[32] and ZIP-code level median income were included. There were significant differences in use by region and comorbidity as described in the Results section. ZIP-code level median income was not significantly related to mammography usage except among women living in ZIP codes with median incomes ≥$60,000, where the OR was 0.95 (0.91 to 0.98).

[a]Base regression adjusted for race/ethnicity, age, SEER site, year, and comorbidities.

[b]Multiple comparisons possible. At least one visit to PCP compared to no visits; at least one visit to OB/Gyn compared to no visits; at least one visit to ER compared to no ER; at least one hospitalization compared to no hospitalizations.

CI, confidence interval; ER, emergency room; OB/Gyn, obstetrician gynecologist; PCP, primary care provider; OR, odds ratio; SEER, Surveillance, Epidemiology, and End Results; SES, socioeconomic status.

1.76), as were women who visited obstetricians/gynecologists (OR=3.18, CI=3.13–3.24). Having been hospitalized decreased the odds of undergoing mammography (OR=0.75, CI=0.74–0.77). Patients with severe comorbidities were in general less likely to be screened, while women with some less-serious comorbidities were more likely to be screened. Women in the Los Angeles and New Mexico SEER areas were least likely to undergo mammography; women in San Francisco were most likely to undergo mammography.

To compare estimates of screening from the present study to widely cited data from the NHIS and BRFSS,[8,9] specific age, race, and ethnicity stratifications not previously published were used, allowing matching to the categories used in this paper, as shown in Figure 2. The NHIS and BRFSS are both based on self-reports. The NHIS is an in-person household interview survey of the civilian, non-institutionalized household population conducted by the National Center for Health Statistics, Centers for Disease Control and Prevention. These data are from the 1998 survey, asking whether a woman had a mammogram for screening purposes within the previous 2 years. The BRFSS is a state-based survey of the non-institutionalized civilian adult population living in households with telephones. Women were asked whether they had had a mammogram, and, if so, when and whether that mammogram had been performed as part of a routine checkup (for screening). The results are for women who reported having a screening mammogram within the past 2 years.[5] For comparability with the NHIS and BRFSS data, the data presented in Figure 2 from the present study do not include age-adjustments. BRFSS data are for 1996–1997. Biennial screening rates based on Medicare data presented here are substantially lower than those reported in the NHIS and BRFSS self-report surveys.

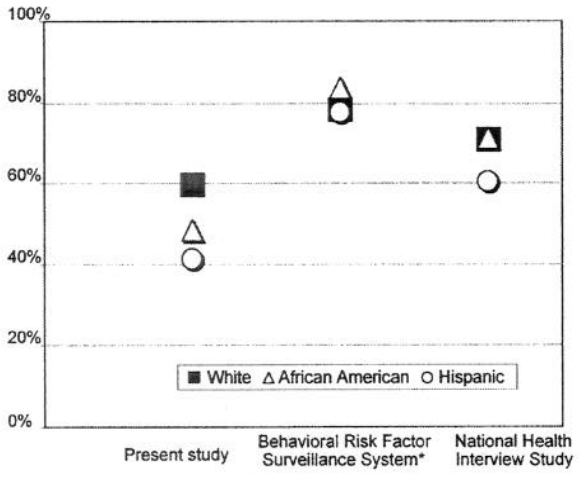


Figure 2. Screening mammography rates for 1997–1998 among women aged 65 to 69 based on the present study, the National Health Interview Study (NHIS), and the Behavioral Risk Factor Surveillance System (BRFSS).

Discussion

The advisability of routine mammography screening in the elderly remains a subject of debate. Most recent analyses and meta-analyses have shown mortality benefit from mammography for older women at least to age 69.[4–7] However, there are few or no randomized trials evaluating the risks and benefits of continued screening in women older than this. Most analyses past age 70 have relied on modeling and have suggested overall benefits from screening mammography in at least some subgroups to age 75 or 79.[37,38] The American Cancer Society, American Geriatrics Society, and U.S. Preventive Services Task Force all recommend continued screening for elderly women unless they have comorbid conditions that would limit their life expectancy or make them poor candidates for breast cancer treatment.[6]

One of the most fundamental questions regarding screening mammography in the elderly remains largely unanswered—how many elderly women are actually screened? The most widely cited estimates of mammography screening suggest that elderly American women are highly screened in aggregate, and that there is no longer a substantial difference by race or ethnicity in the use of screening. For example, self-reported data for 38 states from the BRFSS of 1997[9] suggest that, among women aged 65 to 69 years, 84% of African Americans, 77% of Hispanics, 90% of Asians/Pacific Islanders, and 78% of whites underwent screening mammography within the previous 2 years. Similarly, 1998 data from the NHIS[8] reported that 71%, 60%, and 71% of African-American, Hispanic, and white women, respectively, underwent mammography in the previous 2 years. These data sources are widely cited, including in recent reports by the Institute of Medicine detailing racial and ethnic disparities in diagnosis and treatment[17] and the Cancer Intervention and Surveillance Modeling Network attempting to untangle the effects on mortality of increased screening and improved breast cancer treatments.[3]

However, as shown in Figure 2, the present data suggest that American women are considerably less highly screened than self-reported data have suggested. The discrepancies between Medicare data and self-reported data are particularly stark for some ethnic and racial groups. For example, the present estimates suggest that Asians/Pacific Islanders receive less than half, and Hispanics just more than half of the screening described by the most optimistic self-report study.[9]

The general absence of ethnic or racial disparities in the use of mammography screening based on self-reported data has led authors to conclude either that there are no differences in mammography screening rates by race and ethnicity, or, that if such differences do exist, they are entirely explained by socioeconomic factors[16–18,20,39] despite multiple studies suggesting that self-report studies of mammography substantially overstate the frequency at which women undergo mammography,[21–24] particularly for African-American women.[25–27] In the present study, as shown in Table 3, substantial differences in screening by race and ethnicity persist even when controlling for proxies for health status (30 individual diagnostic groups, hospitalization, and emergency room usage), access to care (primary care provider and obstetrician/gynecologist visits), education, and income. Despite similar coverage for a substantial portion of the cost of mammography in this Medicare-insured population, Figure 1 shows persistent racial and ethnic disparities in mammography through 2001.

These results are suggestive of the possibility that the later stages at presentation among African-American[40] and Hispanic[41,42] women upon diagnosis with breast cancer may be due partly to lower rates of screening mammography. The experience of Asian and Pacific Islander women seems potentially more complicated. Despite having some of the lowest screening rates of any women in the present study, Asian and Pacific Islander[43] women have not been observed to differ significantly from white women in terms of stage at presentation or breast cancer mortality. Of note, Asian and Pacific Islander women are not a homogenous group, and recent analyses have shown Hawaiian,[44–46] Filipina-American,[43–46] and Indian/Pakistani–American[43] women to present with substantially later-stage disease than do white women. The present study could not differentiate among these groups.

A further complicating factor affecting particularly Asian-American and Hispanic women in the present study is the low sensitivity of Medicare data in identifying these ethnic groups, as described in the work of Bach et al.[35] While the sensitivity, specificity, and posi-

tive predictive value are extremely high for the designations white and black in Medicare claims data, exceeding 95% in most cases, the sensitivity for Hispanics, Asians/Pacific Islanders, and Native Americans is considerably lower, with percent sensitivity measurements in the single digits and teens in 1996, rising to 39% and 58% for Hispanics and Asian/Pacific Islanders in 1997 and beyond, and remaining low at 10.9% for Native Americans. Specificity values approached 100% for all groups.[35,47]

As shown in Table 3, including physician/hospital usage in the regression narrows the gap in rates of screening mammography between both African-American and Hispanic and white women. On the other hand, including this proxy for healthcare access does not significantly affect the gap between Asian Americans and whites, perhaps suggesting that issues other than access barriers lead to the particularly low rates seen by this group.

The gap in screening by race and ethnicity seems to be narrowing over time for women aged 65 to 69 in whom the evidence for the benefit of mammography is greatest, but increasing over time for older (aged 75 and above) elderly women. The importance of these trends remains to be seen, as further evidence of the value of mammography screening for the older elderly is established.

There are relatively few other studies using Medicare data to examine mammography screening in the elderly population. One recent study[9] examining the issue of screening rates and race and ethnicity using Medicare claims in Michigan covered a limited geographic area with less ethnic diversity than the current study. Like the present study, that study found substantially lower rates of screening than found in self-report data. Unlike the present study, that study found no significant impact of African-American race as compared to non–African-American race in predicting the use of mammography. However, that study grouped together whites, Asians, Hispanics, and Native Americans into the non–African-American category. Since all those groups have substantially lower screening rates than whites (or African Americans), it is possible that combining them into a single subgroup could have masked real differences between whites and African Americans in the study of Michigan women. Other studies using Medicare claims data have found lower usage by African Americans, including Asch et al.,[48] and a recent study by Bynum et al.,[49] which found that African-American elderly women were screened less than white women, even when controlling for health status.

There are a number of limitations to the present study. Women enrolled in Medicare HMOs were not included. These women may in aggregate be better screened[50] than women outside of HMOs, thus potentially biasing estimated screening rates downward. By 2000–2001, women enrolled in HMO plans constituted approximately 15% of the study population. Thus, even if their screening rates were higher than non–HMO-enrolled women, their exclusion would bias aggregate estimates by only a few percentage points. All bilateral mammograms were included in the study, including those coded with CPT 76091 (bilateral diagnostic). This will have caused the present study to slightly overestimate the use of screening mammography, since approximately 10% of mammograms are done for diagnostic purposes.[8] However, since women with preexisting breast cancer diagnoses were excluded, virtually all of these bilateral "diagnostic" mammograms likely in fact represented screening. For the purposes of conservatism in examining any possible shortfalls in mammography screening, the most-inclusive definition of mammography screening was used.

The fact that some women receive screening mammograms not captured by Medicare claims was not accounted for. A previous analysis involving the present authors[28] suggests that 5.8% of women aged over 65 years receive a mammogram within a given 2-year period that is not captured by Medicare claims. Another analysis,[29] which focused on the sensitivity of Medicare claims for individual mammograms, suggests that Medicare claims capture only 85% of mammograms in the Colorado elderly. That figure underestimates the sensitivity of Medicare for biennial screening; however, since it focuses on Medicare's sensitivity for each individual claim rather than for **any** claim in a 2-year period. The under-capture relevant to the present study is approximately 5.8% for biennial claims.

Another source of error is that SES characteristics were attributable to individuals only at the level of their ZIP code, and this method is imperfect. As would be expected, validation studies suggest that the use of ZIP code–level median income data correctly estimates the overall direction of effect, but may underestimate the magnitude of the effect.[51] The present study failed to show a significant impact of ZIP code–level income data on mammography screening; it is possible that if family income could have been included in a less-aggregated fashion, these results might have differed. Additionally, only elderly women in the 11 SEER areas were studied. Although these areas are widely geographically dispersed, ethnically diverse, and constitute a significant fraction of the U.S. population, they do not represent a statistically random sample of the overall U.S. population.

A final source of error is the designation of race in Medicare data, as outlined above in this section.

Conclusions

This population-based study finds mammography screening rates in the elderly population substantially lower than the most widely cited estimates for elderly women. Analysis of this objective data source shows persistent discrepancies in screening rates for all racial and ethnic groups

when compared with white women. For Asian-American and Hispanic women, the magnitude of the discrepancies in screening is especially large; however, the sensitivity of Medicare claims data for identifying women in these groups is substantially lower than for African-American or white women. For African-American women, the magnitude of the gap is smaller, but it is of concern that the gap has grown over time.

We are grateful to Nancy Breen, PhD, in the Health Services and Economics Branch of the Applied Research Program, Division of Cancer Prevention and Population Science, National Cancer Institute (NCI), and Donald Blackman, PhD, in the Division and Cancer Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion at the Centers of Disease Control and Prevention for providing unpublished data from the NHIS and BRFSS surveys, respectively, for inclusion in this paper.

This study used the linked SEER–Medicare database. The interpretation and reporting of these data are the sole responsibilities of the authors. We acknowledge the efforts of the Applied Research Program, Division of Cancer Prevention and Population Science, NCI; Office of Information Services and Office of Strategic Planning, Health Care Financing Administration; Information Management Services, Inc; and the SEER program tumor registries in the creation of the SEER–Medicare database.

This work was supported in part by the Mt. Zion Health Care Fund Dean's Account, California Breast Cancer Research Fund, U.S. Department of Defense (DAMD179919112), and NCI (K07 CA86032).

No financial conflict of interest was reported by the authors of this paper.

References

1. National Heart, Lung and Blood Institute. Leading causes of death for American women, 2002. Available at: www.nhlbi.nih.gov/health/hearttruth/whatis/index.htm. Accessed January 8, 2006.
2. American Cancer Society. Breast cancer facts & figures 2001–2002. Available at: www.cancer.org/downloads/STT/BrCaFF2001.pdf 2003. Accessed January 8, 2006.
3. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med 2005;353:1784–92.
4. 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;137:347–60.
5. Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography. A meta-analysis. JAMA 1995;273:149–54.
6. Walter LC, Lewis CL, Barton MB. Screening for colorectal, breast, and cervical cancer in the elderly: a review of the evidence. Am J Med 2005;118:1078.
7. Smith RA, Duffy SW, Gabe R, Tabar L, Yen AMF, Chen THH. The randomized trials of breast cancer screening: what have we learned? Radiol Clin North Am 2004;42:793.
8. Breen N, Wagener DK, Brown ML, Davis WW, Ballard-Barbash R. Progress in cancer screening over a decade: results of cancer screening from the 1987, 1992, and 1998 National Health Interview Surveys. JNCI Cancer Spectrum 2001;93:1704–13.
9. Blackman DK, Bennett EM, Miller DS. Trends in self-reported use of mammograms (1989–1997) and Papanicolaou tests (1991–1997)—Behavioral Risk Factor Surveillance System. MMWR Surveill Summ 1999;48:1–22.
10. Van Harrison R, Janz NK, Wolfe RA, Tedeschi PJ, Huang X, McMahon LF Jr. 5-Year mammography rates and associated factors for older women. Cancer 2003;97:1147–55.
11. May DS, Trontell AE. Mammography use by elderly women: a methodological comparison of two national data sources. Ann Epidemiol 1998;8:439–44.
12. Freeman JL, Klabunde CN, Schussler N, Warren JL, Virnig BA, Cooper GS. Measuring breast, colorectal, and prostate cancer screening with Medicare claims data. Med Care 2002;40(suppl 8):IV-36–42.
13. Randolph WM, Mahnken JD, Goodwin JS, Freeman JL. Using Medicare data to estimate the prevalence of breast cancer screening in older women: comparison of different methods to identify screening mammograms. Health Serv Res 2002;37:1643–57.
14. Jazieh AR, Soora I, Mohrmann H. The reality of mammography utilization in the state of Arkansas. J Ark Med Soc 2000;97:132–3.
15. Jazieh AR, Soora I. Mammography utilization pattern throughout the state of Arkansas: a challenge for the future. J Community Health 2001;26:249–55.
16. Lannin DR, Mathews HF, Mitchell J, Swanson MS. Impacting cultural attitudes in African-American women to decrease breast cancer mortality. Am J Surg 2002;184:418–23.
17. Geiger JH. Racial and ethnic disparities in diagnosis and treatment: a review of the evidence and a consideration of causes. In: Smedley BD, Stith AY, Nelson AR, eds. Unequal treatment: confronting racial and ethnic disparities in health care. Washington DC: National Academies Press, 2002:415–54.
18. Qureshi M, Thacker HL, Litaker DG, Kippes C. Differences in breast cancer screening rates: an issue of ethnicity or socioeconomics? J Womens Health Gend Based Med 2000;9:1025–31.
19. Han B, Wells BL, Primas M. Comparison of mammography use by older black and white women. J Am Geriatr Soc 2003;51:203–12.
20. Hegarty V, Burchett BM, Gold DT, Cohen HJ. Racial differences in use of cancer prevention services among older Americans. J Am Geriatr Soc 2000;48:735–40.
21. Suarez L, Goldman DA, Weiss NS. Validity of Pap smear and mammogram self-reports in a low-income Hispanic population. Am J Prev Med 1995;11:94–8.
22. Degnan D, Harris R, Ranney J, Quade D, Earp JA, Gonzalez J. Measuring the use of mammography: two methods compared. Am J Public Health 1992;82:1386–8.
23. Gordon NP, Hiatt RA, Lampert DI. Concordance of self-reported data and medical record audit for six cancer screening procedures. J Natl Cancer Inst 1993;85:566–70.
24. Norman SA, Localio AR, Zhou L, et al. Validation of self-reported screening mammography histories among women with and without breast cancer. Am J Epidemiol 2003;158:264–71.
25. Champion VL, Menon U, McQuillen DH, Scott C. Validity of self-reported mammography in low-income African-American women. Am J Prev Med 1998;14:111–7.
26. Zapka JG, Stoddard A, Maul L, Costanza ME. Interval adherence to mammography screening guidelines. Med Care 1991;29:697–707.
27. Zapka JG, Bigelow C, Hurley T, et al. Mammography use among sociodemographically diverse women: the accuracy of self-report. Am J Public Health 1996;86:1016–21.
28. Smith-Bindman R, Quale C, Chu PW, Rosenberg R, Kerlikowske K. Can Medicare billing claims data be used to assess mammography utilization among women ages 65 and older? Med Care 2006;44:463–70.
29. Mouchawar J, Byers T, Warren M, Schluter WW. The sensitivity of Medicare billing claims data for monitoring mammography use by elderly women. Med Care Res Rev 2004;61:116–27.
30. Surveillance Implementation Group, National Cancer Institute. Cancer surveillance research implementation plan. Appendix B: cancer surveillance in the United States 1999. Available at: http://cancercontrol.cancer.gov/sig/appb.htm. Accessed January 8, 2006.
31. Blustein J. Medicare coverage, supplemental insurance, and the use of mammography by older women. N Engl J Med 1995;332:1138–43.
32. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998;36:8–27.
33. Southern DA, Quan H, Ghali WA. Comparison of the Elixhauser and Charlson/Deyo methods of comorbidity measurement in administrative data. Med Care 2004;42:355–60.
34. Agency for Healthcare Research and Quality. Comorbidity software, 2003. Available at: www.ahrq.gov/data/hcup/comorbid.htm. Accessed August 2003.
35. Bach PB, Guadagnoli E, Schrag D, Schussler N, Warren JL. Patient demographic and socioeconomic characteristics in the SEER–Medicare database applications and limitations. Med Care 2002;40(suppl 8):IV-19–25.

