Title

Identification of Abnormal Screening Mammogram Interpretation Using Medicare Claims Data

Journal

Health Services Research, 50(1)

ISSN

0017-9124

Authors

Hubbard, Rebecca A
Zhu, Weiwei
Balch, Steven
et al.

Publication Date

2015-02-01

DOI

10.1111/1475-6773.12194

Peer reviewed

HSR Health Services Research

DOI: 10.1111/1475-6773.12194
RESEARCH BRIEF

Identification of Abnormal Screening Mammogram Interpretation Using Medicare Claims Data

Rebecca A. Hubbard, Weiwei Zhu, Steven Balch, Tracy Onega, and Joshua J. Fenton

Objective. To develop and validate Medicare claims-based approaches for identifying abnormal screening mammography interpretation.

Data Sources. Mammography data and linked Medicare claims for 387,709 mammograms performed from 1999 to 2005 within the Breast Cancer Surveillance Consortium (BCSC).

Study Design. Split-sample validation of algorithms based on claims for breast imaging or biopsy following screening mammography.

Data Extraction Methods. Medicare claims and BCSC mammography data were pooled at a central Statistical Coordinating Center.

Principal Findings. Presence of claims for subsequent imaging or biopsy had sensitivity of 74.9 percent (95 percent confidence interval [CI], 74.1–75.6) and specificity of 99.4 percent (95 percent CI, 99.4–99.5). A classification and regression tree improved sensitivity to 82.5 percent (95 percent CI, 81.9–83.2) but decreased specificity (96.6 percent, 95 percent CI, 96.6–96.8).

Conclusions. Medicare claims may be a feasible data source for research or quality improvement efforts addressing high rates of abnormal screening mammography.

Key Words. Breast cancer, mammography, Medicare, quality assessment, screening

The recall rate is a widely used measure of screening mammography performance. Abnormal screening mammography, in which a woman is recalled for diagnostic evaluation for possible breast cancer, occurs commonly, affecting about 1 in 10 screening mammograms in the United States (Rosenberg et al. 2006). Screening mammography recall rates in U.S. community practice are much higher than in other countries, although cancer detection rates are similar (Smith-Bindman et al. 2003), suggesting that recall rates in the United States could be lowered without compromising cancer detection.

Benchmarks for acceptable mammography recall rate have been established (Sickles, Wolverton, and Dee 2002; Rosenberg et al. 2006; Carney et al. 2010), and the federal Mammography Quality Standards Act (MQSA) requires mammography facilities to audit radiologist- and facility-level recall rates as a means of quality assurance, although these data are not publicly available (FDA 2014). The recall rate is included in the Medicare Hospital Outpatient Quality Reporting Program, which incents hospital outpatient departments to allow public reporting of claims-based quality metrics in exchange for potentially higher Medicare payments (Centers for Medicare and Medicaid Services 2014). However, to our knowledge, the claims-based recall measure used in this program has not been externally validated. With over 8.5 million women receiving mammograms paid for by Medicare annually (Lee et al. 2009), Medicare claims data might facilitate mammography quality assessment. If coupled with valid measures of breast cancer detection, a validated claims-based recall measure could provide an efficient means of direct public evaluation of this important dimension of mammography quality. However, no direct information on radiologist's interpretation is available in Medicare claims.

Although several studies have attempted to use Medicare claims to estimate measures of screening mammography performance (Welch and Fisher 1998; Tan et al. 2006), claims provide no direct information about abnormal interpretation. Identifying recalled mammograms using claims is challenging because claims only contain information on women who return for subsequent diagnostic evaluation. Estimating facility- or radiologist-specific recall rates is further complicated by small sample sizes, which may lead to unstable estimates and possible errors in the attribution of claims to individual radiologists or facilities.

