

**HHS PUBLIC ACCESS**

Author manuscript

Med Care. Author manuscript; available in PMC 2017 March 01.

Published in final edited form as:

Med Care. 2016 March ; 54(3): e15–e22. doi:10.1097/MLR.0b013e3182a303d7.**Validation of a Medicare Claims-based Algorithm for Identifying Breast Cancers Detected at Screening Mammography****Joshua J. Fenton, MD, MPH¹, Tracy Onega, PhD², Weiwei Zhu, MS³, Steven Balch, MA, MBA³, Rebecca Smith-Bindman, MD, MPH^{4,5}, Louise Henderson, PhD⁷, Brian L. Sprague, PhD⁸, Karla Kerlikowske, MD^{5,6}, and Rebecca A. Hubbard, PhD^{3,9}**¹Department of Family and Community Medicine, Center for Healthcare Research and Policy, and the Comprehensive Cancer Center, University of California, Davis, Sacramento, CA²Norris Cotton Cancer Center, Section of Biostatistics and Epidemiology, Dartmouth Medical School, Lebanon, NH³Group Health Research Institute, Seattle, WA⁴Department of Radiology, University of California, San Francisco, San Francisco, CA⁵Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA⁶Department of Medicine, University of California, San Francisco, San Francisco, CA**Correspondence to:** Joshua Fenton, MD, MPH, Department of Family and Community Medicine, University of California, Davis, 4860 Y Street, Suite 2300, Sacramento, CA 95817, (916) 734-3164; Fax (916) 734-5641, joshua.fenton@ucdmc.ucdavis.edu.**Conflicts of Interest:** None declared.**COMPLETE AUTHOR INFORMATION**

1. Joshua J. Fenton (see above for complete contact information)
2. Tracy Onega, Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth Medical School, Hanover, NH, phone (603) 650-3516, Tracy.L.Onega@Dartmouth.edu
3. Weiwei Zhu, Group Health Research Institute, 1730 Minor Avenue, Seattle, WA 98101, Phone (206) 287-4870, zhu.w@ghc.org
4. Steven Balch, Group Health Research Institute, 1730 Minor Avenue, Seattle, WA 98101, Phone (206) 287-4870, balch.s@ghc.org
5. Rebecca Smith-Bindman, Departments of Radiology and Epidemiology and Biostatistics University of California, San Francisco, Box 1667, San Francisco, CA 94143-1667 (415) 253-2573; fax (415) 885-7876, rebecca.smith-bindman@radiology.ucsf.edu
6. Louise Henderson, Department of Radiology, University of North Carolina, 2006 Old Clinic, CB #7510, Chapel Hill, NC 27599, (919) 843-7799, louise_henderson@med.unc.edu
7. Brian Sprague, Department of Surgery, University of Vermont, UHC Room 4425, 1 South Prospect Street, Burlington, VT, 05401, (802) 656-4112, Brian.Sprague@uvm.edu
8. Karla Kerlikowske, Departments of Internal Medicine and Epidemiology, University of California, San Francisco, VAMC 111A1, San Francisco, CA 94143, (415) 750-2093, karla.kerlikowske@ucsf.edu
9. Rebecca Hubbard, Group Health Research Institute, 1730 Minor Avenue, Seattle, WA 98101, Phone (206) 287-4870, hubbard.r@ghc.org

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

⁷Department of Radiology, University of North Carolina, Chapel Hill, NC

⁸Health Promotion Research, University of Vermont, Burlington, VT

⁹Department of Biostatistics, University of Washington, Seattle, WA

Abstract

Background—The breast cancer detection rate is a benchmark measure of screening mammography quality, but its computation requires linkage of mammography interpretive performance information with cancer incidence data. A Medicare claims-based measure of detected breast cancers could simplify measurement of this benchmark and facilitate mammography quality assessment and research.

Objectives—To validate a claims-based algorithm that can identify with high positive predictive value (PPV) incident breast cancers that were detected at screening mammography.

Research Design—Development of a claims-derived algorithm using classification and regression tree analyses within a random half-sample of Medicare screening mammography claims followed by validation of the algorithm in the remaining half-sample using clinical data on mammography results and cancer incidence from the Breast Cancer Surveillance Consortium (BCSC).

Subjects—Female fee-for-service Medicare enrollees age 68 years and older who underwent screening mammography from 2001 to 2005 within BCSC registries in four states (CA, NC, NH, and VT), enabling linkage of claims and BCSC mammography data (N=233,044 mammograms obtained by 104,997 women).

Measures—Sensitivity, specificity, and PPV of algorithmic identification of incident breast cancers that were detected by radiologists relative to a reference standard based on BCSC mammography and cancer incidence data.

