

# Mammographic Interpretive Volume and Diagnostic Mammogram Interpretation Performance in Community Practice<sup>1</sup>

Sebastien Haneuse, PhD  
Diana S. M. Buist, PhD  
Diana L. Miglioretti, PhD  
Melissa L. Anderson, MS  
Patricia A. Carney, PhD  
Tracy Onega, PhD  
Berta M. Geller, EdD  
Karla Kerlikowske, MD  
Robert D. Rosenberg, MD  
Bonnie C. Yankaskas, PhD  
Joann G. Elmore, MD, MPH  
Stephen H. Taplin, MD, MPH  
Robert A. Smith, PhD  
Edward A. Sickles, MD

<sup>1</sup>From the Department of Biostatistics, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115 (S.H.); Group Health Research Institute, Seattle, Wash (D.S.M.B., D.L.M., M.L.A.); Department of Biostatistics, University of Washington, Seattle, Wash (D.L.M.); Departments of Family Medicine and Public Health and Preventive Medicine, Oregon Health & Science University, Portland, Ore (P.A.C.); Department of Community and Family Medicine, Dartmouth Medical School, Norris Cotton Cancer Center, Lebanon, NH (T.O.); Departments of Family Medicine and Radiology, University of Vermont, College of Medicine, Burlington, Vt (B.M.G.); Department of Medicine (K.K.), Department of Epidemiology and Biostatistics (K.K.), General Internal Medicine Section, Department of Veterans Affairs (K.K.), and Department of Radiology (E.A.S.), University of California, San Francisco, Calif; Department of Radiology, University of New Mexico, Health Sciences Center, Albuquerque, NM (R.D.R.); Department of Radiology, University of North Carolina, Chapel Hill, NC (B.C.Y.); Department of Medicine, University of Washington School of Medicine, Harborview Medical Center, Seattle, Wash (J.G.E.); Division of Cancer Control and Population Science, Applied Research Program, National Cancer Institute, Rockville, Md (S.H.T.); and Cancer Control Science Department, American Cancer Society, Atlanta, Ga (R.A.S.). Received June 7, 2011; revision requested July 24; revision received July 27; accepted August 3; final version accepted August 15. Supported by the American Cancer Society, the Longaberger Company's Horizon of Hope Campaign (SIRGS-07-271-01, SIRGS-07-272-01, SIRGS-07-274-01, SIRGS-07-275-01, SIRGS-06-281-01, ACS A1-07-362), the Agency for Healthcare Research and Quality (HS-10591), and the Breast Cancer Stamp Fund. **Address correspondence to S.H.** (e-mail: [shaneuse@hsph.harvard.edu](mailto:shaneuse@hsph.harvard.edu)).

© RSNA, 2011

## Purpose:

To investigate the association between radiologist interpretive volume and diagnostic mammography performance in community-based settings.

## Materials and Methods:

This study received institutional review board approval and was HIPAA compliant. A total of 117136 diagnostic mammograms that were interpreted by 107 radiologists between 2002 and 2006 in the Breast Cancer Surveillance Consortium were included. Logistic regression analysis was used to estimate the adjusted effect on sensitivity and the rates of false-positive findings and cancer detection of four volume measures: annual diagnostic volume, screening volume, total volume, and diagnostic focus (percentage of total volume that is diagnostic). Analyses were stratified by the indication for imaging: additional imaging after screening mammography or evaluation of a breast concern or problem.

## Results:

Diagnostic volume was associated with sensitivity; the odds of a true-positive finding rose until a diagnostic volume of 1000 mammograms was reached; thereafter, they either leveled off ( $P < .001$  for additional imaging) or decreased ( $P = .049$  for breast concerns or problems) with further volume increases. Diagnostic focus was associated with false-positive rate; the odds of a false-positive finding increased until a diagnostic focus of 20% was reached and decreased thereafter ( $P < .024$  for additional imaging and  $P < .001$  for breast concerns or problems with no self-reported lump). Neither total volume nor screening volume was consistently associated with diagnostic performance.

## Conclusion:

Interpretive volume and diagnostic performance have complex multifaceted relationships. Our results suggest that diagnostic interpretive volume is a key determinant in the development of thresholds for considering a diagnostic mammogram to be abnormal. Current volume regulations do not distinguish between screening and diagnostic mammography, and doing so would likely be challenging.

© RSNA, 2011

Supplemental material: <http://radiology.rsna.org/lookup/suppl/doi:10.1148/radiol.11111026/-/DC1>

**D**iagnostic mammography is a critical tool in the evaluation of abnormal screening mammograms and in the examination of women who have signs or symptoms of breast cancer. Unlike screening mammography, diagnostic mammography usually involves acquisition of additional mammographic images, is typically performed to evaluate a specific area of concern, and frequently involves a population in which cancer is approximately 10 times more prevalent than in the general population (1).