In this study, we sought to validate the use of Medicare claims for identification of abnormal screening mammography using linked reference standard data from the Breast Cancer Surveillance Consortium (BCSC). By combining claims data with clinical information on mammography interpretation, we were able to estimate the accuracy of claims-based

---

Address correspondence to Rebecca A. Hubbard, Ph.D., Group Health Research Institute and Department of Biostatistics, University of Washington, 1730 Minor Ave, Suite 1600, Seattle, WA 98101; e-mail: hubbard.r@ghc.org. Weiwei Zhu, M.S., and Steven Balch, M.A., M.B.A., are with the Group Health Research Institute, Seattle, WA. Tracy Onega, Ph.D., is with the Department of Community and Family Medicine, Norris Cotton Cancer Center, Dartmouth Medical School, Lebanon, NH. Joshua J. Fenton, M.D., M.P.H., is with the Center for Healthcare Policy and Research, UC Davis Comprehensive Cancer Center, Sacramento, CA.

approaches to defining abnormal mammography results. We also compared radiologist recall rate estimates based on reference standard data to estimates based on claims alone. We hypothesized that Medicare claims codes for diagnostic evaluation procedures, breast-related symptoms, and breast cancer diagnoses subsequent to a screening mammogram would accurately predict abnormal mammography and enable accurate estimation of radiologist recall rates.

METHODS

Data

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/). BCSC facilities submit prospectively collected patient and mammography data to regional registries, which link the data to breast cancer outcomes ascertained from cancer registries. Mammography data include radiologist information on examination purpose (screening or diagnostic) and interpretation (normal or abnormal). BCSC sites have received institutional review board approval for 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 compliant, and BCSC sites have received a Federal Certificate of Confidentiality to protect the identities of patients, physicians, and facilities. Women in the BCSC database were matched to Medicare claims data using Social Security number, name, date of birth, and date of death. Among women aged 65 years and older and BCSC enrolled during the study period, over 87 percent were successfully matched to Medicare claims. 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.

SUBJECTS

We identified a sample of screening mammograms in Medicare claims and the BCSC with the same date of service among women who were aged 66 or older on mammography dates from January 1, 1999, to December 31, 2005. Screen-

ing mammograms were identified from Medicare claims data using a previously validated algorithm for distinguishing screening and diagnostic mammograms (Fenton et al. 2012). We chose to identify screening mammograms using Medicare claims codes (rather than BCSC data on the radiologist's indication) so that our data would resemble a sample of mammograms that would be available to researchers who only had access to Medicare claims.

From this matched sample, we selected screening mammograms for women with continuous enrollment in fee-for-service Medicare (parts A and B) for 12 months before and after mammography and with no enrollment in a Medicare HMO during this period. These enrollment criteria ensured complete capture of Medicare claims. Premammogram enrollment was required to facilitate correct classification of screening mammography purpose. Postmammogram enrollment was required for identification of claims codes indicative of an abnormal mammography result. We randomly divided the matched sample into two half-samples, one for algorithm training and one for validation.

Definitions

We defined the reference standard for mammography interpretation using BCSC data on radiologists' assessments and recommendations. A screening mammogram was considered abnormal if the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS®) (American College of Radiology 2003) assessment was 0 (needs additional imaging evaluation); 4 (suspicious abnormality); 5 (highly suggestive of malignancy); or 3 (probably benign finding) with a recommendation for immediate follow-up.

We used Healthcare Common Procedure Coding System (HCPCS) codes, International Classification of Disease (ICD)-9 codes, and diagnosis-related groups (DRG) from Medicare claims to identify service date for procedures potentially suggestive of diagnostic evaluation following an abnormal screening mammogram (see Appendix SA2). These included diagnostic breast imaging tests, breast biopsies, breast symptoms, and breast cancer diagnoses.