36. Garg PP, Frick KD, Diener-West M, Powe NR. Effect of the ownership of dialysis facilities on patients' survival and referral for transplantation. N Engl J Med 1999;341:1653–60.
37. Kerlikowske K, Salzmann P, Phillips KA, Cauley JA, Cummings SR. Continuing screening mammography in women aged 70 to 79 years: impact on life expectancy and cost-effectiveness. JAMA 1999;282:2156–63.
38. Fracheboud J, Boer R, Draisma G, de Bruijn A, Verbeek A, de Koning H Seventy-five years is an appropriate upper age limit for population-based mammography screening. Int J Cancer 2006;118:2020–25.
39. Thompson B, Coronado GD, Solomon CC, McClerran DF, Neuhouser ML, Feng Z. Cancer prevention behaviors and socioeconomic status among Hispanics and non-Hispanic whites in a rural population in the United States. Cancer Causes Control 2002;13:719–28.
40. Lannin DR, Mathews HF, Mitchell J, Swanson MS, Swanson FH, Edwards MS. Influence of socioeconomic and cultural factors on racial differences in late-stage presentation of breast cancer. JAMA 1998;279:1801–7.
41. Zaloznik AJ. Breast cancer stage at diagnosis: Caucasians versus Hispanics. Breast Cancer Res Treat 1997;42:121.
42. Chen F, Trapido EJ, Davis K. Differences in stage at presentation of breast and gynecologic cancers among whites, blacks, and Hispanics. Cancer 1994;73:2838–42.
43. Li CI, Malone KE, Daling JR. Differences in breast cancer stage, treatment, and survival by race and ethnicity. Arch Intern Med 2003;163:49–56.
44. Hsu J, Glaser S, West D. Racial/ethnic differences in breast cancer survival among San Francisco Bay Area women. JNCI Cancer Spectrum 1997;89:1311–2.
45. Meng L, Maskarinec G, Wilkens L. Ethnic differences and factors related to breast cancer survival in Hawaii. Int J Epidemiol 1997;26:1151–8.
46. Meng L, Maskarinec G, Lee J. Ethnicity and conditional breast cancer survival in Hawaii. J Clin Epidemiol 1997;50:1289–96.
47. Arday SL, Arday DR, Monroe S, Zhang J. HCFA's racial and ethnic data: current accuracy and recent improvements. Health Care Financ Rev 2000;21:107–16.
48. Asch SM, Sloss EM, Hogan C, Brook RH, Kravitz RL. Measuring under-use of necessary care among elderly Medicare beneficiaries using inpatient and outpatient claims. JAMA 2000;284:2325–33.
49. Bynum JPW, Braunstein JB, Sharkey P, Haddad K, Wu AW. The influence of health status, age, and race on screening mammography in elderly women. Arch Intern Med 2005;165:2083–8.
50. Carrasquillo O, Lantigua RA, Shea S. Preventive services among Medicare beneficiaries with supplemental coverage versus HMO enrollees, Medicaid recipients, and elders with no additional coverage. Med Care 2001;39:616–26.
51. Gornick ME, Eggers PW, Reilly TW, et al. Effects of race and income on mortality and use of services among Medicare beneficiaries. N Engl J Med 1996;335:791–9.

Can Medicare Billing Claims Data Be Used to Assess Mammography Utilization Among Women Ages 65 and Older?

Rebecca Smith-Bindman, MD,† Chris Quale, PhD,* Philip W. Chu,* Robert Rosenberg, MD,‡ and Karla Kerlikowske, MD§¶*

Background: Medicare data may be a useful source for determining the utilization of mammography among elderly women, but the accuracy of these data has not been established.

Objective: We determined whether Medicare physician billing claims are an accurate reflection of mammography utilization among women ages 65 and older and whether they can be used to assess the use of screening as compared with diagnostic mammography.

Data Sources: Mammography use was assessed using Medicare billing claims and radiology reports from 2 mammography registries; the San Francisco Mammography Registry and the New Mexico Mammography Registry.

Methods: Completeness of the Medicare data was assessed by comparing mammography use based on Medicare, with radiology reports from the mammography registries, which served as the referent standard. Capture rates for Medicare claims for individual mammograms were examined, and women were characterized as having undergone at least 1 mammogram within each 2-year period based on the Medicare data, and these rates were compared with the referent standard. To determine whether Medicare data can distinguish between screening and diagnostic mammography, we performed a classification analysis using the mammography registries screening/diagnostic designation as the referent standard (dependent variable) and Medicare claim information as the independent/predictor variable. On the basis of the mammogram level classification analysis, women were categorized as having been frequently screened (at least 2 screening mammograms spaced by 12 to 36 months), screened (at least 1 screening mammogram), or not screened.

Subjects: Women ages 65 and older, diagnosed with breast cancer between 1992–1999, who had at least 1 mammogram between 1992–1999 were examined.

Results: A total of 3340 mammograms were obtained in 1371 women between 1992 and 1999. Overall, 83% of mammograms obtained by these women had a corresponding billing claim in Medicare. This increased from 65% in 1992 to 90% in 1999. Of women who underwent at least 1 mammogram during each 2-year period per the referent standard, 94% of women were accurately classified by Medicare claims as also having undergone mammography during the same 2-year period. In multivariable analysis, a mammogram was more likely to be associated with a billing claim over time, for women 80 years or older, and for white and Asian as compared with Hispanic women. Neither socioeconomic status nor screening/diagnostic designation affected the likelihood that a mammogram would be associated with a billing claim. The Medicare data accurately categorized a given mammogram as screening or diagnostic for 87.5% of mammograms. Lastly, there was moderate to substantial agreement in the categorization of women as frequently screened, screened or not screened between the 2 data sets (weighted kappa 0.74, 95% confidence interval 0.70–0.78).

Conclusion: Medicare administrative claims are reliable for assessment of mammography utilization and have become more accurate over time. Medicare claims data also provide a mechanism for designating mammography as screening or diagnostic, which subsequently may allow accurate description of a woman's screening history.

Key Words: screening mammography, utilization of mammography, Medicare, SEER-Medicare, elderly women

(*Med Care* 2006;44: 463–470)

From the *Department of Radiology, University of California, San Francisco; †Department of Epidemiology and Biostatistics, University of California, San Francisco; ‡Department of Radiology, University of New Mexico, Albuquerque, New Mexico; §Department of Medicine, University of California, San Francisco; and ¶General Internal Medicine Section, Department of Veterans Affairs, University of California, San Francisco.

Supported by the NCI (K07 CA86032), Breast Cancer Surveillance Consortium co-operative agreement U01CA63740, and The Department of Defense (DAMD179919112 and DAMD 170010193).

Reprints: 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. E-mail: Rebecca.Smith-Bindman@Radiology.UCSF.edu

Copyright © 2006 by Lippincott Williams & Wilkins

ISSN: 0025-7079/06/4405-0463

Several national efforts have encouraged mammography use among elderly women during the last decade, including the expansion of Medicare coverage to reimburse for screening mammography in 1991. As a result, utilization rates of screening mammography among elderly women have increased.[1,2] For purposes of understanding the quality of cancer screening among elderly women, it is important to understand whether mammography is being used appropriately, by whom, and how often. There are few accurate estimates of the actual utilization rates of mammography

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

among elderly women. The most widely cited estimates for mammography use are based on self-report data collected by the Behavior Risk Surveillance System and the National Health Interview Study. These suggest elderly women in aggregate are close to receiving the recommended biennial mammography (at least 1 mammogram within a 2-year period).[1,3] However, the reliability of these data has not been established because women may overestimate their actual use of mammography.[4–9]

The ideal database to assess mammography utilization among elderly women would be population-based and readily available to allow the rapid and reliable assessment of mammography utilization. If found reliable, Medicare data would be an ideal source to study mammography rates among elderly women.[10–13] Most of elderly women in the United States are covered by this government-sponsored insurance policy and it would give an unbiased estimate of mammography utilization. However, Medicare claims have not yet been shown to be reliable for determination of mammography utilization.[14] Specifically, because beneficiaries may pay out of pocket for mammograms, pay for them through private health insurance, or have them paid for through other programs, using Medicare billing claims data to assess mammography utilization could lead to an underestimation of actual utilization.[4] Only a single study has evaluated whether Medicare billing claims accurately capture mammography examinations.[15]

The purpose of this analysis is to determine whether Medicare physician billing claims data can be used to accurately determine the use of mammography among elderly women, and there are 2 separate aims. First, among a group of elderly women in whom we know mammography was obtained based on radiology physician reports, we determined the percentage of mammograms that had an associated Medicare billing claim. The goal of this first objective is to assess whether Medicare data are accurate for determination of dissemination of mammography. Second, it is important to distinguish screening mammograms (mammograms obtained in asymptomatic women to lead to the earlier detection of breast cancer) from diagnostic mammograms (mammograms obtained to further evaluate a breast symptom or a previous mammogram abnormality). The distinction is important because the majority of women with breast cancer will undergo diagnostic mammography, but this does not equate to having had the opportunity to participation in screening. Thus the second purpose of this analysis is to determine whether Medicare physician billing claims can be used to distinguish between screening and diagnostic mammograms and to determine which women have undergone screening mammography.

METHODS

Mammography use among elderly women was assessed in 2 geographic areas using data from 2 sources. First, data were obtained from 2 mammography registries (the San Francisco Mammography Registry and the New Mexico Mammography Registry) that have participated in the Breast Cancer Surveillance Consortium (BCSC), an NCI-funded consortium of mammography registries in the United States.[16–18] These data served as the referent standard for assessment of mammography utilization. Second, Medicare administrative billing claims and enrollment information were obtained from the Center for Medicare and Medicaid Services as part of the linked Surveillance, Epidemiology and End Results (SEER)–Medicare dataset. This dataset includes information on patients diagnosed with cancer.[19] The mammography registries and Medicare data cannot be readily linked for all women as both datasets have been anonymized. However, each dataset has been linked with their regional SEER program data, and thus we were able to link the mammography registry data (SEER-MR) with the Medicare data (SEER-Medicare) using the SEER program identification codes. The study was thus limited to women diagnosed with breast cancer, because we only were able to link information for these women. Medicare began to reimburse beneficiaries for biennial screening mammography beginning in 1991, and we obtained Medicare records on subjects from 1991 to 1999. The UCSF institutional review board, and each of the SEER program and mammography registries approved the study.

Data Sources

SEER-MR

Data were obtained 2 Mammogram Registries; the San Francisco Mammography Registry, which began collecting data in 1985, and the New Mexico registry, which began collecting data in 1989. Both registries were expanded in 1995 as a result of NIH funding for the Breast Cancer Surveillance Consortium, which allowed them to expand mammography facility-based data collection in their respective areas to include nearly all mammography facilities in the same geographic areas that they started data collection before 1995. Although each registry has become increasingly population based over time, the characteristics of those covered did not change over time. Each registry tries to capture all mammograms in their region, limited primarily by practical concerns. These data provided a way to assess mammography use among elderly women based on detailed medical records. These 2 registries were included because both link their mammography data with their regional SEER program.

SEER-Medicare

The SEER-Medicare dataset provided a convenient source to evaluate Medicare claims information. The SEER-Medicare database is a collaborative effort of the National Cancer Institute, the SEER program, and the Center for Medicare and Medicaid Services to create a large population-based source of information for cancer related epidemiologic and health services research.[19] SEER-Medicare data combines cancer information from population based cancer registries with clinical information derived from the Medicare data, which includes billing claims for physician services including mammography. The SEER-Medicare data we obtained for this study includes women in the SEER areas receiving a diagnosis of Breast Cancer between the years 1991 and 1999 and includes all billing claims for these women from 1991 onwards.

© 2006 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

Data Linkage

Information from the SEER-MR and SEER-Medicare data sources were linked for women who resided in the same geographic areas. The women were matched using the tumor ID variable in each data set by each mammography registry.

Subjects

We used information from both SEER-Medicare and SEER-MR to assemble our cohort of study subjects. Because we needed subjects with 1 year of Medicare eligibility before any mammogram to determine variables we used for the classification analysis, we limited our study to women ages 66 and older at diagnosis of breast cancer (so that would have at least 1 year of Medicare eligibility prior to diagnosis); who were diagnosed with cancer between 1992 and 1999 (so that they had a least 1 year to document and characterize mammography before their cancer diagnosis); who had at least 1 mammogram in one of the mammography registries between 1992 and 1999; and who lived in 1 of the 2 geographic areas (n = 2244.) As part of Medicare, women may be enrolled in risk-based HMO plans. Because Medicare does not receive billing claims for physician services for these women, we did not include women who had any months of HMO coverage. Additionally, as Medicare does not receive billing claims during periods where a woman is not Part B eligible, we excluded women who were not Part B eligible for any given month. From the original 2244 eligible women, 568 (25.3%) had all of their mammographic history excluded for either HMO enrollment or non Part B eligibility; most (89.3%) were removed because of mammograms during months of HMO enrollment, 10.5% because of non-Part B enrollment, and 1.2% for both. This yielded a population of 1,676 (74.7%) women who had 1 or more eligible mammograms, and who received those mammograms during years of study. For this group of women, we included all mammograms obtained from 1992 to 1999 that occurred before their diagnosis of cancer. We limited the mammograms to those that the women had received before their breast cancer diagnosis because the completeness of Medicare billing claims for mammography could vary by whether or not women had yet received a diagnosis of breast cancer. These criteria resulted in a population of 1371 women contributing 3340 registry mammograms, and these mammograms were obtained from 1992 to 1999.

Assessment of Mammography Utilization and Other Variables

SEER-MR

The mammography registries include the date of all mammographic examinations, patient survey results,[20] radiologist designation of the mammogram as screening or diagnostic, and mammographic interpretation. We considered a mammogram to have occurred if there was a record of a mammographic examination with a corresponding physician interpretation.

SEER-Medicare

Mammography claims were taken from the outpatient facility file and the physician's claim file.[19] We searched for all claims with 1 of 3 CPT codes: 76090 (Mammography, unilateral), 76091 (Mammography, bilateral) or 76092 (Screening Mammography, bilateral.)[10–13,21–25] We included both screening and diagnostic billing codes because Medicare billing claim codes may not reliably distinguish between screening and diagnostic examinations,[14,24] and because our referent standard based on the mammography registries included both screening and diagnostic examinations.[17,18] When we encountered duplicate claims (more than 1 mammogram for the same woman occurring on the same day but from the different claim sources), we only counted one of those mammograms to avoid overestimating Medicare mammography usage. Thus, at most, a single mammogram was compared for any given day in the study period. Overall, 59.8% of the mammography claims occurred in the physician claim file but not in the outpatient facility file, 3.5% of the claims were in the outpatient file but not in the physician claims, and 36% of claims were in both files.

Age was stratified into 3 age groups, 65–69, 70–79, 80+, and race and ethnicity was grouped as Hispanic and non-Hispanic white, black, and Asian/Pacific Islander using the SEER race "recode B" variable.[26] SES was determined using a combination of median neighborhood income based on census tract and zip code level variables.[26]

Analysis

The completeness of the Medicare data (capture rate) was assessed using as the referent standard all mammograms that were documented in a mammography registry. Using this referent standard, we determined the percentage of mammograms that were documented in the Medicare data. We considered mammograms from the 2 different sources to match if the ID variables matched and the dates from the 2 datasets were within 1 week from each other, although the majority (more than 92%) of the matches were on the same day. We calculated unadjusted Medicare capture rates by registry, age, race, ethnicity, year, and median income and stratified them by whether mammograms were screening or diagnostic. We used the Cochran–Mantel–Haenzel Test for differences in capture rates between screening and diagnostic examinations, controlling for registry, age, race, ethnicity, year, and median income. To estimate what patient variables predict the likelihood of a mammogram being billed to Medicare, all of the above variables were included in a multivariable generalized estimating equation logistic model[27] that accounted for within-woman correlation. All analyses were done using the SAS System version 8.2.[27] Fitting the multivariable model allowed us to examine the differential effects of age, race, ethnicity, year, income and whether an examination was screening or diagnostic on the probability of there being a record of a given mammogram in Medicare.

For the preceding analysis, we calculated Medicare capture rates at the mammogram level. From a health care policy perspective, most professional societies and government agencies recommend biennial mammography for elderly women.[2] To do this, a common way to determine whether women undergo mammography at least once every 2 years is to ask the question, "did you have a mammogram in the last 2 years." Thus, we wanted to ascertain how well

© 2006 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

2-year mammography (ie, at least 1 mammogram every 2 years) is captured in the Medicare claims. The goal of this analysis was to determine whether Medicare data can accurately measure prevalence of mammography examinations within a 2-year period. Thus for this analysis, each 2-year period is considered a separate unit of evaluation. To assess whether 2-year screening is accurately captured by Medicare, within each 2-year period (1992–1993; 1993–1994; etc) we calculated the percentage of women who were classified as having undergone mammography based on the Medicare data, and compared this to the classification based on the referent mammography registry data. For this analysis we included all bilateral mammograms (screening and diagnostic). However, the 2-year screening rates were calculated during the years prior to each woman's breast cancer diagnosis, when none of the women were diagnosed with breast cancer, and where the majority of mammograms were likely obtained for screening.[28]

For the second aim, there were 2 separate components. First, on the mammogram level, we determined how often an individual mammogram was correctly assessed based on the Medicare data as a screening or a diagnostic examination, using the mammography registry data as the referent standard. We needed to restrict the classification analysis to mammograms that matched on the same day (2563 mammograms) because some of the predictor variables (based on Medicare) included consideration of the timing between mammograms, where the day the mammogram occurred was important. Second, on the woman level, we determined whether women could be accurately classified using the Medicare claims data as screened (1 screening mammogram before cancer diagnosis) frequently screened (at least 2 screening mammograms spaced by 12 to 36 months that occurred prior to cancer diagnosis), or not screened with mammography during the study period. To give each woman the opportunity to be characterized as frequently screened, we required that the women we included in this validation analysis have at least 3 years of precancer observation time in which to assess their screening, which reduced the number of women who could be evaluated to 676. A weighted and unweighted kappa statistic was used to estimate the agreement between the 2 data sets in assessing women as frequently screening, screened and not screened using the Cicchetti-Allison formulation.[29] When calculating the weights for the weighted kappa, we used a linear model, resulting in weights 0.5 and 0 for categories 1 and 2 steps removed, following the formula of Cicchetti-Allison. The characterization of women as frequently screened, screened or not screened relied on the assessment of each mammogram as screening or diagnostic, and the mammography registry data provided the referent standard. This involved first performing a classification tree analysis[30] on the matched mammography registry/Medicare mammograms. Using a tree model allowed us to use the data objectively to determine the optimal rule to classify Medicare mammograms as screening or diagnostic based on Medicare claims data. We then used the fitted tree to generate the predicted woman level screening history based on Medicare. This was compared with the "true" screening history as determined by the Mammography registry data.