Results—An algorithm based on subsequent codes for breast cancer diagnoses and treatments and follow-up mammography identified incident screen-detected breast cancers with 92.9% sensitivity (95% CI: 91.0%-94.8%), 99.9% specificity (95% CI: 99.9%-99.9%), and a PPV of 88.0% (95% CI: 85.7%-90.4%).

Conclusions—A simple claims-based algorithm can accurately identify incident breast cancers detected at screening mammography among Medicare enrollees. The algorithm may enable mammography quality assessment using Medicare claims alone.

Keywords

Breast Cancer Screening; Mammography; Validation Studies; Medicare; Quality Assessment

INTRODUCTION

Core measures of mammography performance, such as recall rates and breast cancer detection rates, vary widely across U.S. radiologists and mammography facilities (1). In addition, despite an average recall rate in the U.S. that is nearly twice that of most European nations, breast cancer detection rates after screening mammography are no higher in the

U.S. than in Europe (2). These findings point to potentially remediable gaps in the quality of U.S. mammography services.

One issue may be differences in radiologist mammography experience and annual interpretive volume in the U.S. vs. European countries (3), and some have suggested benchmarks for screening mammography performance as a means of prompting radiologist- or facility-level efforts to improve interpretation (4, 5). The Institute of Medicine has also emphasized the need for high quality measures of provider- and facility-level performance to guide targeted quality improvement efforts (6). The U.S. Medicare program has a particularly large stake in mammography quality improvement. Medicare pays for over 8.5 million mammograms each year at a cost of over \$1 billion annually (7), and over half of incident breast cancers occur among women older than 65 years (8). Currently, Medicare reports facility-level performance of follow-up or recall rates as part of the Hospital Quality Reporting Program. However, additional valid and reliable performance measures are needed to facilitate broader scale improvement.

Medicare claims data are one potential data source for measuring mammography interpretative performance. The breast cancer detection rate, for example, is a widely used measure of radiologist- and facility-level interpretive performance; a cancer detection rate of 2.5 breast cancers per 1000 screening mammograms is considered a minimally acceptable performance level (5). This measure, however, is challenging to implement, because it requires accurate identification of: 1) screening mammograms, 2) subsequent incident breast cancers, and 3) screening mammography interpretation. With regard to identifying screening mammograms using Medicare data, we recently validated a claims-based algorithm for distinguishing screening from diagnostic mammograms (9). The distinction between screening and diagnostic is critical, because the incidence of breast cancer is much higher following diagnostic than screening mammography (4, 10). However, with regard to the second step, there has been uncertainty about the accuracy of claims for identifying incident (rather than prevalent) breast cancers (11, 12). Diagnosis codes for breast cancer are not specific for incident rather than prevalent disease, and procedure codes for breast cancer treatments, such as mastectomy, can occur when women are treated for either benign conditions or recurrent breast cancer. Finally, claims provide no direct information to infer whether mammograms were interpreted as normal or abnormal, potentially limiting the ability to infer that incident cancers were detected at screening mammography.

Nattinger et al. previously validated a clinically informed, four-step algorithm for distinguishing incident from prevalent breast cancers using Medicare claims data (13). Relative to a reference standard based on Surveillance, Epidemiology, and End Results (SEER) cancer registry data, the algorithm achieved a sensitivity of ~80% and a positive predictive value of over 90%. However, the algorithm was not validated in a cohort undergoing screening mammography, so it is uncertain if the algorithm accurately identifies breast cancers in a screening population, **nor** does the algorithm address screening mammography interpretation. Additionally, the algorithm was validated using data from 1994, and it is uncertain if it achieves similar performance with more recent Medicare claims.

We capitalized on the linkage of high-quality mammography and cancer incidence data from the Breast Cancer Surveillance Consortium with Medicare claims to evaluate the performance of claims-based methods for identifying incident breast cancer that followed positive screening mammography. We first assessed the performance the approach described by Nattinger et al. (13) within a screening mammography cohort. (Henceforth, we refer to this as the Nattinger algorithm.) We then assessed whether classification and regression tree analyses could identify an alternative approach with improved performance. We hypothesized that claims-derived algorithms could identify incident cancers following true-positive screening mammography with high predictive value in recent Medicare claims.

METHODS

Data

We used data from Medicare claims files (the Carrier Claims, Outpatient, and Inpatient files) and the Medicare denominator file, which provides demographic, enrollment, and vital status data. While Medicare mammography claims typically appear in the Carrier file, we assessed both the Carrier and Outpatient files to capture the minority of claims present only in the Outpatient file (~3%) (14). We used Healthcare Common Procedure Coding System (HCPCS) procedure codes to identify bilateral mammograms and procedures occurring before and after mammography. Medicare claims also include International Classification of Diseases, 9th Edition, Clinical Modification (ICD-9-CM) diagnosis codes, which we used to identify breast cancer diagnosis codes and breast symptoms.