The technical quality of mammography has improved since the 1992 Mammography Quality Standards Act (2); however, relative to screening mammography, little research has been devoted

to understanding variability in the interpretive performance of diagnostic mammography in community-based settings. Despite studies in which researchers considered characteristics at the level of the woman and mammogram (1,3–8), radiologist (9,10), and facility (11–13), broad unexplained variation remains in the performance of diagnostic mammography (1,9).

One possible source of variation in performance is the interpreting radiologist's volume of work; an Institute of Medicine report called for more research in this area (2). To our knowledge, in only two studies have researchers examined interpretive volume as it relates to diagnostic mammography. Jensen et al (13) noted improved diagnostic performance in Danish clinics with at least one radiologist with a high volume of work. However, they did not examine radiologist-specific volume measures. Miglioretti et al (9) found no association between work volume and diagnostic performance, but the evaluation was limited by the use of a self-reported three-level categorical measure of total volume. Given these limitations, we undertook this study to investigate the association between radiologist interpretive volume and diagnostic mammography performance in community-based settings.

### Materials and Methods

The collection of cancer incidence data used in this study was supported in

### Implication for Patient Care

- Our results suggest that diagnostic interpretive volume is a key determinant in developing thresholds for considering a diagnostic mammogram to be abnormal; however, because of the complex and multifaceted nature of interpretive volume and diagnostic performance, it is difficult to characterize their relationship, which limits the practicality of setting regulatory thresholds for diagnostic mammographic interpretation.


part by several state public health departments and cancer registries throughout the United States. For a full description of these sources, please see <http://breastscreening.cancer.gov/work/acknowledgement.html>. The authors had full responsibility for designing the study; collecting, analyzing, and interpreting the data; writing the manuscript; and deciding to submit it for publication.

The Breast Cancer Surveillance Consortium (BCSC) is a population-based collaborative network of mammography registries (14,15). Each registry collects demographic and clinical data at each mammography visit at participating facilities. Cancer outcomes are ascertained via linkage to a regional or Surveillance, Epidemiology, and End Results cancer registry and to pathology databases. Analyses presented here are based on pooled data from six registries (San Francisco, Calif; North Carolina; New Hampshire; Vermont; western Washington; and New Mexico). Sites received institutional review board approval for study activities and for protection of the identities of women,

### Advances in Knowledge

- Neither annual total volume nor screening volume was consistently associated with diagnostic performance.
- Annual diagnostic volume exhibited a strong association with sensitivity of diagnostic mammography performed for additional imaging or for evaluation of breast concerns or problems ( $P < .001$  and  $P = .049$ , respectively).
- For diagnostic mammography performed to evaluate breast concerns or problems, significant interactions between diagnostic focus and the presence of a self-reported lump were found for each performance measure ( $P = .020$ ,  $P < .001$ , and  $P = .027$  for sensitivity, false-positive rate, and cancer detection rate, respectively).
- False-positive rate exhibited an increasing association with diagnostic focus up to a focus of 20%; beyond this apex, false-positive rate decreased with increasing diagnostic focus.
- No consistent association was found between any volume measure and cancer detection rate.

### Published online before print

10.1148/radiol.11111026 **Content code:** 

**Radiology 2012;** 262:69–79

### Abbreviations:

BCSC = Breast Cancer Surveillance Consortium

CDR = cancer detection rate

CI = confidence interval

FPR = false-positive rate

### Author contributions:

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

### Funding:

This research was supported by the National Cancer Institute (grants 1R01 CA107623, 1K05 CA104699, U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, and U01CA70040).

Potential conflicts of interest are listed at the end of this article

physicians, and facilities. Radiologists provided informed consent. Active consent, passive permission, and/or waivers were obtained at each BCSC site from women who underwent mammography. Procedures were compliant with the Health Insurance Portability and Accountability Act (16).

### Interpretive Volume

Between 2005 and 2006, 214 BCSC radiologists who interpreted mammograms at a BCSC facility completed a self-administered mailed survey (17). The survey asked radiologists to indicate all facilities where they had interpreted mammograms between 2001 and 2005. For each radiologist who interpreted mammograms at non-BCSC facilities, registry staff collected additional volume information at the non-BCSC facilities. Only radiologists with complete volume information from all facilities were included. Our final study sample included 107 radiologists; demographic characteristics and experience of these radiologists resembled those of the original survey respondents (Table E1 [online]).