Patient age and race were defined based on a self-administered survey collected by BCSC facilities at the time of mammography. In statistical analyses, we used age- and race-stratified results to evaluate whether performance of the algorithm was consistent across patient subgroups. Although age and race information was also available in Medicare data, we selected to use information from the BCSC patient questionnaire

because these data are based on self-report and have been carefully validated. Previous research has demonstrated that Medicare enrollment files on race/ethnicity tend to underidentify members of racial/ethnic minority groups (Zaslavsky, Ayanian, and Zaborski 2012). We therefore chose to use BCSC data to define these patient characteristics. Information on facility characteristics, including for-profit status, academic medical center affiliation, and facility type, were obtained from a survey completed by the BCSC registries. Facility type was classified as hospital, specialty center, or nonspecialty center. Specialty centers included radiology private offices and hospital outpatient centers. Nonspecialty centers included obstetrics and gynecology offices, primary care offices, and multispecialty clinics.

Statistical Analysis

We computed descriptive statistics to characterize the study sample stratified by BCSC classification of the mammography result as normal or abnormal. We report the number and percentage of screening mammograms followed by a claims procedure code for diagnostic imaging or biopsy or a diagnosis code for breast-related symptoms, abnormal mammography, or breast cancer diagnoses within 365 days. Among those with a subsequent claims code, we estimated the median and interquartile range (IQR) for the number of days from the screening mammogram to the code.

We first classified mammograms as abnormal if any claims codes for diagnostic mammography, breast ultrasound, breast MRI, or breast biopsy/fine needle aspiration (FNA) were found in the 90 days subsequent to the screening mammogram. Because this is an ad hoc method, we investigated whether a formal, statistical prediction model could improve classification accuracy. We developed a classification algorithm using classification and regression tree (CART) analysis because this method can identify important interactions in predictor variables without requiring prespecification, does not require assumptions about the functional form of the relationship between continuous predictors and the outcome, and returns an algorithm that is simple to communicate and implement in external data sets (Breiman et al. 1984). The outcome variable in our analysis was the BCSC classification of mammography result (normal/abnormal) based on the radiologist's assessment and recommendations. Potential predictors were number of days from the screening mammogram to each of the categories of claims codes described above.

The CART was developed using the training half-sample. The CART selected splits in predictor variables on the basis of the Gini index, and continued splitting until no further splits improved the Gini index by more than 0.00001. The Gini index is a measure of impurity and was defined as the product of the proportion of normal mammograms and the proportion of abnormal mammograms in a terminal node (Breiman et al. 1984). To develop a high sensitivity classification rule, we penalized misclassification of abnormal mammograms as normal approximately 40 times more highly than misclassification of normal mammograms as abnormal. We selected this degree of penalization because in our training sample this was the highest penalty that produced improved sensitivity without unduly compromising specificity. To minimize overfitting, our final algorithm only included the first five splits of the pruned tree.

We quantified the operating characteristics of both classification approaches using the validation half-sample. Performance characteristics included the following: sensitivity (the proportion of abnormal mammograms classified as abnormal); specificity (the proportion of normal mammograms classified as normal); positive predictive value (PPV) (the proportion of mammograms classified as abnormal that were abnormal according to radiologists' assessments and recommendations); and negative predictive value (NPV) (the proportion of mammograms classified as normal that were normal based on radiologists' assessments and recommendations). We computed performance measures overall and stratified by age, race, facility profit status, and radiology practice type. For mammograms misclassified by the claims-based approaches, we investigated the proportion with a subsequent cancer diagnosis within 1 year and the BI-RADS assessment assigned by the radiologist to better understand possible causes of misclassification.

We used data from the validation half-sample to estimate radiologist recall rates based on BCSC data and using each of the two Medicare claims-based approaches. This analysis was restricted to radiologists with at least 100 mammograms included in the validation sample. Medicare claims-based approaches were compared to BCSC estimates using the Spearman correlation coefficient.

We performed statistical analyses using R, version 2.15.1 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

The sample included 387,709 screening mammograms received by 150,395 women. These mammograms were interpreted by 551 radiologists, practicing at 146 facilities. The majority were performed on nonHispanic white women (79.4 percent) who were 66–74 years of age (58.3 percent) (Table 1). Overall, 7.2 percent of mammograms were interpreted as abnormal. Of the categories of claims codes considered, diagnostic mammography, breast ultrasound, and diagnosis codes for breast-related signs or symptoms were the most commonly occurring during the 1-year period after the screening mammogram (see Appendix SA3).