Medicare claims from 1998 to 2006 were linked with BCSC mammography data derived from regional mammography registries in four states (North Carolina; San Francisco Bay Area, CA; New Hampshire; and Vermont) (<http://breastscreening.cancer.gov/>). Claims and BCSC mammography data were matched using a deterministic algorithm derived by the National Center for Health Statistics that is based on social security numbers (when available), names, and birthdates. Among women aged 65 years and older with a BCSC mammogram between 1998 and 2006, 87% were successfully matched to Medicare claims. Women who did not successfully match typically received mammography at facilities that did not report social security numbers to the BCSC.

BCSC facilities transmit prospectively collected patient and mammography data to regional registries, which link the data to breast cancer outcomes ascertained from regional or statewide cancer registries and pathology data. Data are pooled at a central Statistical Coordinating Center (SCC). BCSC mammography data include information on examination purpose (screening vs. diagnostic), radiologist interpretation, and patient socio-demographics and breast health history. The BCSC has established standard definitions for key variables and multiple levels of data quality control and monitoring (15). Each registry and the SCC have received institutional review board approval for either active or passive consenting processes or a waiver of consent to enroll participants, link data, and perform analytic studies. All procedures are Health Insurance Portability and Accountability Act (HIPAA) compliant and all registries and the SCC have received a Federal Certificate of Confidentiality to protect the identities of patients, physicians, and facilities.

Subjects

We identified mammograms captured in both Medicare claims and the BCSC among women who were aged 68 or older on mammography dates from January 1, 2001 to December 31, 2005. We identified screening mammograms (distinguishing them from diagnostic mammograms) using a validated claims-based algorithm based upon the HCPCS mammogram codes, claims for mammography in the prior nine months, and claims with ICD-9-CM codes for breast cancer during the year prior to mammography (9). Relative to a BCSC-derived reference standard, the positive predictive value of the algorithm's screening designation was ~95%. We used the claims-based algorithm (rather than BCSC data) to identify screening mammograms so that results would be potentially generalizable to other Medicare claims-based studies where linkage of claims to mammography registry data is not feasible.

We selected mammograms for women with continuous enrollment in fee-for-service Medicare (parts A and B) for twelve months after mammography and three years prior to mammography, enabling both prospective assessment of outpatient claims for clinical events that might indicate incident breast cancer following abnormal screening mammography, and retrospective assessment for claims indicating prevalent breast cancer. Hence, we excluded women aged 65 to 68 years because only women aged 68 years or greater consistently have three years of prior claims. Following Nattinger, et al. (13), we considered a woman to have prevalent cancer if any ICD-9-CM codes for invasive breast cancer or ductal carcinoma in-situ appeared on claims in the three years prior to mammography (see Table, Supplemental Digital Content 1, which gives claims and diagnostic codes used in developing study variables). Although some women with prevalent breast cancer continue to receive screening mammography, we excluded these mammograms so that our sample represented women undergoing screening who had no breast cancer in the prior three years. We randomly divided the sample into two half-samples, one for training and one for validation of classification and regression tree (CART) analyses.

Reference Standard

We used BCSC cancer registry data to identify incident breast cancers and BCSC mammography data to identify radiologists' mammography interpretation. We defined incident breast cancers as those with a diagnosis date within one year of the date of screening mammography. Following standard BCSC definitions (available at: http://breastscreening.cancer.gov/data/bcsc_data_definitions.pdf), a positive mammogram was defined as one with a Breast Imaging Reporting and Data System (BI-RADS®) assessment of 0 (needs additional imaging evaluation), 3 (probably benign finding) with a recommendation for immediate follow-up, 4 (suspicious abnormality), or 5 (highly suggestive of malignancy)(16). Mammograms with other BI-RADS® assessments were defined as negative, because no immediate evaluation was recommended at the time of screening mammography. A screen-detected breast cancer was defined as a positive mammogram with a diagnosis of incident breast cancer within one year.

Claims-Based Algorithms for Identifying Incident Screen-Detected Breast Cancers

We first adapted the Nattinger algorithm to identify incident breast cancers during the one year following screening mammography (13). The Nattinger algorithm begins by identifying “high-probability” cases based on codes for breast cancer diagnoses and therapies during the year following the index date (in this instance, the date of mammography). Next, the algorithm attempts to exclude women treated for benign breast disease, cancers that were metastatic to the breast, and, finally, women with prevalent breast cancer based on breast cancer diagnosis codes occurring during three years of claims prior to the index date. We classified all cases identified by the algorithm as screen-detected incident cancers. Although we recognize that the Nattinger algorithm was not designed to consider mammography interpretation, we felt that it was valuable to evaluate its performance in the context of recent screening mammography.