For the mammogram level classification analysis, variables generated from the Medicare claims were the independent variables, and the screening/diagnostic designation (the "Gold Standard") of the mammography registry was the dependent variable. For the referent standard, a mammogram was considered diagnostic if it was preceded by another mammogram within the proceeding 9 months, or if the radiologist called the mammogram diagnostic. We considered several variables generated from the Medicare claims to help characterize a given mammogram as screening or diagnostic, including previous mammogram within 9 months, any screening mammography code used, breast cancer related diagnosis or procedure code preceding a mammogram (no code, breast malignancy, patient symptoms, breast diagnostic procedure such as an ultrasound, breast biopsy or breast surgery), number of inpatient/outpatient claims the month before the mammogram, time between mammogram and breast cancer diagnosis, as well as patient age, race and ethnicity. The final fitted tree was reduced from the full tree using cross validation, a technique that "holds back" a fraction of the data to evaluate generated subtrees. Minimization of misclassification of mammograms was used as the reduction (pruning) criteria.[31] In the final fitted tree, patient age,

TABLE 1. Characteristics of Women Included in this Report

	n	Percent
Mean no. mammograms per woman	2.4	
Registry		
A	1004	73.2
B	367	26.8
Age, years		
65–69	205	15.0
70–79	801	58.4
80+	365	26.6
Race/ethnicity		
White	1129	82.4
Black	23	1.7
Asian/Pacific Islander	45	3.3
Hispanic	155	11.3
Other/unknown	19	1.4
Year		
1992	85	6.2
1993	116	8.5
1994	135	9.9
1995	136	9.9
1996	176	12.8
1997	214	15.6
1998	236	17.2
1999	273	19.9
Mean income of community of residence		
<$30,000	248	18.1
$30,000 to $50,000	654	47.7
≥$50,000	412	30.1
Not known	57	4.2

© 2006 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

race and ethnicity and number of inpatient/outpatient claims the month prior to the mammogram did not help to discriminate between a screening or diagnostic examination and were not included. The tree analysis was done using the R statistical language version 1.9.[32] All other analyses were performed by using the SAS system, version 8.2.[27]

RESULTS

A total of 3340 mammograms were obtained on 1371 women ages 66 and older between 1992 and 1999. The characteristics of the women included are provided in Table 1. Overall, 83% of mammograms had a corresponding billing claim in Medicare (screening mammograms 81%, diagnostic mammograms 86%). The capture rates by registry, age, race, ethnicity, year, and median community income are provided in Table 2 as overall rates and stratified by whether examinations were screening or diagnostic. The capture rates increased substantially and significantly over time, which occurred for both screening and diagnostic examinations. Among screening examinations, the capture rate increased from 67% in 1992 to 94% in 1999, and among diagnostic examinations, the capture rate increased from 59% in 1992 to 87% to 1999, Table 2. Capture rates varied by age, and were highest for women age 80 and older. The capture rates also

TABLE 2. The Percent of Mammograms Among Women Ages 66 and Older With an Associated Medicare Billing Claim, by Registry, Age, Race, Ethnicity, Year, and Median Community Income (Overall Capture Rate and Capture Rates Stratified by Whether Examinations Were Screening or Diagnostic Are Shown. The Proportion Reflects Crude Capture Rates.)

	Percent of Mammograms with an Associated Medicare Billing Claim							
	Overall			Stratified by Whether Mammogram Screening or Diagnostic				
	Overall			Screening		Diagnostic		
	%	n	P*	%	n	%	n	P†
Registry								
A	80.8	2547		79.8	1674	82.7	873	
B	88.3	793		84.1	502	95.5	291	
			<0.0001					0.0003
Age, years								
65–69	79.6	858		78.7	596	81.7	262	
70–79	82.1	1884		80.2	1231	85.6	653	
80+	88.3	598		86.3	349	91.2	249	
			0.0001					0.0006
Race/ethnicity								
White	83.3	2788		81.8	1833	85.9	955	
Black	87.8	49		80.7	31	100.0	18	
Asian/Pacific Islander	89.8	88		84.6	52	97.2	36	
Hispanic	77.3	370		73.8	229	82.9	141	
Other/unknown	62.2	45		61.3	31	64.3	14	
			<0.0001					0.0002
Year								
1992	65.1	355		66.8	274	59.3	81	
1993	66.1	433		64.7	329	70.2	104	
1994	81.4	381		77.7	291	93.3	90	
1995	87.6	442		84.9	326	94.8	116	
1996	84.6	475		82.6	310	88.5	165	
1997	91.1	492		91.9	298	89.7	194	
1998	91.6	442		94.5	217	88.9	225	
1999	90.0	320		93.9	131	87.3	189	
			<0.0001					0.24
Mean income of community of residence								
<$30,000	80.1	542		77.3	344	84.9	198	
$30,000 to $50,000	81.7	1685		82.0	662	92.8	332	
≥$50,000	85.6	994		81.3	1100	82.4	585	
Not known	80.7	119		77.1	70	85.7	49	
			0.0186					0.0003
Overall	82.54	3340		80.74	2176	85.9	1164	0.0002

*χ^2 test of differences in Medicare capture rates by registry, age, race, ethnicity, year, and median income of community.
†Cochran-Mantel-Haenzel Test for differences in capture rates between screening and diagnostic examinations, controlling for registry, age, race, ethnicity, year and income.

© 2006 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

varied by race and the capture rates were highest for Asian women. Significant differences existed in the capture rates between screening and diagnostic exams after adjusting individually for registry, age, race, ethnicity and income, but not for year.

In the multivariable analysis, a mammogram was more likely to be associated with a billing claim in the most recent years of the study (with a fairly consistent increase over time), for women age 80 and older, and was less likely to be associated with a billing claim in Hispanic women (Table 3). Interestingly, after adjusting for other variables, neither socioeconomic status nor screening/diagnostic designation affected the likelihood that a mammogram would be associated with a billing claim.

To evaluate whether 2-year mammography rates are accurately captured by the Medicare, we examined how well Medicare claims captured 2-year mammography rates when the mammography registry documented a woman had at least 1 mammogram in that 2-year period. A total of 94.3% of 2-year screening intervals were accurately classified by Medicare claims, and this ranged from 93.8% in 1992 to 94.6% in 1997.

The algorithm that was used for characterizing mammograms as screening or diagnostic examinations is provided in Figure 1. The Medicare data accurately categorized a given mammogram as screening or diagnostic for 88% of mammograms (2268/2593); 87% of screening mammograms were correctly classified (1580/1822), and 89% (688/771) of diagnostic mammograms were correctly classified. The second validation of classification was on the woman level. Overall 85% of the women were correctly classified (572/676), with moderate to substantial agreement in the categorization of women as frequently screened, screened or not screened between the 2 data sets (weighted kappa 0.74 (95% confidence interval 0.70, 0.78); unweighted kappa 0.69 (95% confidence interval 0.64, 0.74; Table 4).

TABLE 3. Influence of Registry, Age, Race, Ethnicity, Year, Median Income, and Type of Examination Screening or Diagnostic on the Likelihood That a Mammogram Will Have an Associated Billing Claim in Medicare

	Odds Ratio	95% CI	*P*
Registry			
A		Referent	
B	1.08	0.72–1.61	0.71
Age, years			
65–69		Referent	
70–79	1.27	0.89–1.81	0.18
80+	1.68	1.11–2.55	0.01
Race/ethnicity			
White		Referent	
Black	0.88	0.32–2.41	0.80
Asian/Pacific Islander	1.16	0.42–3.20	0.77
Hispanic	0.61	0.43–0.87	0.007
Other/unknown	0.35	0.16–0.78	0.001
Year			
1992		Referent	
1993	0.92	0.67–1.25	0.58
1994	2.10	1.46–3.02	<0.0001
1995	3.00	2.10–4.29	<0.0001
1996	2.42	1.72–3.43	<0.0001
1997	4.22	2.83–6.29	<0.0001
1998	5.09	3.28–7.90	<0.0001
1999	4.47	2.77–7.22	<0.0001
Mean income of community of residence			
<$30,000	0.84	0.56–1.26	0.50
$30,000 to $50,000	0.93	0.66–1.30	0.66
≥$50,000		Referent	
Not known	0.84	0.42–1.78	0.63
Type of examination			
Diagnostic		Referent	
Screening	0.97	0.81–1.17	0.78

DISCUSSION

Medicare physician claims can be used to determine whether women have undergone mammography as most mammograms (83%) obtained among elderly women between 1992 and 1999 had a corresponding Medicare billing claim. Medicare capture-rates increased substantially over time, and by 1999, 90% of mammograms among elderly women had a billing claim in Medicare, suggesting that Medicare has become a more reliable reflection of mammography use in recent years. Similarly, these data can be used to determine whether women have undergone at least 1 mammogram within 2-year periods, as 94% of women were correctly classified using the Medicare claims with respect to 2-year mammography rates. Additionally, the Medicare claims could be used to determine whether women had undergone screening or diagnostic evaluation fairly accurately. Overall, 88% of examinations were correctly classified as screening or diagnostic compared with a referent standard. We did not find a large percentage of missing mammography claims, nor large differences by age, race, ethnicity or median community income, although Hispanic women were less likely than white women to have a Medicare billing claim, and women age 80 and older were more likely to have a Medicare billing claim.

What are the implications of this study? Mammography use as assessed by self-report has been found to be very high, and up to 80% of women older than age 65 have reported biennial screening mammography use.[1,3] However, these high self-report rates have not been confirmed with administrative claims.[12,22] The results of this study suggest administrative billings claims capture a large percentage of examinations, and that Medicare data are a fairly reliable method for assessment of mammography utilization among elderly women. These findings support the use of Medicare claims to assess population trends in the use of mammography. Overall, approximately 88% of mammograms are obtained for screening purposes.[28] If the total number of mammograms among elderly women is assessed using Medicare data including the billing codes we have used, than this number can be adjusted downward to determine the number of mammograms obtained for screening.[28]

© 2006 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

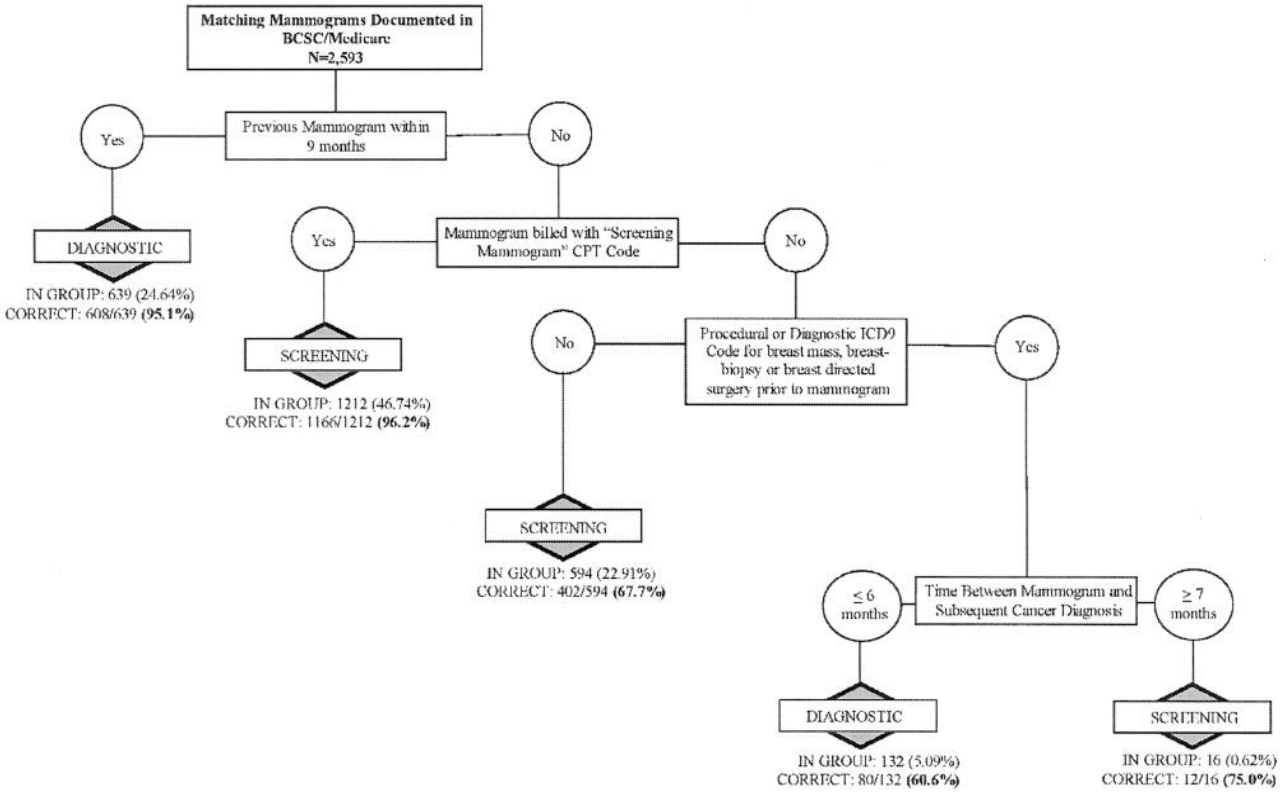


FIGURE 1. Method used to characterize Medicare mammograms as screening or diagnostic, based on results of the CART analysis.

To study patient outcomes of mammography, it is important to be able to distinguish screening from diagnostic evaluations. Virtually all women with breast cancer undergo diagnostic evaluation with mammography (either because of a breast symptom that led to a diagnostic evaluation or because of a screening mammogram that led to a diagnostic evaluation.) Thus it is important to differentiate screening from diagnostic examinations to determine which women were exposed to screening mammography. We have shown that the Medicare data provide detailed information that can be used to distinguish screening from diagnostic examinations. Thus Medicare data may be useful for categorizing women's screening history and for evaluating breast cancer outcomes among women who have undergone screening mammography.

TABLE 4. Characterization of Women as Frequently Screened, Screened and Not Screened With Mammography Based on Mammography Registries and Medicare Billing Claims

	Mammography Registry					
Medicare	**Frequently Screened**		**Screened**		**Not Screened**	
Frequently screened	420	0.62	13	0.02	0	0.00
Screened	37	0.05	128	0.19	46	0.07
Not screened	0	0.00	8	0.01	24	0.04

We included both physician claims for mammography and outpatient facility claims for mammography. Most claims, however, were found in the physician claims file. Including claims from the outpatient file contributed only 3.5% additional mammograms. Thus, it may not be worth the effort to use outpatient facility claims as a tool to identify mammography use, given their relatively limited additional benefit.

This report addresses the accuracy of Medicare claims for assessing mammography utilization. It does not assess the accuracy of mammography or the benefit of mammography among elderly women. Unfortunately, Medicare data cannot be used to assess mammography utilization among women age 50–64, where the evidence regarding the effectiveness of mammography is greatest.[2]

The major strength of this report is that it compares mammography use as assessed by Medicare administrative billing claims, with patient specific medical records including characterization of mammograms as screening or diagnostic examinations. Medicare records provide a readily available and timely method to measure mammography screening rates that are free of recall bias, and we have demonstrated that they are a reliable method to assess mammography utilization. There are several limitations of this study. We did not

© 2006 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

include women enrolled in HMO plans. Unfortunately the use of physician services cannot be ascertained in Medicare beneficiaries enrolled in HMO-types of plans, and thus mammography usage among women enrolled in these types of plans cannot be assessed. We looked at billing claims only in women who were diagnosed with breast cancer. However, the mammograms that we included for these women were only those prior to their diagnosis of cancer, and we suspect these results are fairly generalizable to women who do not have breast cancer. We assessed screening patterns over a relatively short period of time, and did not evaluate whether Medicare claims can accurately predict patterns of mammography screening over longer periods of time. Lastly, we looked at only 2 geographic areas, and even within those areas there was variability in the capture rates. However, we did not find a large difference in the capture rates, and in the multivariate analysis this difference was not significant.

In summary, Medicare data are accurate for assessment of mammography screening, particularly in the more recent years of the study. These data can be reliably used for health services research that focus on mammography, and breast cancer outcomes associated with screening.

ACKNOWLEDGMENTS

This study used the linked SEER-Medicare database. The interpretation and reporting of these data are the sole responsibilities of the authors. We acknowledge the efforts of the Applied Research Branch, Division of Cancer Prevention and Population Science, National Cancer Institute; The office of Information Services and the Office of Strategic Planning, HCFA; Information Management Services, Inc; and the SEER program tumor registries in the creation of the SEER-Medicare database.

REFERENCES

1. Breen N, Wagener DK, Brown ML, et al. Progress in Cancer Screening Over a Decade: Results of Cancer Screening From the 1987, 1992, and 1998 National Health Interview Surveys. *JNCI Cancer Spectrum*. 2001; 93:1704–1713.
2. Fletcher SW, Elmore JG. Clinical practice. Mammographic screening for breast cancer. *N Engl J Med*. 2003;348:1672–1680.
3. Blackman DK, Bennett EM, Miller DS. Trends in self-reported use of mammograms (1989–1997) and Papanicolaou tests (1991–1997)–Behavioral Risk Factor Surveillance System. *MMWR CDC Surveill Summ*. 1999;48:1–22.
4. Degnan D, Harris R, Ranney J, et al. Measuring the use of mammography: two methods compared. *Am J Public Health*. 1992;82:1386–1388.
5. Gordon NP, Hiatt RA, Lampert DI. Concordance of self-reported data and medical record audit for six cancer screening procedures. *J Natl Cancer Inst*. 1993;85:566–70.
6. IOM. *Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis*. Washington, DC: IOM; 2004.
7. Norman SA, Localio AR, Zhou L, et al. Validation of Self-reported screening mammography histories among women with and without breast cancer. *Am J Epidemiol*. 2003;158:264–271.
8. Zapka JG, Bigelow C, Hurley T, et al. Mammography use among sociodemographically diverse women: the accuracy of self-report. *Am J Public Health*. 1996;86:1016–1021.
9. Champion VL, Menon U, McQuillen DH, et al. Validity of self-reported mammography in low-income African-American women. *Am J Prev Med*. 1998;14:111–117.
10. Randolph WM, Goodwin JS, Mahnken JD, et al. Regular mammography use is associated with elimination of age-related disparities in size and stage of breast cancer at diagnosis. *Ann Intern Med*. 2002;137:783–790.
11. Freeman JL, Goodwin JS, Zhang D, et al. Measuring the performance of screening mammography in community practice with Medicare claims data. *Women Health*. 2003;37:1–15.
12. Harrison RV, Janz NK, Wolfe RA, et al. 5-Year mammography rates and associated factors for older women. *Cancer*. 2003;97:1147–1155.
13. Jencks SF, Cuerdon T, Burwen DR, et al. Quality of medical care delivered to Medicare beneficiaries: A profile at state and national levels. *JAMA*. 2000;284:1670–1676.
14. Freeman JL, Klabunde CN, Schussler N, et al. Measuring breast, colorectal, and prostate cancer screening with medicare claims data. *Med Care*. 2002;40:IV36–IV42.
15. Mouchawar J, Byers T, Warren M, et al. The sensitivity of Medicare billing claims data for monitoring mammography use by elderly women. *Med Care Res Rev*. 2004;61:116–127.
16. Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. *Am J Roentgenol*. 1997;169:1001–1008.
17. Barlow W, Lehman C, Zheng Y, et al. Performance of diagnostic mammography for women with signs or symptoms of breast cancer. *J Natl Cancer Insti*. 2002;94:1151–1159.
18. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Intern Med*. 2003;138:168–175.
19. Warren J, Klabunde C, Schrag D, et al. Overview of the SEER-Medicare data: content, research applications and generalizability to the United States elderly population. *Med Care*. 2002;2002:3–18.
20. Breast Cancer Surveillance Consortium Web Site. Available at: http://www.breastscreening.cancer.gov/elements.html#questionnaires. Accessed February 13, 2006.
21. Blustein J. Medicare coverage, supplemental insurance, and the use of mammography by older women. *N Engl J Med*. 1995;332:1138–1143.
22. May DS, Trontell AE. Mammography use by elderly women: a methodological comparison of two national data sources. *Ann Epidemiol*. 1998;8:439–444.
23. Smith-Bindman R, Kerlikowske K, Gebretsadik T, et al. Is screening mammography effective in elderly women? *Am J Med*. 2000;108:112–119.
24. Randolph WM, Mahnken JD, Goodwin JS, et al. Using Medicare data to estimate the prevalence of breast cancer screening in older women: comparison of different methods to identify screening mammograms. *Health Serv Res*. 2002;37:1643–1657.
25. Welch HG, Fisher ES. Diagnostic testing following screening mammography in the elderly [see comments]. *J Natl Cancer Inst*. 1998;90:1389–1392.
26. Bach PB, Guadagnoli E, Schrag D, et al. Patient demographic and socioeconomic characteristics in the SEER-Medicare database applications and limitations. *Med Care*. 2002;40:IV19–IV25.
27. SAS version 8.2. Cary, NC: SAS Institute Inc.; 2003.
28. Brown ML, Fintor L. U.S. screening mammography services with mobile units: results from the National Survey of Mammography Facilities. *Radiology*. 1995;195:529–532.
29. Cicchetti D, Allison T. A new procedure for assessing reliability of scoring EEG sleep recordings. *American Journal of EEG Technology*. 1971;11.
30. Breiman L. *Classification and Regression Trees*. Belmont, CA.: Wadsworth International Group c. 1984;1984.
31. Chambers JM, Hastie TJ. Statistical models in S. pTree-based models. In: Clark LA, Pergibon D, eds. *Wadsworth & Brooks/Cole Computer Science Series*. Pacific Grove, CA: Wadsworth & Brooks/Cole Advanced Books & Software; 1992:377–419.
32. The R Project for Statistical Computing. Available at: http://www.r-project.org. Accessed February 13, 2006.