Claims codes for breast imaging or biopsy/FNA within 90 days subsequent to screening mammography correctly classified abnormal mammograms with 74.9 percent sensitivity and 99.4 percent specificity in the validation half-sample (Table 2). A CART that included claims codes for diagnostic mammography, breast signs and symptoms, breast ultrasound, breast cancer, and abnormal mammography had sensitivity of 82.5 percent and specificity of 96.6 percent in the validation half-sample (Figure 1).

Among mammograms misclassified as abnormal by the presence of claims within 90 days, 5.2 percent were followed by a cancer diagnosis within 1 year. Analogously, 3.3 percent of mammograms misclassified as abnormal by the CART were followed by a cancer diagnosis within 1 year. In both the claims within 90 days and CART approaches, 18 percent of normal mammograms misclassified as abnormal had a BI-RADS assessment of 3 with no recommendation for immediate evaluation. Meanwhile, among abnormal mammograms misclassified as normal, in both the claims within 90 days and CART approaches, <1 percent were followed by a cancer diagnosis within 1 year, and the vast majority (94 percent) had a BI-RADS assessment of 0.

Sensitivity of the claims-based algorithms varied across patient and facility characteristics. Sensitivity was lower for women older than 85 years compared to younger women (Appendix SA4). Sensitivity was also lower for black women and American Indian/Alaskan Native women compared to nonHispanic white women. Sensitivity was similar for mammograms from hospitals and specialty centers and somewhat lower for mammograms from nonspecialty centers. Specificity was similarly high across patient and facility subgroups.

Recall rate estimates for 325 radiologists with at least 100 mammograms included in the validation half-sample agreed well when estimated using BCSC data or Medicare claims (Appendix SA5). On average, recall rates

Table 1: Patient and Screening Mammogram Characteristics Stratified by Mammography Result Based on Radiologist's Assessment and Recommendations Captured by the BCSC

Characteristic	*All (N = 387,709) N (%)*	*Normal (N = 359,922) N (%)*	*Abnormal (N = 27,787) N (%)*
Age			
66–74	226,167 (58.3)	209,398 (58.2)	16,769 (60.4)
75–84	141,740 (36.6)	132,006 (36.7)	9,734 (35.0)
85+	19,802 (5.1)	18,518 (5.2)	1,284 (4.6)
Race			
White, non-Hispanic	307,783 (79.4)	285,220 (79.2)	22,563 (81.2)
Black, non-Hispanic	28,390 (7.3)	26,522 (7.4)	1,868 (6.7)
Asian/Pacific Islander	12,014 (3.1)	11,518 (3.2)	496 (1.8)
American Indian/Alaskan Native	1,591 (0.4)	1,488 (0.4)	103 (0.4)
Hispanic	5,675 (1.5)	5,308 (1.5)	367 (1.3)
Other/mixed/unknown	32,256 (8.3)	29,866 (8.3)	2,390 (8.6)
Year of mammogram			
1999	49,974 (12.9)	46,691 (13.0)	3,283 (11.8)
2000	52,099 (13.4)	48,597 (13.5)	3,502 (12.6)
2001	53,358 (13.8)	49,575 (13.8)	3,783 (13.6)
2002	59,346 (15.3)	54,960 (15.3)	4,386 (15.8)
2003	60,660 (15.7)	56,137 (15.6)	4,523 (16.3)
2004	56,704 (14.6)	52,559 (14.6)	4,145 (14.9)
2005	55,568 (14.3)	51,403 (14.3)	4,165 (15.0)
BI-RADS assessment of screening mammogram			
1	253,839 (65.5)	253,839 (70.5)	–
2	102,706 (26.5)	102,706 (28.5)	–
3	4,014 (1.0)	3,377 (0.9)	637 (2.3)
0	257,22 (6.6)	–	25,722 (92.6)
4	1,261 (0.3)	–	1,261 (4.5)
5	167 (0.0)	–	167 (0.6)
Facility profit status			
Not for profit	98,675 (25.4)	91,722 (25.5)	6,953 (25.0)
For profit	218,208 (56.3)	202,023 (56.1)	16,185 (58.2)
Unknown	70,826 (18.3)	66,177 (18.4)	4,649 (16.7)
Facility associated with academic medical center			
No	326,829 (84.3)	304,305 (84.5)	22,524 (81.1)
Yes	34,697 (8.9)	31,774 (8.8)	2,923 (10.5)
Unknown	26,183 (6.8)	23,843 (6.6)	2,340 (8.4)