We then conducted CART analyses in attempts to identify an algorithm with improved performance (17). CART is a non-parametric decision tree methodology that identifies sequential binary partitions in independent variables to optimize prediction of the dependent variable (in this case, incident breast cancer detected at screening mammography). A potential advantage is that CART can specify the best-performing time intervals following mammography for ascertaining breast cancer diagnosis or treatment codes following incident breast cancers that were detected at mammography.

Potential independent variables for CART analyses were derived from claims on the mammogram date or during a one-year post-mammogram follow-up, including: patient age; **days to any** subsequent mammogram; days to diagnosis codes for breast cancer, ductal carcinoma in-situ, benign breast tumors, secondary cancers of the breast, or a personal history of breast cancer; days to procedural or diagnostic codes for breast biopsies; days to any procedural codes for breast-directed surgery (lumpectomy, partial mastectomy, or mastectomy), axillary lymph node biopsy or resections; and days to any codes for breast radiation (18, 19). (See Table, Supplemental Digital Content 1, for claims and diagnostic codes used in developing study variables.)

We performed CART analyses on the training half-sample of mammograms. The CART algorithm selected splits in independent variables on the basis of the Gini index, and continued growing the tree until no further splits improved the Gini index by more than 0.0001 (17). To minimize over-fitting, we pruned the tree to optimal complexity based on cross-validation.

Analyses of Classification Accuracy

Within the validation sub-sample, we created cross-tabulations to compare the identification of incident screen-detected breast cancers using claims-based algorithms versus the reference standard. We quantified accuracy using: sensitivity (the proportion of incident screen-detected cancers that were identified by the algorithm); specificity (the proportion of mammograms without incident screen-detected cancers that were so classified by the algorithm); positive predictive value (PPV, or the proportion of algorithmically-identified incident screen-detected cancers that were also classified as such by the reference standard);

and negative predictive value (the proportion of mammograms classified as having no incident screen-detected cancer also classified as having no incident screen-detected cancer by the reference standard). As a practical test of the algorithm's accuracy, we compared unadjusted relative rates of breast cancer detection as computed using the CART algorithm and BCSC data (i.e., the reference standard) within subgroups based on age, race/ethnicity, non-urban vs. urban residence, and Medicaid eligibility. A woman was classified as Medicaid eligible if the Medicare denominator file indicated at least one month of Medicaid eligibility in the three years before or one year after the index mammogram. Non-urban vs. urban residence was determined using the Rural-Urban Commuting Area Code for the zip code of the woman's primary residence. We computed 95% confidence intervals around all point estimates. We performed statistical analyses using the rpart package in R, version 2.12.0 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Mammogram Samples

We identified a sample of 233,044 screening mammograms with linked Medicare claims and BCSC records. The mammograms were obtained by 104,997 women who received an average of 2.2 mammograms during the study period (range: 1-7). On the date of mammography, women had a mean age of 75.2 years (SD: 5.3; range 68 to 104). Based on BCSC interpretation and cancer data, radiologists detected breast cancer on 1,384 of the mammograms; the breast cancer detection rate in the sample was 5.9 cancers per 1000 screening mammograms. The patient sample was ethnically diverse with substantial representation in both non-urban and urban settings (Table 1).

Performance of Claims-Based Algorithms

Using CART analyses, we identified a claims-based algorithm that classified women based on whether breast cancers were detected as screening mammography (Figure 1). Assessing claims for one year following screening mammography, the algorithm classifies a mammogram as having detected breast cancer if there is a claim diagnosis code for breast cancer within 123 days and any claim for breast-directed surgery within one year. Alternatively, if there is no claim diagnosis code for breast cancer within 123 days, the algorithm classifies a mammogram as having detected incident breast cancer if there is a claim diagnosis code for ductal carcinoma in-situ within 286 days and a claim for another mammogram within 82 days. The algorithm classifies all other mammograms as not detecting incident breast cancer. (See Supplemental Digital Content 2 for programming code for implementing the algorithm in R.)

We compared the performance of the Nattinger algorithm and the CART-derived algorithm in identifying incident breast cancer detected at screening mammography (Table 2). While the Nattinger algorithm had a sensitivity of 83.75%, a specificity 99.86%, and a PPV of 77.45%, the CART-derived algorithm had a sensitivity of 92.89% and slightly higher specificity (99.93%). Although small, this difference results in fewer cases falsely classified as detected breast cancers using the CART-derived algorithm as compared to the Nattinger algorithm and a higher PPV (88.03% vs. 77.45%).

In analyses stratified by age, race/ethnicity, Medicaid eligibility, and non-urban vs. urban residence, estimates of relative rates of incident detected breast cancer using the algorithm were similar to reference standard estimates based on BCSC data (Table 3).