© 2006 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

Annals of Internal Medicine | ARTICLE

Does Utilization of Screening Mammography Explain Racial and Ethnic Differences in Breast Cancer?

Rebecca Smith-Bindman, MD; Diana L. Miglioretti, PhD; Nicole Lurie, MD, MSPH; Linn Abraham, MS; Rachel Ballard Barbash, MD, MPH; Jodi Strzelczyk, PhD; Mark Dignan, PhD; William E. Barlow, PhD; Cherry M. Beasley, MS, RNCS; and Karla Kerlikowske, MD

Background: Reasons for persistent differences in breast cancer mortality rates among various racial and ethnic groups have been difficult to ascertain.

Objective: To determine reasons for disparities in breast cancer outcomes across racial and ethnic groups.

Design: Prospective cohort.

Setting: The authors pooled data from 7 mammography registries that participate in the National Cancer Institute–funded Breast Cancer Surveillance Consortium. Cancer diagnoses were ascertained through linkage with pathology databases; Surveillance, Epidemiology, and End Results programs; and state tumor registries.

Participants: 1 010 515 women 40 years of age and older who had at least 1 mammogram between 1996 and 2002; 17 558 of these women had diagnosed breast cancer.

Measurements: Patterns of mammography and the probability of inadequate mammography screening were examined. The authors evaluated whether overall and advanced cancer rates were similar across racial and ethnic groups and whether these rates were affected by the use of mammography.

Results: African-American, Hispanic, Asian, and Native American women were more likely than white women to have received inadequate mammographic screening (relative risk, 1.2 [95% CI, 1.2 to 1.2], 1.3 [CI, 1.2 to 1.3], 1.4 [CI, 1.3 to 1.4], and 1.2 [CI, 1.1 to 1.2] respectively). African-American women were more likely than white, Asian, and Native American women to have large, advanced-stage, high-grade, and lymph node–positive tumors of the breast. The observed differences in advanced cancer rates between African American and white women were attenuated or eliminated after the cohort was stratified by screening history. Among women who were previously screened at intervals of 4 to 41 months, African-American women were no more likely to have large, advanced-stage tumors or lymph node involvement than white women with the same screening history. African-American women had higher rates of high-grade tumors than white women regardless of screening history. The lower rates of advanced cancer among Asian and Native American women persisted when the cohort was stratified by mammography history.

Limitations: Results are based on a cohort of women who had received mammographic evaluations.

Conclusions: African-American women are less likely to receive adequate mammographic screening than white women, which may explain the higher prevalence of advanced breast tumors among African-American women. Tumor characteristics may also contribute to differences in cancer outcomes because African-American women have higher-grade tumors than white women regardless of screening. These results suggest that adherence to recommended mammography screening intervals may reduce breast cancer mortality rates.

Ann Intern Med. 2006;144:541-553. www.annals.org
For author affiliations, see end of text.

Breast cancer mortality rates in the United States began to decrease in the 1990s (1) because of increased use of screening mammography and improved breast cancer treatment (2, 3). However, these decreases have primarily benefited non-Hispanic white women, whereas the mortality rate for breast cancer in African-American women changed little (1).

Although racial and ethnic differences in breast cancer mortality rates have been consistently documented (1, 4–9), reasons for the persistence of these differences have been difficult to ascertain (10). Possible explanations include differences in biological characteristics of tumors (11–13); patient characteristics, such as obesity, that may affect prognosis; mammography use (14, 15); timeliness and completeness of breast cancer diagnosis and treatment (16, 17); social factors, such as education, literacy, and cultural beliefs; and economic factors, such as income level and health insurance coverage, that might affect a patient's access to and choices for breast cancer screening and treatment (18–22). Stage at diagnosis, the strongest predictor of breast cancer survival (23), is proportionally higher in all non-Asian minority groups than in white women (8). Although minority women have historically undergone less mammography than white women (14), several recent surveys have found only small differences in mammography use between white and nonwhite women (24, 25). These observations raised doubt that tumors go undiagnosed until later stages in minority women because of infrequent breast cancer screening (26). However, the 2 most widely cited surveys of mammography use are based on self-report and only inquire about recent use, not adherence over time (24, 25).

See also:

Print
Editors' Notes 542
Editorial comment 614
Related article 563
Summary for Patients I-18

Web-Only
Conversion of figures and tables into slides

© 2006 American College of Physicians

Context

Breast cancer mortality rates have fallen but still differ by race and ethnicity. One explanation might be differences in mammography use.

Content

These investigators linked data from mammography registries to tumor registries and showed that African-American and Hispanic women have longer intervals between mammography and are more likely to have advanced-stage tumors at diagnosis and to die of breast cancer than white women. However, in women with similar screening histories, these rates were similar regardless of race or ethnicity.

Implications

Differences in mammography use may explain ethnic disparities in the incidence of advanced-stage breast cancer and in mortality rates.

—The Editors

We explored stage of disease at diagnosis, tumor characteristics (including size and grade), and lymph node involvement among women of different races and ethnicities whose patterns of mammography use were similar. We hypothesized that differences in tumor characteristics may result primarily from differences in mammography use and that women with similar patterns of mammography use may have similar tumor characteristics. We had sufficient sample sizes within each racial and ethnic group and obtained sufficiently detailed data regarding mammography use to permit stratification of the cohort by pattern of mammography use; this technique enabled us to compare tumor characteristics among women with similar screening histories.

METHODS

Data Source

We pooled data from facilities that participate in 7 mammography registries that form the National Cancer Institute–funded Breast Cancer Surveillance Consortium: San Francisco Mammography Registry, San Francisco, California; Group Health Cooperative, Seattle, Washington; Colorado Mammography Project, Denver, Colorado; Vermont Breast Cancer Surveillance System, Burlington, Vermont; New Hampshire Mammography Network, Lebanon, New Hampshire; Carolina Mammography Registry, Chapel Hill, North Carolina; and New Mexico Mammography Project, Albuquerque, New Mexico. The data consisted of information sent to the registries regarding all mammographic evaluations performed at these facilities, including radiology reports and breast health surveys. The surveys, which were completed by patients at each mammography examination, included questions regarding race, ethnicity, presence of breast symptoms, and previous mammography use. Breast cancer diagnoses and tumor characteristics were obtained through linkage with state tumor registries; regional Surveillance, Epidemiology, and End Results programs; and hospital-based pathology services. Previous research has shown that at least 94% of cancer cases are identified through these linkages (27). Each surveillance registry captures most mammography case reports within its respective geographic area, and mammograms in these registries include approximately 2% of mammographic examinations performed in the United States. Each registry obtains annual approval from its institutional review board to collect mammography-related information and to link with tumor registries.

Participants

This study included women without a history of breast cancer who were 40 years of age and older who had undergone mammography at least once for screening or diagnostic purposes between 1996 and 2002 (n = 1 010 515). We categorized the race and ethnicity of the participating women (the "mammography registry cohort") as non-Hispanic white (n = 789 997), non-Hispanic African American/black (n = 62 408), Hispanic (n = 90 642), Asian/Pacific Islander (n = 49 867), or Native American/Native Alaskan (n = 17 601). We excluded women who did not report their race or ethnicity (n = 133 235 [12%]) or reported mixed or other race (n = 6003 [<1%]). Breast cancer was diagnosed in a subset of the women in the mammography registry cohort (**Table 1**).

Characterization of Mammography Use

We included all mammographic evaluations in eligible women that were performed during the study period. We characterized each mammogram that was included in the study by the time interval between that mammogram and the one most recently preceding it. We determined these intervals by using examination dates that were recorded in the database (data were available for 85% of patients) and self-reported dates that the remaining women provided at the time of their examination. The mammography screening intervals were categorized into the following groups:

Table 1. **General Categorization of Study Participants**

Group	Group Size, *n*	Participants Who Received Inadequate Screening, *n* (%)
Mammography registry cohort*	10 10 515	147 810 (14.6)
Subset of women in registry cohort with breast cancer†	17 558	14 217 (19.0)

* Includes all women who had screening or diagnostic mammography performed at least once at a facility included in the Breast Cancer Surveillance Consortium Registry between 1996 and 2002.
† Includes women in the mammography registry cohort with a first diagnosis of breast cancer between 1996 and 2002.

Table 2. **Characteristics of the Study Participants***

Variable	Mammography Registry Cohort (n = 1 010 515)	Women with Breast Cancer (n = 17 558)
Ethnicity, *n* (%)		
White, non-Hispanic	789 997 (78.2)	14 693 (83.7)
African American, non-Hispanic	62 408 (6.2)	994 (5.7)
Hispanic	90 642 (9.0)	1077 (6.1)
Asian/Pacific Islander	49 867 (4.9)	669 (3.8)
Native American	17 601 (1.7)	125 (0.7)
Age, *n* (%)†		
40–49 y	378 311 (37.4)	3574 (20.4)
50–59 y	291 252 (28.8)	4908 (28.0)
60–69 y	176 445 (17.5)	4131 (23.5)
70–79 y	124 408 (12.3)	3551 (20.2)
80 y	40 099 (4.0)	1394 (7.9)
Type of tumor, *n* (%)		
Ductal carcinoma in situ	–	2902 (16.5)
Invasive	–	14 656 (83.5)
Invasive tumors by stage, *n* (%)‡		
Stage 1	–	7591 (56.7)
Stage 2	–	4864 (36.4)
Stage 3	–	677 (5.1)
Stage 4	–	245 (1.8)
Invasive tumors by grade, *n* (%)‡		
Grade 1	–	3004 (25.0)
Grade 2	–	4977 (41.5)
Grade 3	–	3657 (30.5)
Grade 4	–	359 (3.0)
Distribution of advanced-stage invasive tumors, *n* (%)‡§		
White, non-Hispanic	–	4708 (41.9)
African American, non-Hispanic	–	395 (55.1)
Hispanic	–	437 (52.0)
Asian/Pacific Islander	–	199 (41.9)
Native American	–	47 (46.1)
Distribution of high-grade invasive tumors, *n* (%)‡‖		
White, non-Hispanic	–	3212 (32.1)
African American, non-Hispanic	–	324 (50.0)
Hispanic	–	295 (36.6)
Asian/Pacific Islander	–	142 (32.5)
Native American	–	43 (43.4)

* Women with a first diagnosis of breast cancer were a subset of the mammography registry cohort.
† Randomly selected 1 observation per woman to estimate age.
‡ Among women with known stage or grade.
§ Advanced-stage tumors are those in stages 2 through 4.
‖ High-grade tumors are those of grades 3 and 4.

within 1 year (4 to 17 months); 2 years (18 to 29 months); 3 years (30 to 41 months); and 4 years or longer (>41 months). At the time of each mammogram, women completed a breast health survey and provided the date of their last mammogram.

We created 2 classifications for first mammograms. Mammography was classified as a first screening if the radiologist coded the examination as "screening" and the woman reported no breast symptoms. The mammogram was classified as diagnostic if the radiologist coded the examination as "diagnostic" or if the woman reported a breast mass or nipple discharge. Women whose first mammogram was diagnostic were assigned to the "never screened" group.

Of note, a woman could have had mammography more than once during the study period and therefore could contribute more than 1 observation to the analyses. A woman could have observations that were categorized into different mammography screening intervals. For example, a woman could have had her first mammographic evaluation in 1998 and had subsequent mammography in 1999 and 2001. Her first mammogram would have been categorized as a "first screening" or as "diagnostic," depending on the radiologist's indication for that examination and whether the patient reported symptoms. Her second mammogram would have been categorized in the "1 year" group, and her third mammography would have been categorized in the "2 year" group.

Table 3. Rates of Overall Breast Cancer and Large, Advanced-Stage, High-Grade, and Lymph Node–Positive Tumors per 1000 Mammograms by Racial and Ethnic Group*

Ethnicity	Overall Women with Breast Cancer		Women with Large Tumors†		Women with Advanced-Stage Tumors‡		Women with High-Grade Tumors§		Women with Lymph Node–Positive Tumors	
	Rate	Relative Rate (95% CI)	Rate	Relative Rate (95% CI)	Rate	Relative Rate (95% CI)	Rate	Relative Rate (95% CI)	Rate	Relative Rate (95% CI)
White	7.03	Referent	2.43	Referent	2.26	Referent	1.54	Referent	1.51	Referent
African American	7.30	1.04 (0.95–1.13)	2.88	**1.19 (1.06–1.33)**	2.93	**1.30 (1.16–1.45)**	2.17	**1.41 (1.24–1.61)**	1.86	**1.24 (1.09–1.40)**
Hispanic	5.33	**0.76 (0.70–0.83)**	2.21	0.91 (0.81–1.02)	2.08	0.92 (0.82–1.03)	1.42	0.92 (0.80–1.06)	1.36	0.91 (0.80–1.02)
Asian/Pacific Islander	5.33	**0.76 (0.68–0.85)**	1.77	**0.73 (0.62–0.86)**	1.62	**0.72 (0.61–0.84)**	1.37	0.89 (0.73–1.09)	0.94	**0.62 (0.52–0.75)**
Native American	3.76	**0.53 (0.43–0.67)**	1.52	**0.62 (0.46–0.84)**	1.37	**0.61 (0.45–0.83)**	1.23	0.80 (0.58–1.11)	0.97	**0.64 (0.47–0.89)**

* Adjusted for age and registry; values shown in boldface are statistically significant.
† Large tumors were defined as those greater than 15 mm.
‡ Advanced-stage tumors are those in stages 2 through 4.
§ High-grade tumors are those of grades 3 and 4.

Breast Cancer

To determine breast cancer status, we tracked each participant's mammogram for 365 days following the date it had been obtained or until the patient underwent her next mammographic examination (whichever came first). Consequently, each tumor was associated with a single mammogram—that obtained closest to the date of diagnosis. We characterized breast cancer as either invasive or ductal carcinoma in situ. Large tumors were defined as invasive tumors that were 15 mm or larger in diameter. We used the TNM (tumor, node, metastasis) system (which is based on the criteria of the American Joint Committee on Cancer) to classify stage at diagnosis as 0 (ductal carcinoma in situ), 1, 2, 3, or 4 (28); advanced-stage tumors were defined as invasive lesions of stage 2 or higher. High-grade tumors were defined as invasive lesions of grades 3 and 4. Lymph node status was defined as positive, negative, or unknown. Advanced disease was defined as the presence of a large, advanced-stage, high-grade tumor or lymph node–positive tumor at the time of diagnosis.

Statistical Analysis

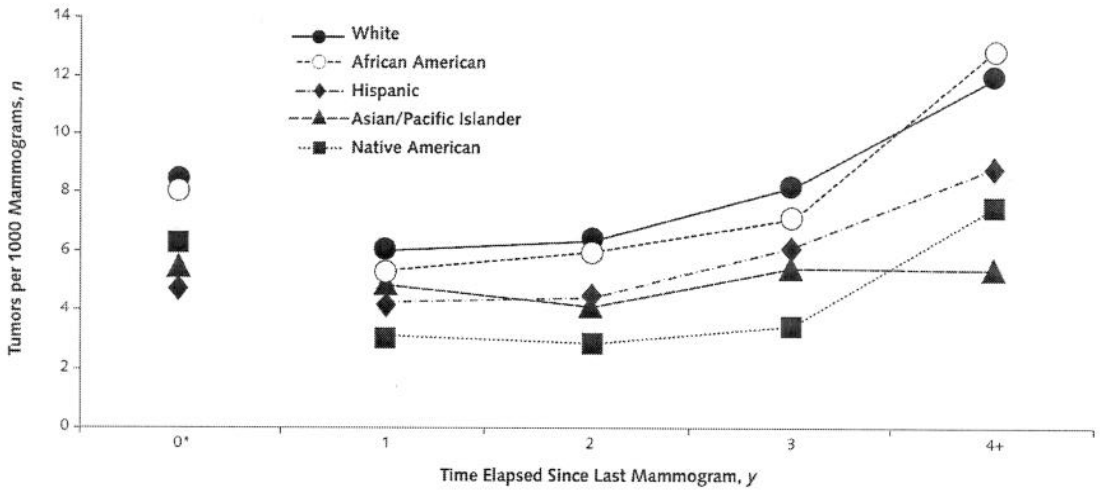
We calculated the frequency distributions of various risk factors for all women in the mammography registry cohort. Among the subset of women with breast cancer (n = 17 558), we calculated the proportion of tumors that were invasive and, among invasive tumors, the proportion that were advanced-stage or high-grade tumors; we then calculated the distribution by race and ethnicity. For all women in the cohort, we evaluated whether overall and advanced cancer rates per 1000 mammograms were similar across racial and ethnic groups after we adjusted for age and registry by using Poisson regression. We then calculated whether adjusted overall and advanced cancer rates per 1000 mammograms were similar across mammography screening interval groups. Because overall and advanced

Table 4. Rates of Overall Breast Cancer and Large, Advanced-Stage, High-Grade, and Lymph Node–Positive Tumors per 1000 Mammograms by Mammography Screening Interval Group*

Screening Interval Group	Overall Women with Breast Cancer		Women with Large Tumors†		Women with Advanced-Stage Tumors‡		Women with High-Grade Tumors§		Women with Lymph Node–Positive Tumors	
	Rate	Relative Rate (95% CI)	Rate	Relative Rate (95% CI)	Rate	Relative Rate (95% CI)	Rate	Relative Rate (95% CI)	Rate	Relative Rate (95% CI)
1 year	5.8	Referent	1.8	Referent	1.7	Referent	1.3	Referent	1.2	Referent
2 years	5.9	1.0 (0.95–1.1)	2.0	1.1 (0.96–1.2)	1.8	1.1 (0.93–1.2)	1.4	1.1 (0.95–1.2)	1.2	1.0 (0.91–1.2)
3 years	7.7	**1.3 (1.2–1.5)**	**2.8**	**1.5 (1.3–1.8)**	**2.6**	**1.5 (1.3–1.8)**	**1.7**	**1.3 (1.1–1.6)**	**1.6**	**1.4 (1.2–1.7)**
4 years or longer	11.4	**2.0 (1.8–2.2)**	**4.8**	**2.6 (2.3–3.0)**	**4.5**	**2.6 (2.3–3.0)**	**2.8**	**2.2 (1.9–2.5)**	**2.8**	**2.5 (2.1–2.8)**
First screening mammogram‖	7.6	**1.3 (1.1–1.5)**	2.3	1.3 (0.99–1.6)	2.1	1.2 (0.97–1.6)	1.2	1.0 (0.73–1.3)	1.3	1.1 (0.87–1.5)

* Tumors diagnosed within 365 days or before the next mammogram (whichever came first) per 1000 mammograms within each interval group. For example, among women in the 1-year mammography screening interval group, 5.82 tumors were diagnosed within the follow-up period per 1000 mammograms. Values shown in boldface are statistically significant; $P < 0.001$ across screening groups.
† Large tumors were defined as those greater than 15 mm.
‡ Advanced-stage tumors are those in stages 2 through 4.
§ High-grade tumors are those of grades 3 and 4.
‖ Rates of cancer in the first screening group reflect the rate of tumors that were diagnosed within 365 days after screening examinations among asymptomatic women who had no previous mammograms.