We measured diagnostic volume, screening volume, and total volume when the radiologist served as the primary reader from each facility for each year between 2001 and 2005. Diagnostic volume included examinations performed for additional evaluation of a prior mammogram, short-interval follow-up, or evaluation of a breast concern or problem. Total volume included all diagnostic and screening examinations; diagnostic and screening mammograms obtained in the same woman and interpreted by the same radiologist on the same day counted as one study (18). This differs from Breast Imaging Reporting and Data System audits (19), where screening and diagnostic mammography performed on the same day independently contribute to total volume. Mammograms with missing indication (2.5%) were attributed to diagnostic or screening volume based on the observed proportion for that reader (18).

For each radiologist and year, each volume measure was summed across all facilities to obtain annual totals;

diagnostic focus is the percentage of total mammograms that were considered diagnostic.

### Interpretive Performance

We included diagnostic mammograms that had indications for additional imaging of an abnormality that was detected at screening or that were obtained to evaluate a breast concern or problem (1). Mammograms obtained in women with a history of breast cancer were excluded.

Examination results were considered positive or negative on the basis of the final Breast Imaging Reporting and Data System assessment at the end of imaging work-up (up to 90 days after acquisition of the index diagnostic mammogram and before biopsy) with use of standard BCSC definitions (20). Women were considered to have breast cancer if a diagnosis of invasive carcinoma or ductal carcinoma in situ was made within 1 year after mammography.

Performance measures included sensitivity, false-positive rate (FPR) per 100 mammograms, and cancer detection rate (CDR) per 1000 mammograms (15).

### Statistical Analyses

All analyses were stratified by indication. Volume measures were analyzed in relation to performance in the subsequent year. For instance, 2005 volume was linked to 2006 performance (18).

We examined the distributions of radiologist- and mammogram-level characteristics with categorized volume measures. To describe overall performance, we calculated performance measures for each radiologist collapsing across all examinations and years. We examined unadjusted associations by estimating performance by using all mammograms within each categorized volume stratum and calculating 95% confidence intervals (CIs) (21).

The observed volume distributions were heavily skewed, with sparse information at very high volumes. To ensure stable estimation, we restricted our modeling to mammograms with annual average volume measures in the following ranges: total volume, 480–7000 mammograms; screening volume, 480–6000 mammograms; diagnostic volume, 0–2500

mammograms; and diagnostic focus, 1%–40%.

To estimate the adjusted effect of volume on diagnostic performance, we fit separate logistic regression models for each performance measure. Models for sensitivity included mammograms obtained in patients in whom a cancer diagnosis was made within 1 year after mammography, with final assessment (positive or negative) as the outcome. Models for FPR were also used in the final assessment but were based on mammograms obtained in patients in whom no cancer diagnosis was made within 1 year after mammography. Models for CDR included all mammograms and used the outcome of whether cancer was diagnosed within 1 year after positive final assessment. All models were adjusted for BCSC registry. Patient-level characteristics were age, family history, self-reported presence of a lump, and time since last mammographic examination. Radiologist-level characteristics were years of experience with mammogram interpretation and percentage of time spent in breast imaging. Breast density was not included because of the high rate of missing data.

To avoid making restrictive assumptions about the shape of the volume-performance relationships, we used natural cubic splines with one knot at the midpoint of the prespecified volume ranges (22). To standardize interpretations, we set the following reference levels: 2000 mammograms for total volume, 1500 mammograms for screening volume, 1000 mammograms for diagnostic volume, and 20% for diagnostic focus. The tables show estimated adjusted odds ratios and 95% CIs; Appendix E1 (online) provides additional detail.

We hypothesized that a self-reported lump could modify the effect of volume on interpretive performance for mammograms that showed a breast concern or problem. Thus, we tested for an interaction between self-reported lump and each volume measure. For volume and performance combinations with significant interactions, we present lump stratum-specific results; otherwise, we present results from the main effects-only model (with adjustment for self-reported lump).