continued

Table 1. *Continued*

Characteristic	*All (N = 387,709) N (%)*	*Normal (N = 359,922) N (%)*	*Abnormal (N = 27,787) N (%)*
Facility type			
Hospital	184,435 (47.6)	171,645 (47.7)	12,790 (46.0)
Specialty center	136,178 (35.1)	126,208 (35.1)	9,970 (35.9)
Nonspecialty center	35,652 (9.2)	33,055 (9.2)	2,597 (9.3)
Other/unknown	31,444 (8.1)	29,014 (8.1)	2,430 (8.7)

BCSC, Breast Cancer Surveillance Consortium; BI-RADS, Breast Imaging Reporting and Data System.

Table 2: Performance of Classification Based on Claims Codes for Breast Imaging or Biopsy/FNA within Ninety Days or Classification and Regression Tree Analysis for Classifying Screening Mammograms as Normal versus Abnormal Assessed in Validation Half-Sample ($N = 193{,}855$)

Performance Measures	*Claims Codes for Breast Imaging or Biopsy Estimate (95% CI)*	*CART Estimate (95% CI)*
Sensitivity	74.9 (74.1, 75.6)	82.5 (81.9, 83.2)
Specificity	99.4 (99.4, 99.5)	96.6 (96.6, 96.7)
PPV	90.9 (90.4, 91.5)	65.3 (64.6, 66.0)
NPV	98.1 (98.0, 98.2)	98.6 (98.6, 98.7)

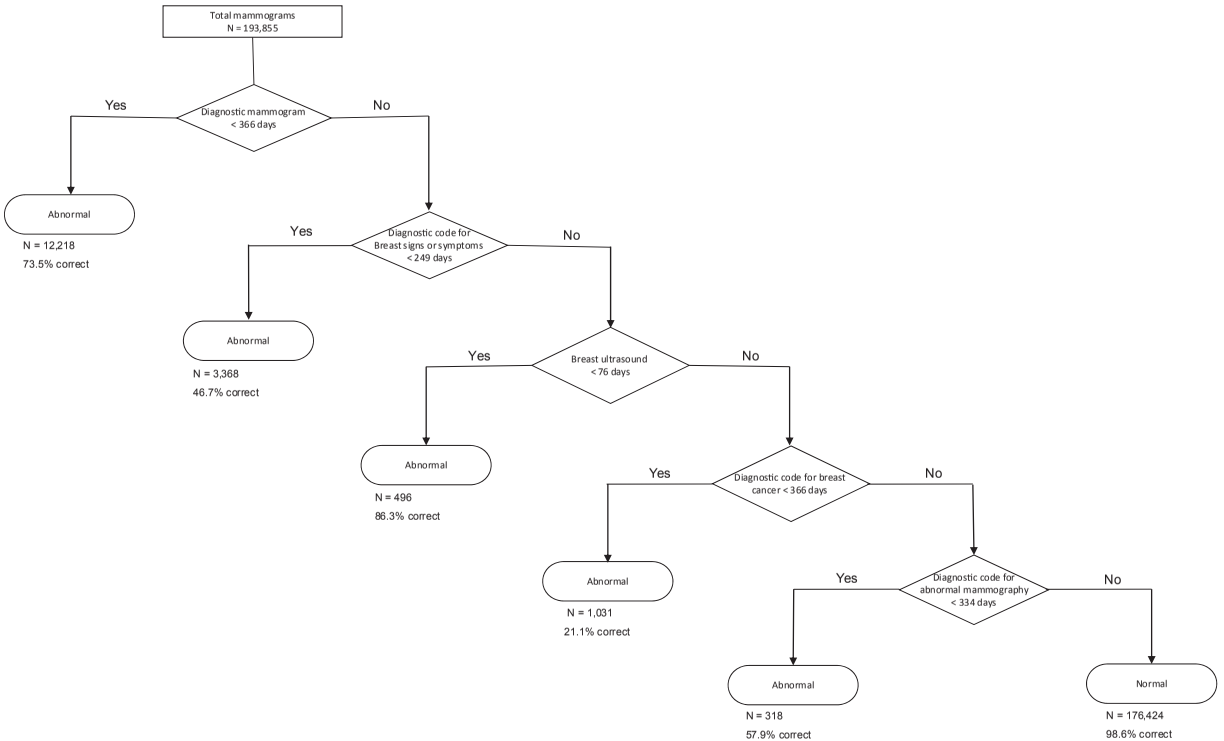
CART, classification and regression tree; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