Figure 1. **Overall breast cancer rates per 1000 mammograms by racial and ethnic group and mammography screening interval group, adjusted to the age and registry distribution of the mammography registry cohort.**



	Relative Rate (95% CI) of Breast Cancer				
White	Referent	Referent	Referent	Referent	Referent
African American	1.00 (0.74–1.33)	**0.87 (0.77–0.98)**	0.96 (0.80–1.14)	0.86 (0.64–1.15)	1.07 (0.86–1.32)
Hispanic	**0.57 (0.40–0.81)**	**0.74 (0.66–0.82)**	**0.69 (0.58–0.82)**	**0.74 (0.57–0.96)**	**0.73 (0.60–0.90)**
Asian/Pacific Islander	**0.66 (0.49–0.90)**	**0.80 (0.70–0.92)**	**0.66 (0.54–0.81)**	**0.66 (0.47–0.94)**	**0.45 (0.30–0.67)**
Native American	0.76 (0.30–1.97)	**0.51 (0.37–0.69)**	**0.46 (0.28–0.75)**	**0.41 (0.20–0.86)**	0.63 (0.39–1.01)

Rates were calculated as the number of tumors occurring within 365 days after mammography or before the next mammogram, whichever came first. Values shown in boldface are statistically significant. *Includes asymptomatic women who were undergoing their first-ever mammogram for screening purposes only.

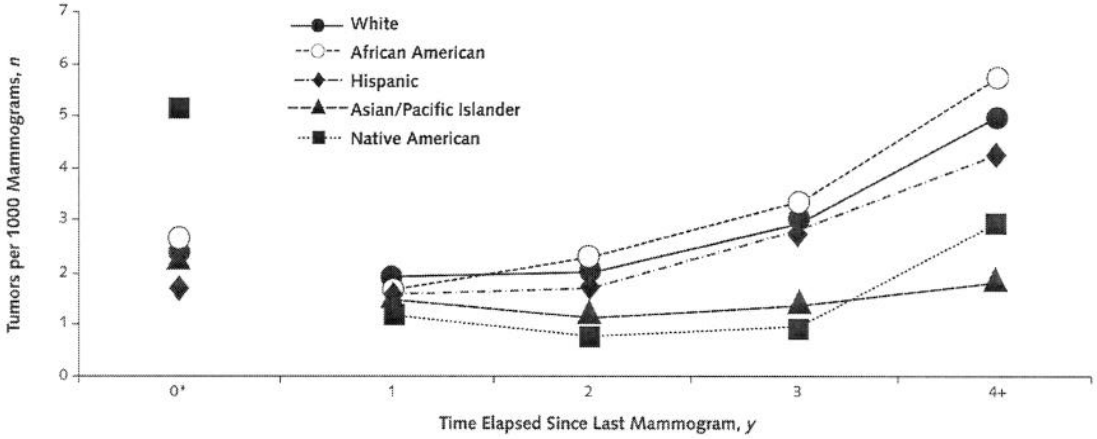
cancer rates varied across racial and ethnic groups ($P < 0.001$) and by previous mammography use ($P < 0.001$), and because mammography use potentially varied by race and ethnicity, we modeled cancer rates among similarly screened women in each ethnic group. We used Poisson regression to adjust for age and registry; an interaction term between race and ethnicity and previous mammography use was included in the Poisson model to allow for possible differences in the association between ethnicity and cancer rates by mammography group.

As previously described, a woman could have had mammography more than once during the study period and therefore could contribute more than 1 observation to the Poisson regression models. To account for possible correlation among such repeated mammography within women in the analysis, we estimated the scale parameter of the Poisson distribution from the data, allowing for underdispersion or overdispersion.

On the basis of the fitted Poisson model, we used marginal standardization (also called predictive margins) (29, 30) to estimate the adjusted total cancer rates and the prevalence of large, advanced-stage, high-grade tumors with lymph node involvement per 1000 mammograms by racial and ethnic group. We calculated the average of the rates that were estimated from the Poisson regression model for each combination of ethnicity, age, and registry; these averages were weighted by the proportion of women in the study with that particular combination of age and registry values. Because the same weights were used for each ethnic group, we adjusted these rates to a common population with distributions of age and registry that were equal to the distributions observed in our study sample. Similar methods were used to calculate adjusted rates by mammography screening interval and by ethnicity within mammography screening interval groups.

We explored differences in mammography use by racial and ethnic groups for the entire mammography registry cohort and among the subset of women with diagnosed breast cancer. For each woman in the mammography registry cohort, we randomly selected 1 mammogram to include in the analysis because we did not want to overcount women who were screened frequently. For women with breast cancer, we included the mammogram that was performed closest to the date of diagnosis. We classified mammograms from women with breast cancer and those without into mammography screening interval groups by using the previously described methods. In addition, we classified women as "frequently screened" if they had had previous

Figure 2. **Rates of large (>15 mm) tumors per 1000 mammograms by racial and ethnic group and mammography screening interval group, adjusted to the age and registry distribution of the mammography registry cohort.**



	Relative Rate (95% CI) of Large Tumors				
White	Referent	Referent	Referent	Referent	Referent
African American	1.11 (0.70–1.76)	0.87 (0.72–1.05)	1.10 (0.85–1.42)	1.13 (0.77–1.65)	1.16 (0.88–1.54)
Hispanic	0.77 (0.47–1.27)	**0.83 (0.71–0.98)**	0.84 (0.65–1.07)	0.98 (0.71–1.36)	0.87 (0.67–1.13)
Asian/Pacific Islander	0.86 (0.54–1.36)	**0.79 (0.63–0.98)**	**0.56 (0.40–0.79)**	**0.46 (0.25–0.85)**	**0.37 (0.21–0.68)**
Native American	2.19 (0.87–5.51)	**0.65 (0.42–0.99)**	**0.41 (0.19–0.89)**	0.33 (0.10–1.08)	0.61 (0.32–1.17)

Rates were calculated as the number of tumors occurring within 365 days after mammography or before the next mammogram, whichever came first. Values shown in boldface are statistically significant. *Includes asymptomatic women who were undergoing their first-ever mammogram for screening purposes only.

mammography within 30 months of the mammogram that was included in the analysis. We classified women as "inadequately screened" if their first mammogram was diagnostic, if they were 55 years of age or older at the time of their first screening mammogram, or if they had not had a mammogram within 42 months before the mammogram that was included in the analysis.

We used log-binomial regression (adjusting for age and registry) to examine racial and ethnic differences in mammography screening interval groups in the entire mammography registry cohort and in the subset of women with breast cancer. When we examined the risk associated with absence of screening among the women with breast cancer, the log-binomial model would not converge; therefore, we used the modified Poisson regression approach of Zou (31) for estimation. All analyses were performed with SAS statistical software, version 8.2 (SAS Institute, Inc., Cary, North Carolina). Details of all analytic methods and procedures are available on request from the authors.

Role of the Funding Sources

National Cancer Institute staff with relevant expertise participated in several aspects of this project, including design, collection, management, analysis and interpretation of the data, and manuscript preparation. The National Cancer Institute provided support to this project through the Breast Cancer Surveillance Consortium cooperative agreements that underwent peer review. The final version of the manuscript was approved by all study authors and principal investigators of the Breast Cancer Surveillance Consortium sites. The remaining funding organizations were not involved in the design, analysis, or interpretation of the data.

RESULTS

Between 1996 and 2002, 1 010 515 women who were 40 years of age and older had 2 588 479 eligible mammograms within the Breast Cancer Surveillance Consortium. Of these, 17 558 women received a first-time diagnosis of breast cancer. Demographic characteristics are shown in **Table 2**. Overall, 83.5% of the tumors detected were invasive; of these, 43% were stage 2 or higher and 33% were grade 3 or 4. Advanced-stage tumors were more likely to be diagnosed in African-American and Hispanic women than in white women ($P < 0.001$ for both), and high-grade tumors were more likely to be diagnosed in African-American ($P < 0.001$), Hispanic ($P = 0.008$), and Native American ($P = 0.017$) women than in white women.

Mammography Registry Cohort

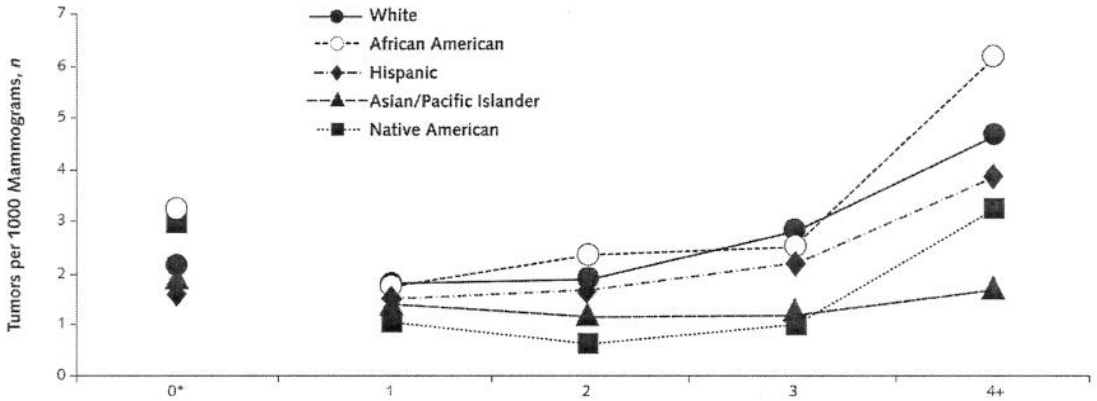
Overall breast cancer rates differed significantly by race and ethnicity (**Table 3**). Whereas 7 tumors were diagnosed per 1000 mammograms in white women, cancer rates were

significantly lower for Hispanic (5.3 tumors per 1000 mammograms; relative rate, 0.76 [CI, 0.70 to 0.83]), Asian (5.3 tumors per 1000 mammograms; relative rate, 0.76 [CI, 0.68 to 0.85]), and Native American women (3.8 tumors per 1000 mammograms; relative rate, 0.53 [CI, 0.43 to 0.67]). Overall cancer rates in African-American women were not significantly different from those in white women (relative rate, 1.04 [CI, 0.95 to 1.13]); however, African-American women had significantly higher rates of large tumors (relative rate, 1.19 [CI, 1.06 to 1.33]), advanced-stage tumors (relative rate, 1.30 [CI, 1.16 to 1.45]), high-grade tumors (relative rate, 1.41 [CI, 1.24 to 1.61]), and lymph node involvement (relative rate, 1.24 [CI, 1.09 to 1.40]). Asian and Native American women had significantly lower rates of large, advanced-stage tumors and lymph node involvement than white women. Of interest, Hispanic women had lower overall breast cancer rates than white women but advanced cancer rates were similar in both groups.

We explored whether overall breast cancer rates and advanced cancer rates varied by use of mammography (**Table 4**). As we expected, the total cancer rates and rates of large, advanced-stage, lymph node–positive tumors increased in relation to the length of time between mammographic examinations ($P < 0.001$ for trend for each outcome). For the outcome of large tumors, the rate increased from 1.8 tumors per 1000 mammograms for women with a 1-year screening interval to 2.0 tumors for women with a 2-year interval (relative rate, 1.1 [CI, 0.96 to 1.21]). The rate further increased to 2.8 tumors per 1000 mammograms for a 3-year interval (relative rate, 1.5 [CI, 1.3 to 1.8]) and to 4.8 tumors per 1000 mammograms for an interval of 4 or more years (relative rate, 2.6 [CI, 2.3 to 3.0]). The incidence of cancer and advanced tumors (except for high-grade tumors) within 1 year of first screening mammography fell between the values for the 2-year and 3-year screening interval groups.

For all racial and ethnic groups, the total cancer rates and rates of large, advanced-stage, high-grade, and lymph node–positive tumors increased with the length of time elapsed since the patient's previous mammogram (**Figures 1** through **5**). In general, rates of advanced breast cancer in white and African-American women did not differ when we compared women who were screened at the same intervals (**Figures 2** through **4**). African-American women had rates of large, advanced-stage, and lymph node–positive tumors that were similar to those of white women with the same screening history (**Figures 2** through **4**). There were several exceptions, however. African-American women had significantly higher rates of high-grade tumors than white women at all screening frequencies (**Figure 5**). Among women undergoing their first screening mammography,

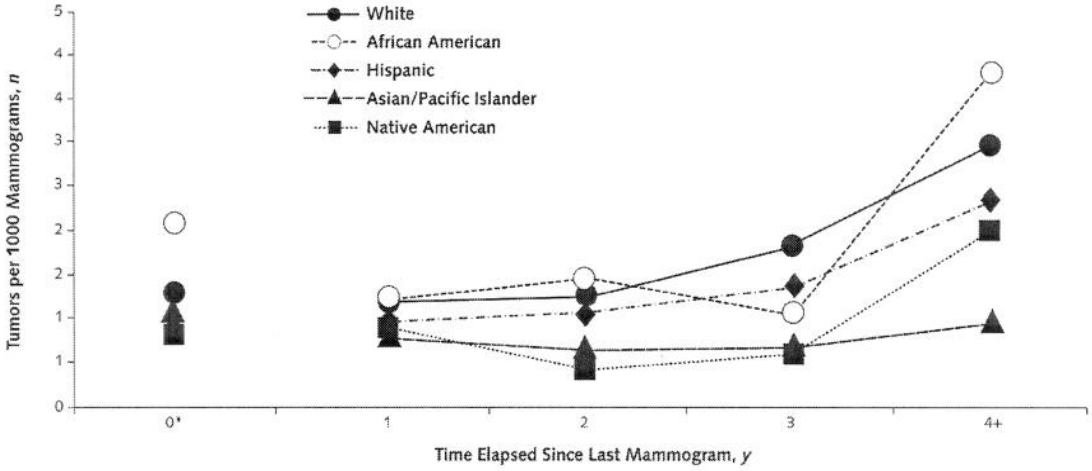
Figure 3. **Rates of advanced-stage tumors per 1000 mammograms by racial and ethnic group and mammography screening interval group, adjusted to the age and registry distribution of the mammography registry cohort.**



	Relative Rate (95% CI) of Advanced-Stage 1				
White	Referent	Referent	Referent	Referent	Referent
African American	**1.57 (1.03–2.39)**	0.99 (0.83–1.19)	1.24 (0.96–1.60)	0.88 (0.58–1.35)	1.30 (0.99–1.71)
Hispanic	0.80 (0.49–1.33)	0.88 (0.76–1.03)	0.87 (0.68–1.11)	0.77 (0.54–1.11)	0.81 (0.62–1.06)
Asian/Pacific Islander	0.81 (0.50–1.32)	**0.76 (0.61–0.95)**	**0.59 (0.42–0.83)**	**0.42 (0.22–0.79)**	**0.35 (0.19–0.65)**
Native American	1.47 (0.46–4.67)	**0.64 (0.42–0.99)**	**0.33 (0.13–0.79)**	0.35 (0.11–1.09)	0.71 (0.39–1.30)

Rates were calculated as the number of tumors occurring within 365 days after mammogram or before the next mammogram, whichever came first. Values shown in boldface are statistically significant. *Includes asymptomatic women who were undergoing their first-ever mammogram for screening purposes only.

Figure 4. **Rates of lymph node–positive tumors per 1000 mammograms by racial and ethnic group and mammography screening interval group, adjusted to the age and registry distribution of the mammography registry cohort.**



	Relative Rate (95% CI) of Lymph Node–Positive Tumors				
White	Referent	Referent	Referent	Referent	Referent
African American	**1.63 (1.02–2.59)**	1.02 (0.84–1.24)	1.17 (0.88–1.55)	0.60 (0.34–1.05)	1.28 (0.95–1.73)
Hispanic	0.85 (0.50–1.45)	**0.83 (0.70–0.98)**	0.84 (0.64–1.09)	0.77 (0.52–1.13)	0.78 (0.59–1.05)
Asian/Pacific Islander	0.77 (0.45–1.31)	**0.64 (0.50–0.82)**	0.50 (0.34–0.73)	**0.36 (0.17–0.74)**	**0.32 (0.16–0.62)**
Native American	1.75 (0.13–4.38)	0.75 (0.49–1.15)	**0.38 (0.16–0.91)**	0.35 (0.10–1.21)	0.68 (0.35–1.33)

Rates were calculated as the number of tumors occurring within 365 days after mammography or before the next mammogram, whichever came first. Values shown in boldface are statistically significant. *Includes asymptomatic women who were undergoing their first-ever mammogram for screening purposes only.

African-American women also had higher rates of advanced-stage and lymph node–positive tumors. Among women screened at an interval of 4 or more years, African-American women had higher rates of advanced-stage and lymph node–positive tumors; this finding bordered on statistical significance.

Asian and Native American women had significantly lower overall breast cancer rates and lower rates of large, advanced-stage, and lymph node–positive tumors than white women when we stratified the cohort by mammography screening intervals. Consequently, the relative rate of advanced disease did not change substantially in these groups after we accounted for mammography frequency. Hispanic women also had significantly lower overall cancer rates than white women when we accounted for screening history. In addition, Hispanic women had lower advanced cancer rates than white women when we subdivided the groups by mammography screening interval, but these differences did not reach statistical significance.

Mammography Use

Because of the racial and ethnic differences in the rates of advanced cancer, we explored variations in mammography use as a possible cause (**Table 5**). We analyzed these trends in both the mammography registry cohort and the subset of women with breast cancer.

Mammography Registry Cohort

White women were more likely to be frequently screened at an interval of 1 to 2 years. Compared with 72% of white women, only 63% to 68% of African-American, Hispanic, Asian, and Native American women were frequently screened (age- and site-adjusted relative risk, 0.86 to 0.93). African-American, Hispanic, and Asian women were more likely to have never undergone screening; therefore, they were more likely to have their first mammogram because of a physical examination finding or breast symptom (age- and site-adjusted relative risk, 1.3 to 1.7). Nonwhite women were also more likely to receive inadequate mammographic screening (age- and site-adjusted relative risk, 1.2 to 1.4).

Subset of Women with Breast Cancer

When we evaluated mammography use among women with breast cancer, the magnitude of the racial and ethnic differences tended to be greater than those seen in our analysis of the entire mammography registry cohort. For example, 24% to 34% of African-American, Hispanic, and Native American women were inadequately screened before their breast cancer diagnosis, compared with 18% of white women and 19% of Asian women. After we adjusted for age and registry, all racial and ethnic groups were sig-

nificantly more likely than white women to receive inadequate screening. The adjusted relative risk was 1.6 (CI, 1.5 to 1.8) for African-American women, 1.5 (CI, 1.3 to 1.7) for Hispanic women, 1.4 (CI, 1.2 to 1.7) for Asian women, and 1.5 (CI, 1.2 to 2.0) for Native American women.