DISCUSSION

In a sample of screening mammograms with corresponding Medicare claims and mammography registry data, a claims-based algorithm identified screening mammograms with detected incident breast cancers with a sensitivity greater than 90%, a specificity over 99.9%, and a PPV of 88%. Algorithm performance was generally similar across population subgroups and enabled accurate estimates of between-group relative rates of incident breast cancer detection.

Both the CART-derived algorithm and the algorithm proposed by Nattinger, et al. had very high specificities compared to the BCSC-derived reference standard (13). A high specificity is crucial for a cancer detection algorithm. With a rare outcome such as breast cancer, even a slight decrement in specificity could markedly diminish the measure's PPV, as the number of false-positives may quickly accumulate in comparison to the number of true-positives. Nevertheless, sensitivity is also important to PPV. In this screening mammography cohort, the PPV of the CART approach (88.0%) exceeded that of the Nattinger approach (77.5%) due to a combination of slightly higher specificity and higher sensitivity (93.9% vs. 83.7%).

The PPV of 88.0% for the CART approach implies that approximately 1 of 9 cancers identified by the algorithm as incident detected breast cancers are wrongly identified as such. In the validation set, this included 87 cases (of a total 116,522 mammograms). As shown in Figure 1, most of these false-positives occurred among women with breast cancer diagnosis codes within 123 days and breast-directed surgery within one year of screening mammography. Using BCSC data, we found that ~36% (n=31) of these women had interval breast cancers that were not detected at screening mammography yet were diagnosed clinically within 365 days of screening. Additionally, we found that a small fraction of exams (<13%) had breast cancer diagnosis codes on claims but ultimately had benign biopsy results; providers for these women may have used breast cancer diagnosis codes while "ruling out" breast cancer. (Specific numbers suppressed to protect subject confidentiality.) Despite exclusion of mammograms for women with breast cancer diagnosis codes on claims during the three prior years, a small fraction of mammograms (<13%) were performed on women with prevalent breast cancers based on BCSC data and who received breast-directed surgery during the year following screening mammography, conceivably for local recurrences.

Meanwhile, the algorithm did not identify 7.1% of women whose breast cancers were truly detected at mammography (49 of 689 detected incident cancers in the validation set). As shown in Figure 1, many of these women may have had delays in the appearance of diagnostic codes for invasive breast cancer or ductal carcinoma in-situ or delays in receipt of either breast-directed surgery or subsequent mammography. Nattinger et al. similarly found that a small fraction of women with incident breast cancer according to SEER data did not

receive breast cancer surgery; these women were more likely than women who received surgery to lack breast cancer diagnostic codes in Medicare claims (13).

We recognize that cut-points in the CART algorithm may seem arbitrary, such as the initial node in which a breast cancer diagnosis code is assessed as occurring within 123 days of screening mammography. While there is no clinical justification for this cut-point, it yields the optimal classification of cases as assessed by the Gini index and validated in a testing sub-sample. Thus, a shorter or longer cut-point would compromise overall classification accuracy, most likely with degradation in PPV. At the same time, we believe the general structure and flow of the decision tree is clinically intuitive as one would expect diagnosis codes for breast cancer with subsequent procedure codes for treatment to be reliable predictors of incident breast cancer. The assessment for procedure codes for subsequent mammography among women with diagnosis codes for ductal carcinoma in-situ is perhaps less intuitive, although most women with newly diagnosed ductal carcinoma in-situ will indeed receive diagnostic mammography within 82 days of screening mammography.

Several potential uses of the CART algorithm are conceivable. First, the algorithm may be useful in claims-based studies of screening mammography outcomes, including studies of new screening technologies, such as digital mammography or computer-aided detection, on incident breast cancer detection. Because the algorithm does not require linkage with cancer registry data, investigators may apply it in unlinked Medicare claims, enabling sample sizes that would not be achievable with linked data such as the SEER-Medicare data. Second, the algorithm could conceivably enable claims-based estimation of radiologist- or facility-level quality metrics, specifically breast cancer detection rates, facilitating quality improvement efforts in mammography among Medicare enrollees. We recognize that the algorithm's PPV of ~88% does not confer certainty regarding a provider's true-positive interpretation. The algorithm may nevertheless allow identification of provider's with extremely low breast cancer detection rates. In any application, investigators must carefully consider potential impacts of algorithmic misclassification on study results and plan suitable sensitivity analyses.

Although BCSC data on whether breast cancers were detected at mammography may be imperfect, leading to inaccuracy in the reference standard, the BCSC mammography data undergo rigorous quality control and are linked with high-quality regional cancer registries. Our results also derive from mammography claims of fee-for-service Medicare enrollees within four U.S. regional mammography registries. Algorithms may not generalize to non-Medicare claims or to Medicare enrollees outside these regions. Because study algorithms require three years of prior claims to exclude prevalent breast cancers, their validity among women younger than age 68 years is uncertain. We also recognize that CART analyses may be prone to over-fitting. Nevertheless, our analysis included variables that are clinically meaningful, and we cross-validated the CART-derived algorithm within the training sample. We also pruned the tree based on theory and practice, and the final tree was highly predictive in the test sample. Finally, in late 2014, Medicare will transition from ICD-9 to ICD-10 coding. While the mapping of the algorithm's breast cancer diagnosis and procedure codes should be unambiguous, the performance of the algorithm would ideally be re-evaluated after the transition to ICD-10.