were slightly underestimated by the approach that identified abnormal mammograms based on any breast-related claims within 90 days and slightly overestimated by the CART. Correlation of BCSC estimates with the algorithm based on any claims within 90 days was 0.84 (95 percent confidence interval [CI]: 0.78, 0.89) and with the CART was 0.82 (95 percent CI: 0.76, 0.87).

DISCUSSION

In this validation study, we evaluated two claims-based approaches for identifying abnormal screening mammograms. An approach based on the presence of claims for diagnostic breast imaging or biopsy within 90 days of a screening mammogram achieved good classification accuracy. Overall, 75 percent of abnormal mammograms and 99 percent of normal mammograms were

Figure 1: Classification and Regression Tree for Identifying Abnormal Screening Mammography Interpretation Based on Claims Subsequent to Screening Mammography



Numbers under Terminal Nodes are Total Number of Mammograms Classified in the Node and Percent Correctly Classified within the Validation Half-Sample.

accurately classified by this simple approach. A CART increased sensitivity to 83 percent but at the cost of decreased specificity. Using either approach to estimate radiologist recall rates displayed good correlation with recall rate estimates based on clinical data.

We found that normal mammograms falsely classified as abnormal were more likely to be followed by a cancer diagnosis within the next year than correctly classified normal mammograms. They were also more likely to have had an initial BI-RADS assessment of 3 compared to other normal mammograms. This suggests that algorithmic misclassification of mammograms as abnormal may be due to women returning for short interval follow-up examinations or diagnostic work-up for a missed cancer that had become symptomatic. Abnormal mammograms misclassified as normal may have occurred because women did not receive recommended diagnostic evaluation within a timely interval following the screening mammogram.

We observed some differences in sensitivity of claims-based approaches by patient age and race. These may reflect differences in the proportion of women within age or racial subgroups receiving timely follow-up care. Timely receipt of follow-up previously has been shown to vary according to patient characteristics (Elmore et al. 2005). We also observed poorer sensitivity for mammograms performed in nonspecialty centers. However, these facilities comprised less than 10 percent of the study sample. Observed differences between facility types may be due to differences in the characteristics of patients served by these facilities. Risk-adjustment should therefore be used to account for difference in case mix variation across providers when estimating recall rates.

Our algorithms may be useful for several purposes. First, they could be used to estimate facility or radiologist recall rates for purposes of quality assessment. In our comparison of claims-based and BCSC estimates of recall rate we found that claims-based measures correlated well with those based on clinical data. However, the approach based on any claims within 90 days tended to underestimate recall rates, while the CART overestimated recall rates. Since Medicare's current measure of outpatient imaging efficiency uses a shorter 45-day window following screening and only includes subsequent claims for mammography, breast ultrasound, or breast MRI, it would underestimate recall rates to an even greater degree. Assessment of relative performance based on our claims-based recall measures may be reasonable, but comparisons to fixed benchmarks would be inappropriate due to the under- and overestimation of absolute recall rates noted above. Nevertheless, our validation study provides general support for Medicare's current claims-based

approach to measuring provider-level recall rates within hospital outpatient departments and suggests that Medicare claims could be used to expand this measure to community-based mammography facilities.