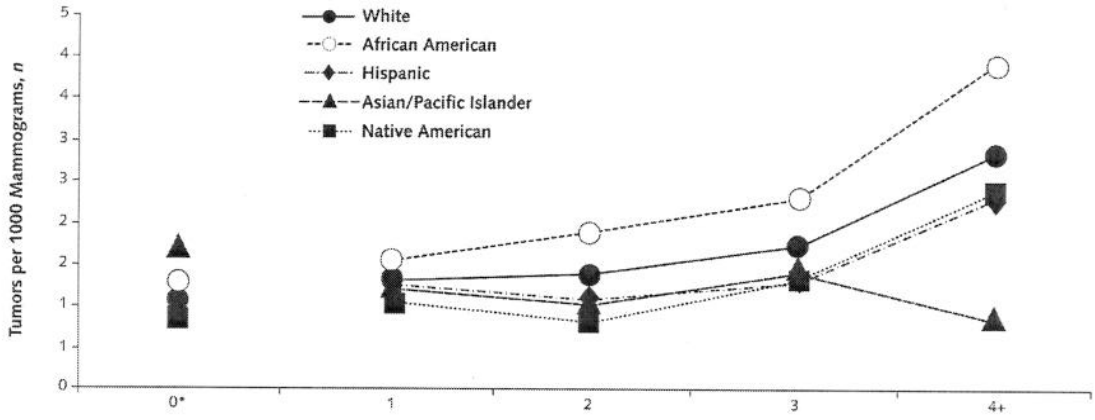
Discussion

There are well-described differences in breast cancer outcomes by race and ethnicity (1). To understand the reasons for the racial disparities in breast cancer outcomes, we studied mammography use in detail among a large cohort of women who had at least 1 mammography examination within the Breast Cancer Surveillance Consortium. Consistent with previous studies, we found that large, advanced-stage tumors and lymph node–involved tumors were more likely to be diagnosed in African-American women than in white women; however, these differences were attenuated or eliminated when comparisons were stratified by extent of previous mammography use. This suggests that inadequate use of screening mammography may be an important reason for higher rates of advanced disease in African-American women. Analysis of mammography use supports this conclusion because nonwhite women are more likely than white women to receive inadequate mammographic screening. On the basis of our analysis of mammography use among women with breast cancer, 34% of African-American women received inadequate mammographic screening before their diagnosis of breast cancer. This suggests that underuse of mammography may be an ongoing issue, particularly among African-American women.

Asian and Native American women had significantly lower rates of large, advanced-stage, and lymph node–positive tumors than white women, and these differences persisted even after accounting for previous mammography screening. This finding is consistent with the overall lower rate of breast cancer in these women (32). Of interest, our analysis of mammography use also suggested that many Asian and Native American women have received inadequate mammographic screening. Increased use of mammography among Asian and Native American women may further reduce their burden of advanced disease.

Because of the nature of the database we used, we believe that our analysis of screening frequency among women with breast cancer may better reflect mammography use in the general population than do our data from the entire cohort. The Breast Cancer Surveillance Consortium registries capture virtually all women in their respective geographic areas who underwent mammography and

Figure 5. **Rates of high-grade tumors per 1000 mammograms by racial and ethnic group and mammography screening interval group, adjusted to the age and registry distribution of the mammography registry cohort.**



Relative Rate (95% CI) of High-Grade Tumors

White	Referent	Referent	Referent	Referent	Referent
African American	1.08 (0.62–1.89)	**1.22 (1.02–1.45)**	**1.39 (1.09–1.79)**	1.32 (0.88–1.96)	**1.37 (1.02–1.84)**
Hispanic	1.02 (0.58–1.78)	0.90 (0.75–1.07)	0.87 (0.66–1.13)	0.74 (0.48–1.14)	0.81 (0.59–1.10)
Asian/Pacific Islander	1.46 (0.88–2.43)	0.95 (0.74–1.22)	0.74 (0.52–1.06)	0.81 (0.44–1.48)	**0.30 (0.13–0.70)**
Native American	0.80 (0.12–5.14)	0.83 (0.55–1.26)	0.60 (0.30–1.21)	0.73 (0.29–1.86)	0.82 (0.42–1.58)

Rates were calculated as the number of tumors occurring within 365 days after mammography or before the next mammogram, whichever came first. Values shown in boldface are statistically significant. *Includes asymptomatic women who were undergoing their first-ever mammogram for screening purposes only.

Table 5. **History of Mammography Use by Participants***

Ethnic Group	Participants, *n*	Mammographic Screening Interval Group†					
		Time since Previous Mammography				First Screening Mammography, *n* (%)	Never Screened, *n* (%)‖
		1 Year, *n* (%)	2 Years, *n* (%)	3 Years, *n* (%)	>4 Years, *n* (%)		
White							
	Cohort: 789 997	383 129 (48)	187 859 (24)	68 282 (9)	91 115 (12)	52 329 (7)	7283 (0.9)
	Cases: 14 693	7953 (54)	2981 (20)	994 (7)	1481 (10)	436 (3)	848 (5.8)
African American							
	Cohort: 62 408	25 647 (41)	13 904 (22)	5845 (9)	7750 (12)	8153 (13)	1109 (1.8)
	Cases: 994	394 (40)	175 (18)	66 (7)	129 (13)	73 (7)	157 (15.8)
Hispanic							
	Cohort: 90 642	38 033 (42)	20 232 (22)	8732 (10)	12 188 (13)	10 287 (11)	1170 (1.3)
	Cases: 1077	522 (48)	188 (17)	90 (8)	140 (13)	49 (5)	88 (8.2)
Asian							
	Cohort: 49 867	18 658 (37)	12 887 (26)	4607 (9)	4293 (9)	8870 (18)	552 (1.1)
	Cases: 669	330 (49)	136 (20)	46 (7)	36 (5)	66 (10)	55 (8.2)
Native American							
	Cohort: 17 601	7678 (44)	4309 (24)	1956 (11)	2526 (14)	1002 (6)	130 (0.7)
	Cases: 125	56 (45)	23 (18)	10 (8)	24 (19)	6 (5)	6 (4.8)

* One observation per woman was randomly selected from all mammography performed between 1996 and 2002.
† We characterized each mammogram on the basis of the time interval between that test and that woman's most recent preceding test. The mammography intervals were categorized into the following mammography utilization groups: 1 year (4–17 months), 2 years (18–29 months), 3 years (30–41 months), and 4 years or longer (> 42 months). Asymptomatic women who were undergoing their first mammogram were categorized as "first screening" if the mammogram was obtained for screening purposes.
‡ Includes women who had no screening mammography (never screened), those who had their first mammogram at 55 years of age or older, and those who had not been screened in the past 42 months.
§ Calculated by comparing white women with each group of nonwhite women. Comparisons were made between women in the entire mammography registry cohort and between the subset of women with breast cancer.
‖ Women were categorized as "never screened" if the included first mammogram was performed to evaluate a breast symptom.
¶These data are a subset of the First Screening Mammography column.

virtually all women found to have breast cancer (assuming that women with breast cancer received some type of diagnostic work-up). However, the consortium registries do not capture women who have never undergone mammography. Consequently, the mammography registry cohort probably underrepresents the percentage of all women who are truly inadequately screened. In comparison, the Breast Cancer Surveillance Consortium includes virtually all women with breast cancer (both those screened with mammography and those who only underwent diagnostic mammography); therefore, we believe the percentage of these women who were considered "inadequately screened" will more closely approximate the percentage of all women who are truly inadequately screened.

Several widely cited studies have found that self-reported annual or biennial mammography use within the years of our study was quite high and that racial and ethnic differences in mammography use have generally disappeared (24, 25). Previous research has shown that self-reported data are prone to overstatement, particularly among minority women (33–38). Our results are primarily based on data from mammography facilities instead of from patient self-reports, which may account for the differences between our results and those reported by others.

We found indirect evidence of biological differences across racial and ethnic groups. Differences in tumor grade across racial and ethnic groups did not disappear after we accounted for mammography use (11, 13). African-American women had more high-grade tumors than white women regardless of screening frequency; this finding is consistent with clinical expectation. Mammography use might not influence tumor grade (which reflects biological differences), whereas delayed diagnosis (because of underuse of mammography or delay in subsequent diagnostic evaluation) affects size, stage, and lymph node status. The differences in tumor grade are clinically important because grade is an important predictor of breast cancer survival (23).

Differences in biological characteristics of tumors, mammography use, and access to and utilization of cancer treatments all probably contribute to the observed differences in breast cancer mortality rates across racial and ethnic groups. Differences in access to care and use of breast cancer treatment (7, 39), as well as response to similar treatment (40), have been identified across racial and ethnic groups; therefore, improved adherence to recommended mammography screening intervals may not eliminate all of the differences in breast cancer mortality rates between nonwhite and white women. However, the strongest predictor of breast cancer survival is stage of disease at diagnosis (22). In addition, patients with breast cancer that was diagnosed by screening mammography have better long-term survival than those with tumors of the same stage that were found by other means (41, 42). Improved

Table 5—Continued

Women with First Mammography at Age >55 y, *n* (%)¶	Women Who Were Inadequately Screened, *n* (%)‡	Relative Risk (95% CI)§					
		Women Who Were Frequently Screened	*P* Value	Women Who Were Never Screened	*P* Value	Women Who Were Inadequately Screened	*P* Value
11 121 (1)	109 519 (14)	Referent	–	Referent	–	Referent	–
254 (2)	2 583 (18)	Referent	–	Referent	–	Referent	–
2462 (4)	11 321 (18)	0.90 (0.90–0.91)	<0.001	1.7 (1.6–1.8)	<0.001	1.2 (1.2–1.2)	<0.001
55 (6)	341 (34)	0.81 (0.77–0.85)	<0.001	2.2 (1.9–2.6)	<0.001	1.6 (1.5–1.9)	<0.001
2658 (3)	16 016 (18)	0.90 (0.89–0.90)	<0.001	1.6 (1.5–1.7)	<0.001	1.3 (1.2–1.3)	<0.001
27 (3)	255 (24)	0.88 (0.84–0.92)	<0.001	1.6 (1.2–2.0)	<0.001	1.5 (1.3–1.7)	<0.001
3208 (6)	8053 (16)	0.86 (0.85–0.87)	<0.001	1.3 (1.2–1.5)	<0.001	1.4 (1.3–1.4)	<0.001
37 (6)	128 (19)	0.88 (0.84–0.93)	<0.001	1.8 (1.4–2.5)	<0.001	1.4 (1.2–1.7)	<0.001
245 (1)	2901 (16)	0.93 (0.92–0.94)	<0.001	0.98 (0.82–1.16)	0.79	1.2 (1.1–1.2)	<0.001
4 (3)	34 (27)	0.86 (0.75–0.98)	0.022	0.84 (0.39–1.84)	0.66	1.5 (1.2–2.0)	0.003

adherence to recommended mammography screening intervals among African-American and Hispanic women would probably reduce the prevalence of advanced-stage disease and therefore translate into a reduction in breast cancer mortality rates in these populations.

Over the past 2 decades, federal and state government health agencies and private health care organizations (such as the American Cancer Society) have striven to increase mammography use. As a result, overall rates of mammography use have increased (24, 25). In addition, outreach programs have made efforts to increase mammography use among underserved women. For example, the National Breast and Cervical Cancer Early Detection Program (which is run by the Centers for Disease Control and Prevention) offers mammography to underserved women who are 50 years of age and older. However, some minority groups still do not seem to undergo regular screening mammography. This is probably attributable to a combination of factors, including persistent financial barriers (22), difficulty in accessing facilities that perform mammography, and multiple personal and cultural reasons for choosing not to undergo screening (43–47). Additional work needs to be done to try to understand how to improve outreach and provide screening to these women.

Rates of large, advanced-stage, high-grade, and lymph node–positive tumors are important outcomes because they are strongly associated with breast cancer mortality. In comparison, proportions of advanced-stage tumors are more difficult to interpret because they may not account for differences in mammography use and potential overdiagnosis of nonpalpable lesions by mammography (15). In this paper, we assessed cancer characteristics among similarly screened women and evaluated rates of advanced disease so that overdiagnosis would not be a confounding factor. After we accounted for variations in mammography screening, the proportion (data not shown) and rates of advanced cancer were similar across all racial and ethnic groups.

The main strengths of our study are its large size; its geographic, racial, and ethnic diversity; and the detail and objectivity of our data concerning mammography use, breast symptoms, and cancer outcomes. Our study also has 3 limitations. First, the participants all underwent mammography within the Breast Cancer Surveillance Consortium, and these women may not be representative of all women. However, previous analyses have found that women in the geographic regions included in the Breast Cancer Surveillance Consortium are representative of women across the United States (48). Second, we evaluated the interval between each mammogram and the preceding mammogram but did not evaluate each woman's pattern of mammography use over time. Third, information regarding mammography use was based on most women's actual medical records; however, we relied on self-reports for the

remaining women. Although this strategy may have introduced some error in a woman's mammography characterization, the discrepancy probably was not large or biased in the direction to support our results. Overestimation of the recency of mammographic examinations probably would have attenuated our results. We did control for age and registry but did not examine mammography accuracy or how it might vary by race or ethnicity; this omission might contribute to reported differences in cancer outcomes. However, accuracy of screening mammography among white women and African-American women (49) and Asian women (32) has been reported to be similar. We did not adjust for insurance status or socioeconomic status, both of which are associated with mammography use.

In conclusion, frequency of mammography screening correlates inversely with severity of breast tumor characteristics at diagnosis. Furthermore, it seems that many women do not undergo routine mammography. Increased adherence to recommended mammography screening intervals, particularly among never-screened or infrequently screened women, may enable discovery of tumors before they have progressed to an advanced stage and may result in decreased mortality rates.

From the University of California and San Francisco Veterans Affairs Medical Center, San Francisco, California; Group Health Cooperative, Seattle, Washington; RAND Corporation, Arlington, Virginia; National Cancer Institute, Bethesda, Maryland; University of Colorado Health Sciences Center, Denver, Colorado; University of Kentucky Prevention Research Center, Lexington, Kentucky; Cancer Research and Biostatistics, Seattle, Washington; and University of North Carolina, Pembroke, North Carolina.

Grant Support: By the Mount Zion Dean's Account, the California Breast Cancer Research Program, The Department of Defense Congressionally Directed Medical Research Programs, The National Cancer Institute (K07 CA86032), and National Cancer Institute–funded Breast Cancer Surveillance Consortium cooperative agreements (U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040.)

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Rebecca Smith-Bindman, MD, Department of Radiology, University of California, 1600 Divisadero Street, San Francisco, CA 94115; e-mail, Rebecca.Smith-Bindman@Radiology.UCSF.edu.

Current author addresses and author contributions are available at www.annals.org.

References

1. **Ries L, Eisner M, Kosary C, Hankey BF, Miller BA, Clegg L, et al.** SEER Cancer Statistics Review, 1973-1999. Accessed at http://seer.cancer.gov/csr/1973_1999/ on 20 February 2006.

2. **Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al.** Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005;353:1784-92. [PMID: 16251534]

3. **Anderson LM, May DS.** Has the use of cervical, breast, and colorectal cancer screening increased in the United States? Am J Public Health. 1995;85:840-2. [PMID: 7762721]

4. **Chu KC, Lamar CA, Freeman HP.** Racial disparities in breast carcinoma survival rates: seperating factors that affect diagnosis from factors that affect treatment. Cancer. 2003;97:2853-60. [PMID: 12767100]

5. **Chu KC, Tarone RE, Kessler LG, Ries LA, Hankey BF, Miller BA, et al.** Recent trends in U.S. breast cancer incidence, survival, and mortality rates. J Natl Cancer Inst. 1996;88:1571-9. [PMID: 8901855]

6. **Natarajan N, Nemoto T, Mettlin C, Murphy GP.** Race-related differences in breast cancer patients. Results of the 1982 national survey of breast cancer by the American College of Surgeons. Cancer. 1985;56:1704-9. [PMID: 4027901]

7. **Eley JW, Hill HA, Chen VW, Austin DF, Wesley MN, Muss HB, et al.** Racial differences in survival from breast cancer. Results of the National Cancer Institute Black/White Cancer Survival Study. JAMA. 1994;272:947-54. [PMID: 8084062]

8. **Li CI, Malone KE, Daling JR.** Differences in breast cancer stage, treatment, and survival by race and ethnicity. Arch Intern Med. 2003;163:49-56. [PMID: 12523916]

9. **Joslyn SA, West MM.** Racial differences in breast carcinoma survival. Cancer. 2000;88:114-23. [PMID: 10618613]

10. **Breen N, Kessler LG, Brown ML.** Breast cancer control among the underserved—an overview. Breast Cancer Res Treat. 1996;40:105-15. [PMID: 8888156]

11. **Chen VW, Correa P, Kurman RJ, Wu XC, Eley JW, Austin D, et al.** Histological characteristics of breast carcinoma in blacks and whites. Cancer Epidemiol Biomarkers Prev. 1994;3:127-35. [PMID: 7519506]

12. **Elledge RM, Clark GM, Chamness GC, Osborne CK.** Tumor biologic factors and breast cancer prognosis among white, Hispanic, and black women in the United States. J Natl Cancer Inst. 1994;86:705-12. [PMID: 7908990]

13. **Trock BJ.** Breast cancer in African American women: epidemiology and tumor biology. Breast Cancer Res Treat. 1996;40:11-24. [PMID: 8888149]

14. **Burns RB, McCarthy EP, Freund KM, Marwill SL, Shwartz M, Ash A, et al.** Black women receive less mammography even with similar use of primary care. Ann Intern Med. 1996;125:173-82. [PMID: 8686974]

15. **McCarthy EP, Burns RB, Coughlin SS, Freund KM, Rice J, Marwill SL, et al.** Mammography use helps to explain differences in breast cancer stage at diagnosis between older black and white women. Ann Intern Med. 1998;128:729-36. [PMID: 9556466]

16. **Chang SW, Kerlikowske K, Napoles-Springer A, Posner SF, Sickles EA, Perez-Stable EJ.** Racial differences in timeliness of follow-up after abnormal screening mammography. Cancer. 1996;78:1395-402. [PMID: 8839544]

17. **Facione NC.** Delay versus help seeking for breast cancer symptoms: a critical review of the literature on patient and provider delay. Soc Sci Med. 1993;36:1521-34. [PMID: 8327915]

18. **Gordon NH, Crowe JP, Brumberg DJ, Berger NA.** Socioeconomic factors and race in breast cancer recurrence and survival. Am J Epidemiol. 1992;135:609-18. [PMID: 1580237]

19. **Singh GK, Miller BA, Hankey BF.** Changing area socioeconomic patterns in U.S. cancer mortality, 1950-1998: Part II—lung and colorectal cancers. J Natl Cancer Inst. 2002;94:916-25. [PMID: 12072545]

20. **Wells BL, Horm JW.** Stage at diagnosis in breast cancer: race and socioeconomic factors. Am J Public Health. 1992;82:1383-5. [PMID: 1415866]

21. **Yood MU, Johnson CC, Blount A, Abrams J, Wolman E, McCarthy BD, et al.** Race and differences in breast cancer survival in a managed care population. J Natl Cancer Inst. 1999;91:1487-91. [PMID: 10469750]

22. **Ward E, Jemal A, Cokkinides V, Singh GK, Cardinez C, Ghafoor A, et al.** Cancer disparities by race/ethnicity and socioeconomic status. CA Cancer J Clin. 2004;54:78-93. [PMID: 15061598]

23. **Rosenberg J, Chia YL, Plevritis S.** The effect of age, race, tumor size, tumor grade, and disease stage on invasive ductal breast cancer survival in the U.S. SEER database. Breast Cancer Res Treat. 2005;89:47-54. [PMID: 15666196]