Study strengths include the inclusion of large mammography claim samples from geographically diverse settings that were linked with high-quality external mammography data, yielding rigorous validation analyses with excellent precision. Because cancer registries such as SEER encompass only 25% of the U.S. population (20), the alternative algorithms may enable mammogram sampling for research or quality improvement across the entire Medicare program regardless of claims linkage with cancer registry data.

We found that a simple, Medicare claims-based algorithm can identify with high predictive value women with incident breast cancers following a positive screening mammogram. Applied to Medicare claims alone without cancer registry linkage, the algorithm may be useful in claims-based studies of screening mammography. The potential for using the algorithm in provider-level quality assessment warrants evaluation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Funding sources

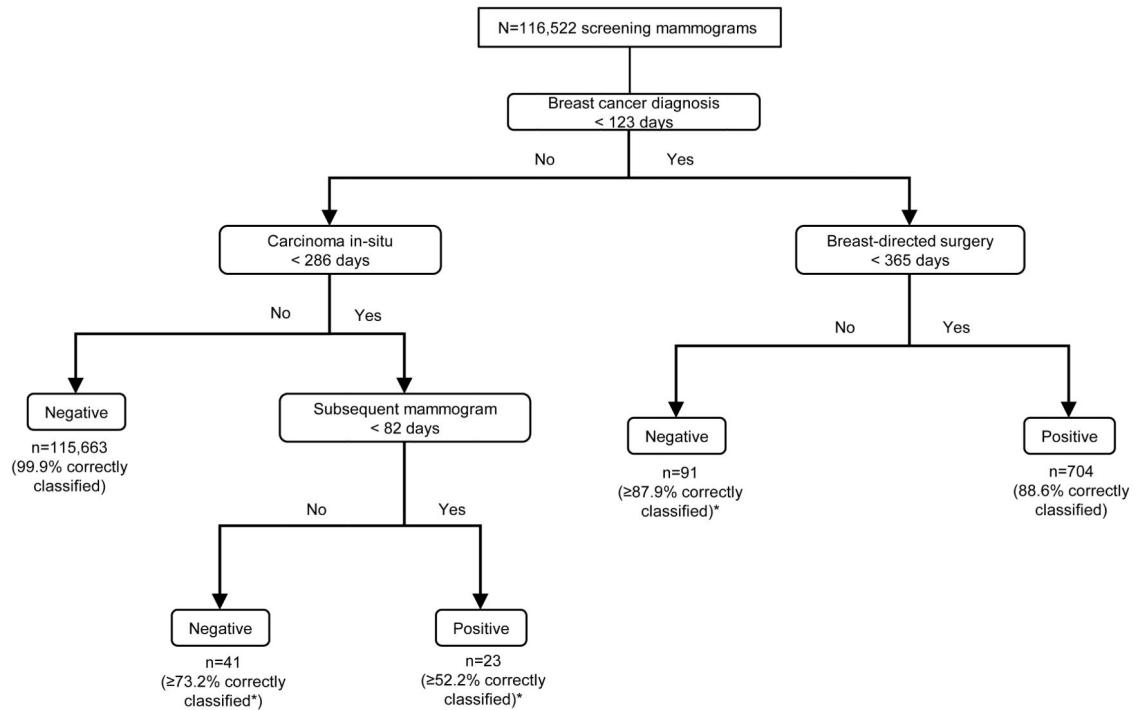
This work was supported by the National Cancer Institute-funded Breast Cancer Surveillance Consortium (U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040, HHSN261201100031C) and the National Cancer Institute-funded grant R21CA158510. The collection of cancer data used in this study was supported in part by several state public health departments and cancer registries throughout the U.S. For a full description of these sources, please see: <http://www.breastscreening.cancer.gov/work/acknowledgement.html>. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

We also thank the participating women, mammography facilities, and radiologists for the data they have provided for this study. A list of the BCSC investigators and procedures for requesting BCSC data for research purposes are provided at: <http://breastscreening.cancer.gov/>.