An alternative use of these algorithms is for evaluation of mammography performance in the Medicare population broadly, across regions, or between patient subgroups. Previous studies of abnormal mammography using Medicare claims have classified mammograms as abnormal based on the presence of codes for diagnostic mammography, breast ultrasound, and breast biopsy or FNA (Welch and Fisher 1998; Tan et al. 2006). Our validation study found that this approach has good specificity (99 percent) but fails to identify about one in four abnormal mammograms. Previous research has demonstrated that a claims-based approach with good specificity but imperfect sensitivity can still enable estimation of unbiased *relative* recall rates between regions or providers (Chubak et al. 2012).

Strengths of this study include the large sample size. Our data comprised mammography information from over 140 mammography facilities and over 500 radiologists in a geographically diverse population. Our study also draws on the strength of a unique linkage of Medicare claims to detailed, high-quality mammography data, including information on the indication for examination and radiologists' assessments and recommendations. Limitations of the study include use of Medicare claims data collected from 2000 to 2005. Performance may not be equivalent in other claims databases or years. However, we believe performance is likely to be similar in more recent Medicare claims data because no notable changes in practice patterns for follow-up of abnormal mammography have occurred between 2005 and the present.

We have validated two approaches to using Medicare claims data to identify abnormal screening mammography interpretation. This research may facilitate use of Medicare claims for estimating provider-level recall rates. However, in evaluating screening mammography accuracy, recall rates would ideally be considered in conjunction with breast cancer detection rates (Fenton et al. 2013). If claims-based metrics of breast cancer detection can be validated at the provider-level, the recall rate algorithms described herein may facilitate robust research and quality improvement efforts in screening mammography.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Dr. Hubbard, Ms. Zhu, and Mr. Balch were previously supported in part by a grant from GE Healthcare. Because

GE Healthcare is a manufacturer of mammography machines and other breast imaging devices, some readers may see this support as relevant to the current manuscript. However, this grant did not support the current research and in no way influenced any of the findings included in this manuscript. This work was supported by the National Cancer Institute–funded grant R21CA158510 and the National Cancer Institute–funded Breast Cancer Surveillance Consortium (HHSN261201100031C). The collection of cancer data used in this study was supported in part by several state public health departments and cancer registries throughout the United States. 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.

Disclosures: None.

Disclaimers: None.

References

American College of Radiology. 2003. *Breast Imaging Reporting and Data System (BI-RADS) Breast Imaging Atlas*. Reston, VA: American College of Radiology.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*. New York: Chapman & Hall.

Carney, P. A., E. A. Sickles, B. S. Monsees, L. W. Bassett, R. J. Brenner, S. A. Feig, R. A. Smith, R. D. Rosenberg, T. A. Bogart, S. Browning, J. W. Barry, M. M. Kelly, K. A. Tran, and D. L. Miglioretti. 2010. "Identifying Minimally Acceptable Interpretive Performance Criteria for Screening Mammography." *Radiology* 255 (2): 354–361.

Centers for Medicare and Medicaid Services. 2014. "Hospital Outpatient Quality Reporting (OQR) Program Overview" [accessed on February 13, 2014]. Available at http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1191255879384

Chubak, J., O. Yu, G. Pocobelli, L. Lamerato, J. Webster, M. N. Prout, M. Ulcickas Yood, W. E. Barlow, and D. S. Buist. 2012. "Administrative Data Algorithms to Identify Second Breast Cancer Events Following Early-Stage Invasive Breast Cancer." *Journal of the National Cancer Institute* 104 (12): 931–940.

Elmore, J. G., C. Y. Nakano, H. M. Linden, L. M. Reisch, J. Z. Ayanian, and E. B. Larson. 2005. "Racial Inequities in the Timing of Breast Cancer Detection, Diagnosis, and Initiation of Treatment." *Medical Care* 43 (2): 141–148.