24. **Blackman DK, Bennett EM, Miller DS.** Trends in self-reported use of mammograms (1989–1997) and Papanicolaou tests (1991–1997)—Behavioral Risk Factor Surveillance System. MMWR CDC Surveill Summ. 1999;48:1-22. [PMID: 10526871]

25. **Breen N, Wagener DK, Brown ML, Davis WW, Ballard-Barbash R.** Progress in cancer screening over a decade: results of cancer screening from the 1987, 1992, and 1998 National Health Interview Surveys. J Natl Cancer Inst. 2001;93:1704-13. [PMID: 11717331]

26. **Jones BA, Kasl SV, Curnen MG, Owens PH, Dubrow R.** Can mammog-

raphy screening explain the race difference in stage at diagnosis of breast cancer? Cancer. 1995;75:2103-13. [PMID: 7697601]
27. **Ernster VL, Ballard-Barbash R, Barlow WE, Zheng Y, Weaver DL, Cutter G, et al.** Detection of ductal carcinoma in situ in women undergoing screening mammography. J Natl Cancer Inst. 2002;94:1546-54. [PMID: 12381707]
28. **Fleming I, Cooper J, Henson D.** AJCC Cancer Staging Manual. 5th ed. New York: Lippincott–Raven; 1997.
29. **Lane PW, Nelder JA.** Analysis of covariance and standardization as instances of prediction. Biometrics. 1982;38:613-21. [PMID: 7171691]
30. **Graubard BI, Korn EL.** Predictive margins with survey data. Biometrics. 1999;55:652-9. [PMID: 11318229]
31. **Zou G.** A modified Poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004;159:702-6. [PMID: 15033648]
32. **Kerlikowske K, Creasman J, Leung JW, Smith-Bindman R, Ernster VL.** Differences in screening mammography outcomes among White, Chinese, and Filipino women. Arch Intern Med. 2005;165:1862-8. [PMID: 16157830]
33. **Degnan D, Harris R, Ranney J, Quade D, Earp JA, Gonzalez J.** Measuring the use of mammography: two methods compared. Am J Public Health. 1992; 82:1386-8. [PMID: 1415867]
34. **May DS, Trontell AE.** Mammography use by elderly women: a methodological comparison of two national data sources. Ann Epidemiol. 1998;8:439-44. [PMID: 9738690]
35. **Gordon NP, Hiatt RA, Lampert DI.** Concordance of self-reported data and medical record audit for six cancer screening procedures. J Natl Cancer Inst. 1993;85:566-70. [PMID: 8455203]
36. **Zapka JG, Bigelow C, Hurley T, Ford LD, Egelhofer J, Cloud WM, et al.** Mammography use among sociodemographically diverse women: the accuracy of self-report. Am J Public Health. 1996;86:1016-21. [PMID: 8669504]
37. **Champion VL, Menon U, McQuillen DH, Scott C.** Validity of self-reported mammography in low-income African-American women. Am J Prev Med. 1998;14:111-7. [PMID: 9631162]
38. **Peek ME, Han JH.** Disparities in screening mammography. Current status, interventions and implications. J Gen Intern Med. 2004;19:184-94. [PMID: 15009798]
39. **Haggstrom DA, Quale C, Smith-Bindman R.** Differences in the quality of breast cancer care among vulnerable populations. Cancer. 2005;104:2347-58. [PMID: 16211547]
40. **Bach PB, Schrag D, Brawley OW, Galaznik A, Yakren S, Begg CB.** Survival of blacks and whites after a cancer diagnosis. JAMA. 2002;287:2106-13. [PMID: 11966385]
41. **Shen Y, Yang Y, Inoue LY, Munsell MF, Miller AB, Berry DA.** Role of detection method in predicting breast cancer survival: analysis of randomized screening trials. J Natl Cancer Inst. 2005;97:1195-203. [PMID: 16106024]
42. **Joensuu H, Lehtimaki T, Holli K, Elomaa L, Turpeenniemi-Hujanen T, Kataja V, et al.** Risk for distant recurrence of breast cancer detected by mammography screening or other methods. JAMA. 2004;292:1064-73. [PMID: 15339900]
43. **Pearlman DN, Rakowski W, Ehrich B, Clark MA.** Breast cancer screening practices among black, Hispanic, and white women: reassessing differences. Am J Prev Med. 1996;12:327-37. [PMID: 8909641]
44. **Selvin E, Brett KM.** Breast and cervical cancer screening: sociodemographic predictors among White, Black, and Hispanic women. Am J Public Health. 2003;93:618-23. [PMID: 12660207]
45. **Fox SA, Stein JA.** The effect of physician-patient communication on mammography utilization by different ethnic groups. Med Care. 1991;29:1065-82. [PMID: 1943268]
46. **Stein JA, Fox SA, Murata PJ.** The influence of ethnicity, socioeconomic status, and psychological barriers on use of mammography. J Health Soc Behav. 1991;32:101-13. [PMID: 1861047]
47. **Tu SP, Yasui Y, Kuniyuki AA, Schwartz SM, Jackson JC, Hislop TG, et al.** Mammography screening among Chinese-American women. Cancer. 2003;97: 1293-302. [PMID: 12599238]
48. **Sickles EA, Miglioretti DL, Ballard-Barbash R, Geller BM, Leung JW, Rosenberg RD, et al.** Performance benchmarks for diagnostic mammography. Radiology. 2005;235:775-90. [PMID: 15914475]
49. **Gill KS, Yankaskas BC.** Screening mammography performance and cancer detection among black women and white women in community practice. Cancer. 2004;100:139-48. [PMID: 14692034]

Current Author Addresses: Dr. Smith-Bindman: Department of Radiology, University of California, San Francisco, 1600 Divisadero Street, Room C250, San Francisco, CA 94115-1667.

Dr. Miglioretti and Ms. Abraham: Group Health Cooperative Center for Health Studies, 1730 Minor Avenue, Suite 1600, Seattle, WA 98101.

Dr. Lurie: RAND Corporation, 1200 South Hayes Street, Arlington, VA 22202.

Dr. Ballard Barbash: National Cancer Institute, Applied Research Program, Division of Cancer Control and Population Sciences, 6130 Executive Boulevard, Bethesda, MD 20892-7344.

Dr. Strzelczyk: School of Medicine, University of Colorado Health Sciences Center, Denver, CO 80262.

Dr. Dignan: University of Kentucky Prevention Research Center, 2365 Harrodsburg Road, Suite B100, Lexington, KY 40504.

Dr. Barlow: Cancer Research and Biostatistics, 1730 Minor Avenue, Suite 1900, Seattle WA 98101.

Ms. Beasley: Department of Nursing, University of North Carolina Pembroke, 1 University Drive, Pembroke, NC 28372.

Dr. Kerlikowske: General Internal Medicine Section, Department of Veterans Affairs, University of California, San Francisco, 4150 Clement Street, 111A1, San Francisco, CA 94121.

Author Contributions: Conception and design: R. Smith-Bindman, D.L. Miglioretti, N. Lurie, R.B. Barbash, J. Strzelczyk, M. Dignan, W.E. Barlow, K. Kerlikowske.

Analysis and interpretation of the data: R. Smith-Bindman, D.L. Miglioretti, N. Lurie, L. Abraham, R.B. Barbash, W.E. Barlow, K. Kerlikowske.

Drafting of the article: R. Smith-Bindman, D.L. Miglioretti, N. Lurie, R.B. Barbash, M. Dignan, W.E. Barlow, K. Kerlikowske.

Critical revision of the article for important intellectual content: R. Smith-Bindman, D.L. Miglioretti, N. Lurie, R.B. Barbash, J. Strzelczyk, M. Dignan, W.E. Barlow, C.M. Beasley, K. Kerlikowske.

Final approval of the article: R. Smith-Bindman, D.L. Miglioretti, N. Lurie, R.B. Barbash, M. Dignan, W.E. Barlow, C.M. Beasley, K. Kerlikowske.

Provision of study materials or patients: M. Dignan, K. Kerlikowske.

Statistical expertise: R. Smith-Bindman, D.L. Miglioretti, L. Abraham, W.E. Barlow.

Obtaining of funding: R. Smith-Bindman, N. Lurie, K. Kerlikowske, D.L. Miglioretti, W.E. Barlow.

Administrative, technical, or logistic support: R. Smith-Bindman.

Collection and assembly of data: R. Smith-Bindman, L. Abraham, C.M. Beasley, K. Kerlikowske, D.L. Miglioretti.

TITLE PAGE

Corresponding author:

Elana Curtis, MB ChB, MPH, FAFPHM[1]

Contact Information: Faculty of Medical and Health Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand. Phone 0064 9 3737599 Fax 0064 9 3035947 Email: e.curtis@auckland.ac.nz

Additional authors:

Chris Quale, PhD [2 3]

David Haggstrom, MD MAS [4]

Rebecca Smith-Bindman, MD [2 3 5]

[1] Visiting 2004-2005 Harkness Fellow in Healthcare Policy (The Commonwealth Fund)

[2] Department of Radiology, University of California, San Francisco, California, USA

[3] Department of Epidemiology and Biostatistics, University of California, San Francisco, California, USA

[4] Department of Internal Medicine, Indiana University School of Medicine, Regenstrief Institute, and Roudebush Veterans' Affairs Medical Center, Indianapolis, IN, United States USA

[5] Department of Obstetrics, Gynecology and Reproductive Medicine, University of California, San Francisco, California, USA

Shortened Title: **R/E Differences in Breast Cancer Survival**

Full Title: **R/E Differences in Breast Cancer Survival: How Much Is Explained By Screening, Tumor Severity, Biology, Treatment, Co-morbidities and Demographics?**

Manuscript category: Original Article

Acknowledgements: This work was supported by the Commonwealth Fund Harkness Fellowship in Healthcare Policy, Department of Defense DAMD 17-99-1-9112, Mount Zion Health Care Fund Dean's Account, the Division of Internal Medicine, University of California. California Breast Cancer Research Program, National Cancer Institute (K07 CA 86032) Additional support was received from Ngā Pae o Te Māramatanga, University of Auckland. Thanks also goes to Dr. Andrew Bindman (University of San Francisco, California), Dr. Ashley Bloomfield (Ministry of Health, New Zealand) and Ms Bridget Robson (Te Ropū Rangahau Hauora a Eru Pōmare, New Zealand).

This study used the linked SEER-Medicare database. The interpretation and reporting of these data are the sole responsibilities of the authors. We acknowledge the efforts of the Applied Research Branch, Division of Cancer Prevention and Population Science, National Cancer Institute; The office of Information Services and the Office of Strategic Planning, HCFA; Information Management Services, Inc; and the SEER program tumor registries in the creation of the SEER-Medicare database.

Financial Disclosures/Conflict of Interests: None

Page count: 35 *Table count:* 2 *Figure Count:* 2

CONDENSED ABSTRACT (2 sentences):

This study found large differences by race and ethnicity in breast cancer survival for elderly women; screening mammography, tumor severity, biology, treatment, co-morbidities and demographics all contributed to these differences. Controlling for these predictor variables reduced almost all of the difference between African American and White women in the All stages analysis and reduced, but did not eliminate, disparities when the analysis was limited to stage II/III disease.

ABSTRACT (word count 289)

Background

The reasons for racial and ethnic differences in breast cancer survival have been difficult to disentangle.

Methods

SEER-Medicare data were used to identify 41,020 women aged ≥68 years with incident breast cancer between 1994-1999 including African American (2,479), Hispanic (1,172), Asian/Pacific Island (1,086), and White women (35,878). A Cox Proportional Hazards Model assessed overall and stage specific (0/I, II/III and IV) racial and ethnic differences in breast cancer survival after adjusting for mammography screening, tumor characteristics at diagnosis, biologic markers, treatment, co-morbidity and demographics.

Results

African American women had worse survival than White women, although controlling for predictor variables reduced the difference between African

American and White women among all stage breast cancer (HR 1.08, [95% CI 0.97-1.20]). Adjustment for predictors reduced, but did not eliminate, disparities in the analysis limited to women diagnosed with stage II/III disease (HR 1.30, [95% CI 1.10-1.54]). Screening mammography, tumor characteristics at diagnosis, biologic markers, and treatment each produced similar reduction in hazard ratios for women with stage II/III cancers. Asian and Pacific Island women had better survival than White women before and after accounting for all predictors (Adjusted All stage HR 0.61, [95% CI 0.47-0.79]); Adjusted stage II/III (HR 0.61, [95% CI 0.47-0.79]). Hispanic women had better survival than White women in All and Stage II/III analysis (All stage HR 0.88, [95% CI 0.75-1.04]); stage II/III (HR 0.88, [95% CI 0.75-1.04]) although these findings did not reach statistical significance. There was no significant difference in survival by race and ethnicity among women diagnosed with stage IV disease.

Conclusions

Predictor variables contribute to, but do not fully explain, racial and ethnic differences in breast cancer survival. Future analyses should further investigate the role of biology, demographics and disparities in quality of care.

Keywords:

1. breast cancer
2. survival
3. race
4. ethnicity

5. mammography screening
6. tumor severity
7. biology
8. treatment
9. co-morbidities
10. demographics

INTRODUCTION

Breast cancer is the most common cancer and the second leading cause of cancer death among women in the US.[1] Although there has been an overall reduction in breast cancer mortality rates in the United States since the 1990s, most of this benefit has been experienced by White (W) women.[2] African American (AA) and Hispanic (H) women remain more likely to be diagnosed with poor prognostic breast cancers (i.e. late stage, large size, lymph node positive, estrogen receptor negative) with AA women experiencing worse survival than W women, a disparity that has increased rather than decreased over time.[3] In contrast, Asian/Pacific Island (A/PI) women tend to have better prognostic breast cancers (i.e. early stage, small size, lymph node negative, estrogen receptor positive) and better survival than W women.[4]

The reasons for the persistence of these racial and ethnic disparities have been difficult to disentangle. Possible explanations include differences in screening mammography leading to differences in the stage and size of tumors at diagnosis[5, 6], tumor biology, inadequate receipt of appropriate breast cancer treatment [5] and underlying patient co-morbidities and socio-economic factors.[7-11]

A number of studies have used Surveillance, Epidemiology and End Results (SEER) program information to explore these issues.[3, 4, 12-18] Although SEER data include valuable regional information on a large, geographically diverse

population, it lacks detailed information on several important factors, such as screening mammography use and underlying co-morbidities that may impact survival.[19] This study aims to overcome these problems by using the combined SEER-Medicare dataset to explore the contribution of screening, tumor characteristics (i.e. stage, size, and lymph node status), biology, treatment type, co-morbidities and demographics to R/E differences in breast cancer survival for elderly women.

METHODS

Data source

Data were obtained from the SEER-Medicare database that combines clinical information with claims information from Medicare, the primary health insurer for 97% of the US population aged ≥65 years.[19] SEER cancer registry data includes information on cancer characteristics and treatment from 11 SEER sites representing ~14% of the US population.[19] Medicare data include health insurance claims data that allows assessment of measurements not possible in SEER data alone, including mammography utilization, co-morbidities, and full treatment.

Study Population

Our study population included all female Medicare beneficiaries diagnosed with incident breast cancer between January 1, 1994 and December 31, 1999. We limited the study to women with at least three years of Medicare enrollment prior to their cancer diagnosis in order to assess the use of screening mammography and the presence of co-morbidities, thus restricting the sample to women aged 68 and older. Medicare does not receive claims for women enrolled in HMO plans, making it impossible to assess the predictor variables in these women. Therefore, we restricted the sample to women enrolled in Medicare's Fee For Service plans.[19] Similarly, mammography coverage is included as part of Part B enrollment and women without Part B enrollment were excluded as we

could not assess mammography screening or outpatient treatment or co-morbidities in these women.[19] Overall 17,364 women were excluded because they had either HMO enrollment or Part B non-enrollment. A total of 41,020 women met inclusion criteria and were included in the analysis.

Variables

Age was calculated using SEER data and grouped into five categories (68-70, 71-75, 76-80, 81-85 and >85). Race and ethnicity was classified according to SEER data as African American, Hispanic, Asian/Pacific Island, Other/Unknown and White. The SEER ethnicity variable used to identify H women is based on a surname-matching algorithm that has greater sensitivity than ethnicity recorded in Medicare data.[20] Race in SEER data is determined from medical records and registration information. The A/PI category represents Chinese, Japanese, Filipino and Hawaiian women. The Other/Unknown category includes Native American data that were not isolated due to small numbers.

Mammography Screening

Each woman was characterized based on her pattern of mammography utilization in the 3 years prior to her breast cancer diagnosis.[21] Because most women with a new diagnosis of breast cancer will undergo mammography around the time of diagnosis, either as a screening examination or as part of the evaluation of known or suspected cancer, we characterized a woman's screening history based on the timing of mammography before this peri-cancer mammogram.[21] The intervals were categorized as within 1 year, 1-2 years, 2-3

years and >3 years (this category included women who had remote mammography or no screening mammography prior to their diagnosis). Adequate screening was defined as screening received within 2 years and inadequate screening as screening received >2 years prior to the peri-cancer mammography, or no screening prior to the time of diagnosis.

Tumor Characteristics at Diagnosis

Tumor characteristics were determined using SEER data and included American Joint Committee on Cancer categories describing cancer stage (0-IV, tumor size (mm) and lymph node status (positive versus negative/unknown).

Biological Measures

We assessed estrogen receptor status using SEER data as (1) positive versus (2) negative (3) borderline (4) unknown and (5) none done.

Histological grade was characterized using SEER data as per the American Joint Committee on Cancer categories as grade I-IV.[22]

Breast Cancer Treatment

Receipt of breast cancer treatment was determined using both SEER and Medicare data.[5, 23-26] In cases of disagreement on the type of surgery the most invasive surgery, i.e. mastectomy as opposed to lumpectomy, noted in either SEER or Medicare was used. Categories included (1) breast conserving surgery

without radiation (2) breast conserving surgery with radiation (3) mastectomy and (4) no surgery.

Co-Morbidity

R/E differences in co-morbidities may produce disparities in the use of mammography screening and breast cancer treatment.[27] Co-morbidity was measured with the Charlson co-morbidity index derived from Medicare inpatient and physician/supplier claims.[28]

Socioeconomic Factors

Income was used as a marker of socio-economic status (SES) and was obtained from SEER data using the median income of residence as recorded in a woman's ZIP code. The type of community a woman lives in may influence her access to breast cancer services and was determined via SEER data using the assigned metropolitan statistical area of (1) rural (2) less rural (3) urban (4) metropolitan and (5) big metropolitan.[29]

Statistical Analysis

A Cox Proportional Hazards Model was used to determine time from breast cancer diagnosis to cancer-specific death among all women with breast cancer and stratified by stage at diagnosis (0/I, II/III, IV). These groups were created because we expected to see similar outcomes among women with stage 0/I and stage IV disease (i.e. excellent prognosis for women with early stage, poor

prognosis for late stage disease) and different outcomes for women with stage II/III breast cancer.

The base model controlled for patient age and SEER site. Additional variables were sequentially added to increasingly more adjusted models, including utilization of screening mammography (adequate versus inadequate); tumor characteristics including size (≤15mm versus ≥16mm), lymph node status (positive versus negative/unknown), stage (0/I versus II/III/IV, except in the stage specific models) biological measures including grade (I/II versus III/IV) and estrogen receptor status (positive versus negative), breast cancer treatment, co-morbidities and demographic variables (income/community type).