REFERENCES

1. Elmore JG, Jackson SL, Abraham L, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology*. 2009; 253:641–651. [PubMed: 19864507]
2. Smith-Bindman R, Chu PW, Miglioretti DL, et al. Comparison of screening mammography in the United States and the United Kingdom. *JAMA*. 2003; 290:2129–2137. [PubMed: 14570948]
3. Buist DS, Anderson ML, Haneuse SJ, et al. Influence of annual interpretive volume on screening mammography performance in the United States. *Radiology*. 2011; 259:72–84. [PubMed: 21343539]
4. Rosenberg RD, Yankaskas BC, Abraham LA, et al. Performance benchmarks for screening mammography. *Radiology*. 2006; 241:55–66. [PubMed: 16990671]
5. Carney PA, Sickles EA, Monsees BS, et al. Identifying minimally acceptable interpretive performance criteria for screening mammography. *Radiology*. 2010; 255:354–361. [PubMed: 20413750]
6. Nass, S.; Ball, J., editors. Improving breast imaging quality standards. National Academies Press; Washington, DC: 2005.
7. Gross CP, Long JB, Ross JS, et al. The cost of breast cancer screening in the Medicare population. *JAMA internal medicine*. 2013:1–7.

8. Ries, LAG.; Melbert, D.; Krapcho, M., et al., editors. SEER Cancer Statistics Review, 1975-2004. National Cancer Institute; Bethesda, MD: 2007.
9. Fenton JJ, Zhu W, Balch S, et al. Distinguishing screening from diagnostic mammograms using Medicare claims data. *Med Care*. 2012 In press.
10. Sickles EA, Miglioretti DL, Ballard-Barbash R, et al. Performance benchmarks for diagnostic mammography. *Radiology*. 2005; 235:775–790. [PubMed: 15914475]
11. 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. [PubMed: 12546290]
12. Freeman JL, Zhang D, Freeman DH, et al. An approach to identifying incident breast cancer cases using Medicare claims data. *J Clin Epidemiol*. 2000; 53:605–614. [PubMed: 10880779]
13. Nattinger AB, Laud PW, Bajorunaite R, et al. An algorithm for the use of Medicare claims data to identify women with incident breast cancer. *Health Serv Res*. 2004; 39:1733–1749. [PubMed: 15533184]
14. Smith-Bindman R, Quale C, Chu PW, et al. Can Medicare billing claims data be used to assess mammography utilization among women ages 65 and older? *Med Care*. 2006; 44:463–470. [PubMed: 16641665]
15. 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. [PubMed: 9308451]
16. American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS) Breast Imaging Atlas. American College of Radiology; Reston, VA: 2003.
17. Lemon SC, Roy J, Clark MA, et al. Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. *Ann Behav Med*. 2003; 26:172–181. [PubMed: 14644693]
18. Virnig BA, Warren JL, Cooper GS, et al. Studying radiation therapy using SEER-Medicare-linked data. *Med Care*. 2002; 40:IV-49–54.
19. Cooper GS, Virnig B, Klabunde CN, et al. Use of SEER-Medicare data for measuring cancer surgery. *Med Care*. 2002; 40:IV-43–48.
20. 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:IV-3–18.



*Cell sizes ≤ 11 are suppressed to protect patient confidentiality.

Figure 1. Algorithm for identifying incident breast cancers detected at screening mammography

Figure shows algorithmic allocation of 116,522 screening mammograms in the test mammogram set with the number allocated and percentage correctly classified in each terminal node. An algorithmic classification of “positive” signifies that the algorithm classified the mammogram as detecting an incident breast cancer, while a “negative” classification signifies that no breast cancer was detected at screening. Timing of all claims events are in relation to the date of screening mammography. To protect patient confidentiality, cell sizes of less than or equal to 11 are suppressed (and related numbers and percentages are given as a range).

Table 1
Characteristics of Medicare Enrollees Receiving Screening Mammograms

Characteristic	All (N=233,044 mammograms)	Breast Cancer Not Detected* (N=231,660)	Breast Cancer Detected* (N=1,384)
	n (%)	n (%)	n (%)
Age			
68-74	117,747 (50.5)	117,100 (50.4)	647 (52.4)
75-84	101,033 (43.4)	100,383 (43.5)	650 (42.1)
85+	14,264 (6.1)	14,177 (6.2)	87 (5.6)
Race/ethnicity			
White, non-Hispanic	187,254 (80.3)	186,117 (80.2)	1137 (82.3)
Black	16,967 (7.3)	16,856 (7.3)	111 (6.7)
Asian/Pacific Islander	6,812 (2.9)	6,780 (3.0)	32 (1.6)
Other/mixed/unknown [†]	22,011 (9.4)	21,907 (9.5)	104 (9.5)
Year of Mammogram			
2001	43,418 (18.6)	43,177 (18.7)	241 (17.4)
2002	48,234 (20.7)	47,934 (20.7)	300 (21.0)
2003	49,270 (21.1)	48,958 (21.1)	312 (21.9)
2004	46,003 (19.7)	45,734 (19.7)	269 (19.8)
2005	46,119 (19.8)	45,857 (19.8)	262 (19.9)
Medicaid eligibility			
No	20,2161 (86.8)	200,963 (86.6)	1,198 (88.5)
Yes	30,883 (13.2)	30,697 (13.4)	186 (11.5)
Non-urban vs. urban residence[‡]			
Non-urban	120,578 (51.7)	119,905 (51.6)	673 (53.9)
Urban	106,611 (45.8)	105,937 (45.9)	674 (43.7)
Unknown	5,855 (2.5)	5,818 (2.5)	37 (2.4)