FDA. 2014. "Mammography Quality Standards Act" [accessed on February 13, 2014]. Available at http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/Regulations/ucm110906.htm

Fenton, J. J., W. Zhu, S. Balch, R. Smith-Bindman, P. Fishman, and R. A. Hubbard. 2012. "Distinguishing Screening from Diagnostic Mammograms Using Medicare Claims Data." *Medical Care.* [Epub ahead of print] doi: 10.1097/MLR.0b013e318269e0f5

Fenton, J. J., T. Onega, W. Zhu, S. Balch, R. Smith-Bindman, L. Henderson, B. L. Sprague, K. Kerlikowske, and R. A. Hubbard. 2013. "Validation of a Medicare Claims-Based Algorithm for Identifying Breast Cancers Detected at Screening Mammography." *Medical Care.* [Epub ahead of print] doi: 10.1097/MLR.0b013e3182a303d7

Lee, D. W., P. E. Stang, G. A. Goldberg, and M. Haberman. 2009. "Resource Use and Cost of Diagnostic Workup of Women with Suspected Breast Cancer." *Breast Journal* 15 (1): 85–92.

Rosenberg, R. D., B. C. Yankaskas, L. A. Abraham, E. A. Sickles, C. D. Lehman, B. M. Geller, P. A. Carney, K. Kerlikowske, D. S. Buist, D. L. Weaver, W. E. Barlow, and R. Ballard-Barbash. 2006. "Performance Benchmarks for Screening Mammography." *Radiology* 241 (1): 55–66.

Sickles, E. A., D. E. Wolverton, and K. E. Dee. 2002. "Performance Parameters for Screening and Diagnostic Mammography: Specialist and General Radiologists." *Radiology* 224 (3): 861–9.

Smith-Bindman, R., P. W. Chu, D. L. Miglioretti, E. A. Sickles, R. Blanks, R. Ballard-Barbash, J. K. Bobo, N. C. Lee, M. G. Wallis, J. Patnick, and K. Kerlikowske. 2003. "Comparison of Screening Mammography in the United States and the United Kingdom." *Journal of the American Medical Association* 290 (16): 2129–2137.

Tan, A., D. H. Freeman Jr, J. S. Goodwin, and J. L. Freeman. 2006. "Variation in False-Positive Rates of Mammography Reading among 1067 Radiologists: A Population-Based Assessment." *Breast Cancer Research and Treat* 100 (3): 309–318.

Welch, H. G., and E. S. Fisher. 1998. "Diagnostic Testing Following Screening Mammography in the Elderly." *Journal of the National Cancer Institute* 90 (18): 1389–1392.

Zaslavsky, A. M., J. Z. Ayanian, and L. B. Zaborski. 2012. "The Validity of Race and Ethnicity in Enrollment Data for Medicare Beneficiaries." *Health Services Research* 47 (3 Pt 2): 1300–1321.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix SA2: Claims Codes for Procedures and Diagnoses Assessed as Predictors of Abnormal Mammography.

Appendix SA3: Frequency and Timing of Breast-Related Medicare

Claims after Screening Mammography by Normal versus Abnormal Reference Standard Classification.

Appendix SA4: Sensitivity and Specificity of Claims-Based Measures of Abnormal Mammography Interpretation Stratified by Patient Age and Racial/Ethnic Sub Groups and Facility Characteristics Evaluated in the Validation Half-Sample.

Appendix SA5: Plot of Radiologist Recall Rates Estimated Based on the Presence of Any Medicare Claims for Breast-Related Procedures in Ninety Days after the Mammogram (Left) or the Medicare Claims-Based CART (Right) versus Recall Rates Estimated Using Clinical Data from the Breast Cancer Surveillance Consortium (BCSC) on Abnormal Mammography. Point Estimates and 95 percent Confidence Intervals Are Plotted for Each Radiologist.