We investigated the interaction between screening and age (age ≤75 years versus >75 years) as screening for very elderly women is not routinely recommended and the impact of screening on survival may vary by age. We also investigated the interaction between screening and stage, as the impact of screening on survival could matter differently for different stage disease. We explored the interaction between stage and radiation to account for the fact that radiotherapy for early stage 0/I is not indicated. These potential interactions were included in the model to account for differences in (a) the use of mammography screening by age and stage and (b) appropriate use of radiation by stage.

Women with unknown R/E were included in the statistical modeling. All analyses were conducted using SAS version 8.2.

RESULTS

A total of 41,020 women were included. Of these women, 6% (n=2479) were AA, 2.7% (n=1086) A/PI, 2.9% (n=1172) H, and 87.5% (n=35,878) W (Table 1). 40.1% of the women were aged ≤75 years, with A/PI (54.1%) and H women (46.4%) having a greater proportion of younger women compared to AA (41.4%) and W women (40.0%). AA and H women had a greater proportion of women living in areas with income levels <$30,000 (47% and 21% respectively) compared to A/PI and W women (5% each).

Overall, more than half of this cohort (58%) received inadequate mammography screening with AA (64%) and H (68%) women being more likely to be inadequately screened than A/PI (57%) and W (57%) women. African American and Hispanic women were also more likely to have advanced and poor prognostic tumors at the time of diagnosis.

AA women had the highest proportion of women who received no surgical treatment (11.5% versus 8.2% H, 6.8% W and 3.3% A/PI) or who received breast-conserving surgery without radiation (19.5% versus 15.8%W, 11.9% H, and 11.9% A/PI).

Cancer-specific mortality

Hazard ratios for cancer-specific death are provided in Table 2. The baseline model adjusts for age and SEER site only, while the fully adjusted model adjusts

for mammography utilization, tumor characteristics, tumor biology, treatment type, co-morbidities, and demographics. Results are presented for All stage Disease and then stratified by stage (0/I, II/III, IV).

All Stages Results

For All stages of disease, AA women had a significantly higher risk of death than W women at baseline (HR 1.63 [95% CI 1.48-1.80]), however this was considerably reduced, and no longer significant (HR 1.08, [95% CI 0.97-1.20]) after full adjustment. H women had a significantly higher risk of death at baseline (HR 1.24, [95% CI 1.06-1.46]), yet after full adjustment had a non-significant 12% lower risk of death than W women (HR 0.88, [95% CI 0.75-1.04]). A/PI women had a 41% significantly lower risk of death than W women at baseline (HR 0.59, [95% CI 0.35-0.77]) that changed very little after full adjustment (HR 0.61, [95% CI 0.47-0.79]).

Stage Specific Results

In the fully adjusted models for stage 0/I, only A/PI women had a statistically significant different risk of death than W women (HR 0.44, [95% CI 0.24-0.81]). For all other groups there were no significant differences in stage 0/I (AA HR: 1.19, [95% CI 0.91-1.55], H HR 0.85, [95% CI 0.57-1.27]).

For stage II/III, AA women had a 66% higher risk of mortality than W women at baseline (HR 1.66, [95% CI 1.43-1.93]). While controlling for all variables reduced

the mortality risk substantially, it did not eliminate their increased hazard ratio leaving a 30% increased risk after full adjustment (HR 1.30, [95% CI 1.10-1.54]). H women had a non-significant similar risk of death at baseline (HR 1.07, [95% CI 0.84-1.36)] that reduced to a 10% non-significant lower risk of death with full adjustment (HR 0.90, [95% CI 0.71-1.15]). A/PI women had a 37% lower risk of mortality than W women after full adjustment (HR 0.63, [95% CI 0.43-0.93]), showing little change from baseline (HR 0.69, [95% CI 0.46-1.10]).

There were no statistically significant differences by R/E for stage IV results.

Contributing Factors

Tumor severity accounted for ~1/3rd (29%) of the mortality reduction from baseline to full adjustment between AA and W women in the all stages analysis, followed by screening (15%), treatment (5%), biology (4%), co-morbidities (2%), and demographics (2%) (Figure 1). For H women, screening (16%) and tumor severity (15%) accounted ~1/3rd of the difference.

For stage II/III results, screening mammography (8%), tumor severity (6%), biology (9%), and breast cancer treatment (6%) all produced a similar sized reduction in the mortality difference for AA women (Figure 2). A similar pattern was seen for H women.

It should be noted that the order the variables were placed in the model would have effected the percentage reduction in hazard ratio attributable to each variable. However, when we made changes in the order of the variables within the model, we saw relatively little difference in the magnitude of effect.

DISCUSSION

This study found large differences by R/E in breast cancer survival for elderly women. Screening mammography, tumor severity, biology, treatment, co-morbidities and demographics all contributed to these differences. Controlling for these predictor variables reduced almost all of the difference between AA and W women in the all stages analysis and reduced, but did not eliminate, disparities in the stage II/III analysis.

In contrast, A/PI women had a consistently better survival profile than W women in all analyses that did not reduce with the addition of predictor variables. This may be because there were no large differences between A/PI women and W women in the predictor variables measured. No statistically significant differences in survival were observed between H and W women, although the overall pattern shows H women having better survival than W women after full adjustment in both all stages and stage II/III analyses.

All women diagnosed with stage IV breast cancers had a similar risk of death regardless of R/E. This likely reflects the small number or women diagnosed with stage IV breast cancers and the poor prognosis in this group. Mention state 0/1 here

Screening mammography is known to reduce the death rate from breast cancer in the general US population[30] and disparities in screening

mammography have been shown to heavily contribute to R/E disparities in breast cancer survival.[6, 31, 32] Our findings show persistent differences by R/E in the utilization of screening mammography. These results are consistent with recent reports[21, 33] and may be due to the persistent underutilization of mammography in this elderly population.

Our findings show that, although the use of adequate screening will reduce differences in survival between AA and W women, it will not eliminate them. Despite not being the sole remedy, screening nevertheless remains important. It is possible that if H women had greater utilization of adequate screening then their risk of breast cancer death could drop further below W women as is seen for A/PI women. AA differences in screening interval accounted for a considerable portion of the mortality difference with W women. Addressing this disparity should therefore reduce AA:W differences in breast cancer survival.[34]

Stage at diagnosis, the strongest predictor of breast cancer survival, is known to contribute to R/E survival disparities.[4] Our findings support this hypothesis given the importance of tumor severity in reducing mortality disparities between AA and W women.

Biological markers (tumor grade, estrogen receptor status) are known to be different between AA and W women, suggesting that they may also contribute to differences in survival.[18] In this study, biological markers had an effect of

similar magnitude on survival results as did screening, tumor characteristics, and treatment. In exploring the role of biology further, it is important to understand the determinants of tumor biology given the ongoing debate as to whether biology reflects genetic causes versus exposure to less favorable environmental conditions.[35, 36]

We found persistent differences in treatment by race and ethnicity, with African American women more likely to have poor quality treatment. Adjusting for differences in treatment receipt reduced AA:W differences in survival. This finding has been observed elsewhere and may reflect differential access to optimal care, including receipt of adjuvant radiotherapy with breast conserving surgery.[21, 26, 37, 38] R/E disparities in access to optimal treatment, and therefore survival, may reflect differences in healthcare access, regional variations, or a higher disease burden.[5] However, our findings among a Medicare population accounted for insurance type, SEER site, type of community, co-morbidities, and tumor severity and still found differences.

Of note, co-morbidities and demographics contributed relatively little to overall R/E differences noted. However, these factors could each act in a common pathway to other factors as underserved/poor women are less likely to undergo mammography.[39]. Because the exact contribution of each variable to the reduction in hazard ratio remains dependent on the model order it is important to

consider these findings within the larger context that highlights a similar reduction in hazard risk for screening, biology and treatment.

Despite the large number of variables considered in our model, there were persistent differences in stage II/III disease. Other untested hypotheses should be considered, particularly the possibility of R/E differences in the timeliness and quality of breast cancer care.[40] Reasons for these differences may reflect the patient-physician relationship, patient preferences and/or institutional and interpersonal racism across healthcare service provision.[40-42] Differences in quality may occur in some contexts more often than others, partially explaining why AA:W differences persist within stage II/III only.[40] For example, R/E differences in the quality of breast cancer care are unlikely to occur when all women are expected to do well or less well i.e. early and late stage but become unmasked in situations where timeliness and quality are more likely to effect outcomes, i.e. stage II/III cancers.

Strengths and Limitations

Due to small numbers of mortality events, we chose R/E categories that represented aggregates of unique sub-populations that may mask disparities within each category. This is of particular concern for Native American, A/PI and H categories because sub-groups within these categories (i.e. Hawaiian, Mexican, South/Central American, Puerto Rican) have greater breast cancer mortality and poorer survival than W women.[4, 43]. The potential for underserved sub-groups

within the African-American population also exists with Caribbean women having a different screening profile than other women identified as AA. [44, 45] Future analyses should disaggregate the R/E categories so that we can better understand these findings.[11, 43]

The use of aggregate based measures to assign individual socio-economic status is suboptimal.[20] However, aggregate measures of income at the community level remains valuable as community-level predictors of SES have been shown to be a strong predictor of health care services and health outcomes for all individuals living in that community (both high and low income).[46] Our population all have health insurance and we expect that the importance of demographics may differ in a less insurance homogenous sample.

Data on chemotherapy use was incomplete, and quality indicators such as timeliness are difficult to measure within the SEER-Medicare database.[19] Additional biological markers were not available, i.e. HER2 status, CYP1A1, P53 mutations, although there is no conclusive evidence of marker differences between AA and W women.[9]

This analysis improves on studies using SEER data alone as a more accurate assessment of screening exposure was used (rather than self report that overestimates exposure), and co-morbidity and treatment information were

included.[21, 47, 48] This study therefore responds to calls in the literature for analyses that include these variables.[7, 9, 10, 36, 49]

Conclusion

Understanding the reasons for R/E disparities in breast cancer survival remains complex. Despite the complexity, there are still areas in which health policy should clearly intervene. Enough evidence now exists such that policy initiatives should urgently be applied to improve access to adequate screening for AA and H women and increase access to appropriate treatment for AA women in particular. Further multidisciplinary investigation into the role of biology, demographics and potential disparities in quality of care are required in both young and elderly cohorts of women.

REFERENCES:

1. Jemal A, Murray T, Ward E, et al. Cancer statistics, 2005. *CA Cancer J Clin,* 2005;55:10-30.
2. Weir H, Thun M, Hankey B, et al. Annual report to the nation on the status of cancer, 1975-2000, featuring the uses of surveillance data for cancer prevention and control. *J Natl Cancer Inst,* 2003;95:1276-1299.
3. Ghafoor A, Jemal A, Ward E, et al. Trends in breast cancer by race and ethnicity. *CA Cancer J Clin,* 2003;53:342-355.
4. Li CI, Malone KE, Daling JR. Differences in breast cancer stage, treatment, and survival by race and ethnicity. *Arch Intern Med,* 2003;163:49-56.
5. Haggstrom D, Quale C, Smith-Bindman R. Differences in the Quality of Breast Cancer Care Among Vulnerable Populations. *Cancer,* 2005;104:2347-2358.
6. McCarthy E, Burns R, Coughlin S. Mammography use, breast cancer stage at diagnosis, and survival among older women. *J Am Geri Soc,* 2000;48:1226-1233.
7. Newman LA, Griffith KA, Jatoi I, et al. Meta-Analysis of Survival in African American and White American Patients With Breast Cancer: Ethnicity Compared With Socioeconomic Status. *J Clin Oncol,* 2006;24:1342-1349.
8. Royak-Schaler R, Rose D. Mammography screening and breast cancer biology in African American women - a review. *Cancer Detect Prev,* 2002;26:180-191.

9. Polite BN, Olopade OI. Breast cancer and race: a rising tide does not lift all boats equally. *Perspect Biol Med,* 2005;48:166-175.

10. Campbell JB. Breast cancer-race, ethnicity, and survival: a literature review. *Breast Cancer Res Treat,* 2002;74:187-192.

11. Lin SS, Clarke CA, Prehn AW, et al. Survival differences among Asian subpopulations in the United States after prostate, colorectal, breast, and cervical carcinomas. *Cancer,* 2002;94:1175-1182.

12. Baquet CR, Commiskey P. Socioeconomic factors and breast carcinoma in multicultural women. *Cancer,* 2000;88:1256-1264.

13. Edwards MJ, Gamel JW, Vaughan WP, et al. Infiltrating ductal carcinoma of the breast: the survival impact of race. *J Clin Oncol,* 1998;16:2693-2699.

14. Joslyn SA. Racial differences in treatment and survival from early-stage breast carcinoma. *Cancer,* 2002;95:1759-1766.

15. Joslyn SA. Hormone receptors in breast cancer: racial differences in distribution and survival. *Breast Cancer Res Treat,* 2002;73:45-59.

16. Li CI, Malone KE, Daling JR. Differences in breast cancer hormone receptor status and histology by race and ethnicity among women 50 years of age and older. *Cancer Epidemiol Biomarkers Prev,* 2002;11:601-607.

17. Joslyn SA, West MM. Racial differences in breast carcinoma survival. *Cancer,* 2000;88:114-123.

18. Henson DE, Chu KC, Levine PH. Histologic grade, stage, and survival in breast carcinoma: comparison of African American and Caucasian women. *Cancer,* 2003;98:908-917.

19. Warren J, Klabunde C, Schrag D. Overview of the SEER-Medicare Data: Content, Research Applications, and Generalizability to the United States Elderly Population. *Med Care,* 2002;40:3-18.

20. Bach P, Guadagnoli E, Schrag D. Patient demographic and socioeconomic characteristics in the SEER-Medicare database. *Med Care,* 2002;40:19-25.

21. Smith-Bindman R, Miglioretti DL, Lurie N, et al. Does Utilization of Screening Mammography Explain Racial and Ethnic Differences in Breast Cancer? *Ann Int Med,* 2006;144:541-553.

22. Rosenberg J, Chia YL, Plevritis S. The Effect of Age, Race, Tumor Size, Tumor Grade, and Disease Stage on Invasive Ductal Breast Cancer Survival in the U.S. SEER database. *Cancer Res Treat,* 2005;89:47-54.

23. Cooper G, Virnig B, Klabunde C. Use of SEER-Medicare data for measuring cancer surgery. *Med Care,* 2002;40:43-47.

24. Du X, Freeman J, Warren J. Accuracy and completeness of Medicare claims data for surgical treatment of breast cancer. *Med Care,* 2000;38:719-727.

25. Cooper G, Zhong Y, Stange K. Agreement of Medicare claims and tumor registry data for assessment of cancer-related treatment. *Med Care,* 2000;38:411-421.

26. Ballard-Barbash R, Potosky A, Hankey B. Factors associated with surgical and radiation therapy for early stage breast cancer in older women. *J Natl Cancer Inst,* 1996;88:716-26.

27. West DW, Satariano WA, Ragland DR, et al. Comorbidity and breast cancer survival: a comparison between black and white women. *Ann Epidemiol,* 1996;6:413-419.

28. Klabunde CN, Potosky AL, Legler JM, et al. Development of a comorbidity index using physician claims data. *J Clin Epidemiol,* 2000;53:1258-1267.

29. Tropman S, Ricketts T, Paskett E. Rural breast cancer treatment: evidence from the Reaching Communities for Cancer Care (REACH) project. *Breast Cancer Res Treat,* 1999;56:59-66.

30. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med,* 2005;353:1784-1792.

31. McCarthy E, Burns R, Coughlin S. Mammography use helps to explain differences in breast cancer stage at diagnosis between older black and white women [see comments]. *Ann Intern Med,* 1996;125:173-182.

32. Freeman H. The meaning if race in cancer of the breast. *The Cancer Journal from Scientific American,* 1997:76-77.

33. Kagay C, Quale C, Smith-Bindman R. Screening mammography in the American elderly. *Am J Prev Med,* 2006;31:142-149.

34. Newman LA. Breast Cancer in American Women. *The Oncologist,* 2005;10:1-14.

35. Kreiger N. Refiguring "race": epidemiology, racialised biology, and biological expressions of race relations. *Int J Health Services,* 2000;30:211-216.

36. Newman LA, Mason J, Cote D, et al. African-American ethnicity, socioeconomic status, and breast cancer survival: a meta-analysis of 14 studies involving over 10,000 African-American and 40,000 White American patients with carcinoma of the breast. *Cancer,* 2002;94:2844-2854.

37. Mandelblatt J, Kerner J, Hadley J. OPTIONS (Outcomes and Preferences for Treatment in Older Women Nationwide Study). Variations in breast carcinoma treatment in older medicare beneficiaries: is it black or white. *Cancer,* 2002;95:1401-1414.

38. Prehn AW, Topol B, Stewart S, et al. Differences in Treatment Patterns for Localised Breast Carcinoma among Asian/Pacific Islander Women. *Cancer,* 2002;95:2268-2275.

39. O'Malley M, Earp J, Hawley S, et al. The association of race/ethnicity, socioeconomic status, and physician recommendation for mammography: who gets the message about breast cancer screening? *Am J Public Health,* 2001;91:49-54.

40. Smedley B, Stith A, Nelson A, eds. *Unequal Treatment-confronting racial and ethnic disparities in health care.* 2002, Institute of Medicine, National Acadamy Press: Washington.

41. Maly R, Leake B, Silliman R. Health Care Disparities in Older Patients with Breast Carcinoma. Informational Support from Physicians. *Cancer,* 2003;97:1517-1527.

42. Katz S, Lantz PM, Janz NK, et al. Patient Involvement in Surgery Treatment Decisions for Breast Cancer. *J Clin Oncol,* 2005;23:5526-5533.

43. Meng L, Maskarinec G, Wilkens L. Ethnic differences and factors related to breast cancer survival in Hawaii. *Int J Epidemiol,* 1997;26:1151-1158.

44. Garbers S, Chiasson MA. Breast cancer screening and health behaviors among African American and Caribbean Women in New York City. *J Health Care Poor Underserved,* 2006;17:37-46.

45. Consedine NS, Magai C, Horton D, et al. Health belief model factors in mammography screening: testing for interactions among subpopulations of Caribbean women. *Ethn Dis,* 2005;15:444-452.

46. Krieger N, Chen JT, Waterman PD, et al. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures--the public health disparities geocoding project. *Am J Public Health,* 2003;93:1655-71.

47. Gordon NP, Hiatt RA, Lampert DI. Concordance of self-reported data and medican record audit for six cancer screening procedures. *J Natl Cancer Inst,* 1993;85:566-570.

48. McGovern P, Lurie N, Margolis K, et al. Accuracy of Self-Report of mammography and Pap Smear in a Low-Income Urban Population. *Am J Prev Med,* 1998;14:201-208.

49. Shavers VL, Harlan LC, Stevens JL. Racial/ethnic variation in clinical presentation, treatment, and survival among breast cancer patients under age 35. *Cancer,* 2003;97:134-147.

Tables and Figures

Table 1. Comparison of various characteristics among 41,020 women with breast cancer by R/E, 1994-1999.

Footnote:

k = x$1000

Table 2. Hazard ratios of cancer-specific mortality after breast cancer diagnosis by stage and R/E, 1994-1999.

Footnotes:

All analyses include the Unknown R/E category.

Figures in bold are statistically significant

Figure 1. Risk of cancer-specific death by R/E for all women i.e. all stages, 1994-1999.

Figure 2. Risk of cancer-specific death by R/E for women with stage 2/3 breast cancers at time of diagnosis, 1994-199