* Incident breast cancers were identified using regional or statewide cancer registries and were considered detected if Breast Cancer Surveillance Consortium radiologists' Breast Imaging and Reporting Data System (BI-RADS) assessment was either 0, 3 with a recommendation for immediate further evaluation, 4, or 5.

[†] Other race/ethnicity includes American Indian/Alaska Native and Hispanic as well as other race/ethnicities.

[‡] Non-urban vs. urban residence defined based on Rural Urban Continuum Codes.

Table 2
Performance of Claims-based Algorithm for Identifying Incident Detected Breast Cancers following Screening Mammography in Validation Mammogram Set (N=116,522 mammograms)

Algorithmic Designation	Algorithm described by Nattinger, et al.(12)		Newly described CART algorithm	
	BCSC Reference Standard	BCSC Reference Standard	Incident Detected Breast Cancer	BCSC Reference Standard
Incident Detected Breast Cancer	577	168	640	87
No Incident Cancer Detected	112	115,665	49	115,746
Performance Measures, % (95% CI)				
Sensitivity	83.75 (80.99, 86.50)		92.89 (90.97, 94.81)	
Specificity	99.86 (99.83, 99.88)		99.93 (99.91, 99.94)	
PPV	77.45 (74.45, 80.45)		88.03 (85.67, 90.39)	
NPV	99.90 (99.89, 99.92)		99.96 (99.95, 99.97)	

Abbreviations: BCSC=Breast Cancer Surveillance Consortium; CART=Classification and regression tree; PPV=Positive Predictive Value; NPV=Negative Predictive Value.

Table 3
Unadjusted Breast Cancer Detection Rates and Relative Rates during Screening Mammography within BCSC Subpopulations using Algorithm versus BCSC Reference Standard

Characteristic	Breast Cancer Detection Rates (per 1000) (95% CI)		Relative Rates (95% CI)	
	Algorithm	BCSC Reference Standard	Algorithm	BCSC Reference Standard
Age, y				
68-74	5.9 (5.3, 6.5)	5.5 (4.9, 6.1)	1.0 (ref)	1.0 (ref)
75-84	6.7 (6.0, 7.4)	6.4 (5.7, 7.1)	1.14 (0.98, 1.32)	1.16 (0.99, 1.35)
85+	6.3 (4.5, 8.2)	6.2 (4.4, 8.1)	1.09 (0.80, 1.48)	1.14 (0.83, 1.56)
Race/ethnicity				
White, non-Hispanic	6.3 (5.8, 6.8)	6.0 (5.5, 6.5)	1.0 (ref)	1.0 (ref)
Black	7.0 (5.2, 8.8)	6.8 (5.0, 8.5)	1.11 (0.85, 1.45)	1.14 (0.87, 1.49)
Asian/Pacific Islander*	5.1 (3.1, 8.1)	5.1 (3.1, 8.1)	0.89 (0.56, 1.40)	0.94 (0.59, 1.48)
Other/mixed/unknown†	5.3 (4.0, 6.7)	5.0 (3.7, 6.3)	0.84 (0.65, 1.10)	0.83 (0.63, 1.10)
Medicaid eligibility				
No	6.3 (5.8, 6.8)	5.9 (5.5, 6.4)	1.0 (ref)	1.0 (ref)
Yes	6.0 (4.8, 7.2)	5.7 (4.5, 6.9)	0.95 (0.77, 1.18)	0.96 (0.77, 1.20)
Non-urban vs. urban residence ‡				
Non-urban	5.2 (4.6, 5.8)	5.1 (4.5, 5.6)	1.0 (ref)	1.0 (ref)
Urban	7.3 (6.6, 8.0)	6.8 (6.1, 7.5)	1.40 (1.21, 1.63)	1.33 (1.15, 1.55)

Abbreviations: BCSC=Breast Cancer Surveillance Consortium, CI=Confidence Interval

* Among Asian/Pacific Islanders, the algorithm identified the same breast cancers that were identified by the BCSC, yielding identical breast cancer detection rates in this subpopulation.

† Other race/ethnicity includes American Indian/Alaska Native and Hispanic as well as other race/ethnicities.

‡ Non-urban vs. urban residence defined based on Rural Urban Continuum Codes.