**Table 1**

**Distribution of Radiologist Characteristics across 107 Radiologists by Average Annual Diagnostic Volume**

Characteristic	Overall Distribution	Average Annual Diagnostic Volume					
		<100 Mammograms	100–199 Mammograms	200–299 Mammograms	300–499 Mammograms	500–999 Mammograms	≥1000 Mammograms
Overall distribution		14	10	22	32	11	10
<b>Demographics</b>							
Age at survey (y)							
<45	24	20	18	38	24	33	0
45–54	29	27	9	25	29	50	36
≥55	47	53	73	38	47	17	64
Sex							
Male	66	87	64	88	59	42	45
Female	34	13	36	12	41	58	55
Works full time (≥40 h/wk)							
No	25	27	36	9	27	25	36
Yes	75	73	64	91	73	75	64
Unknown	2	0	0	4	3	0	0
Affiliation with academic medical center							
No affiliation	73	60	64	88	88	50	40
Adjunct	8	13	0	12	3	0	30
Primary	19	27	36	0	9	50	30
Unknown	1	0	0	0	0	0	9
Experience							
Time since graduated residency (y)							
<10	16	7	18	29	12	25	0
10–19	30	27	18	29	39	17	36
≥20	54	67	64	42	48	58	64
Unknown	1	0	0	0	3	0	0
Combined variable of fellowship training and duration of mammogram interpretation (y)							
No fellowship, <10	17	20	18	25	19	0	0
No fellowship, 10–19	32	13	27	42	41	18	33
No fellowship, ≥20	48	67	55	33	41	55	67
Fellowship, <10	3	0	0	0	0	27	0
Fellowship, ≥10	5	0	0	0	6	8	18
Time working in breast imaging (%)							
<20	26	29	30	52	21	8	0
20–39	25	14	40	30	36	8	0
40–79	15	7	10	0	15	58	9
80–100	34	50	20	17	27	25	91
Unknown	4	7	9	4	3	0	0
Interpretive volume							
Average annual screening volume							
<480 mammograms	3	20	0	0	0	0	0
480–999 mammograms	26	53	64	42	9	0	0
1000–1499 mammograms	16	13	27	21	21	0	0
1500–1999 mammograms	23	7	0	33	41	17	0
2000–2999 mammograms	13	7	0	4	21	33	9
≥3000 mammograms	19	0	9	0	9	50	91
Average annual total volume							
<480 mammograms	2	13	0	0	0	0	0
480–999 mammograms	16	53	55	12	0	0	0
1000–1499 mammograms	22	20	36	46	18	0	0
1500–1999 mammograms	13	7	0	21	24	0	0

Table 1 (continues)

Table 1 (continued)

## Distribution of Radiologist Characteristics across 107 Radiologists by Average Annual Diagnostic Volume

Characteristic	Overall Distribution	Average Annual Diagnostic Volume					
		<100 Mammograms	100–199 Mammograms	200–299 Mammograms	300–499 Mammograms	500–999 Mammograms	≥1000 Mammograms
2000–2999 mammograms	22	0	0	21	47	25	0
3000–4999 mammograms	14	7	9	0	9	58	27
≥5000 mammograms	10	0	0	0	3	17	73
Average annual diagnostic focus (%)							
≤10	20	80	36	12	6	0	0
11–15	23	13	27	46	18	17	9
16–20	36	7	27	33	53	58	9
21–25	12	0	9	4	18	8	36
>25	9	0	0	4	6	17	45
Average annual no. of facilities where mammograms were interpreted							
1	30	27	45	42	18	33	27
>1–2	45	67	36	25	50	50	45
>2–3	14	7	18	17	21	8	0
>3	11	0	0	17	12	8	27
Amount of volume and performance data (y)							
1	7	13	18	8	6	0	0
2	7	20	9	4	0	17	0
3	16	20	27	21	3	17	27
4	21	27	9	12	26	33	9
5	50	20	36	54	65	33	64

Note.—All data are percentages. Unknown percentages are based on all 107 radiologists; remaining percentages are based on the number with the characteristic known.

Generalized estimating equations were used to estimate model parameters (23,24). To account for within-radiologist correlation, we adopted a working exchangeable correlation structure and calculated standard errors, two-sided *P* values, and 95% CIs with the robust sandwich estimator (23). Sensitivity analyses included individually excluding each registry to ensure none overly influenced results, not adjusting for radiologist-level characteristics, and adjusting for breast density. Statistical significance was based on two-sided *P* values at the .05 level. All analyses were performed by using R software (version 2.12.0; R Foundation for Statistical Computing, Vienna, Austria) (25).

## Results

Our study included 107 radiologists who interpreted 117 136 diagnostic mammograms (46 369 obtained because additional imaging was required, 70 767

obtained because of a breast concern or problem) obtained in 98 667 women. Most radiologists (86%) contributed at least 3 years of volume data, performance data, or both (50% contributed 5 years of volume data), for a total of 426 reader-years.

The median age of radiologists was 54 years (age range, 37–72 years); most radiologists were men (66%), worked full time (75%), and had interpreted mammograms for at least 10 years (84%) (Table 1). Approximately 34% of the radiologists spent most (80%–100%) of their time working in breast imaging. The median average annual diagnostic volume was 315 mammograms; 15 (14%) of the 107 radiologists interpreted fewer than 100 diagnostic mammograms on average; 11 (10%) interpreted at least 1000 mammograms. Readers with the highest average annual diagnostic volume (≥1000 mammograms) were more likely to be women (55%, compared with 34% overall), spend more

time in breast imaging (91% spend 80%–100% of their time in breast imaging, compared with 34% overall), and have higher total and screening volume and a higher diagnostic focus (Table 1).

Table 2 presents characteristics of women who underwent mammography for the purpose of additional imaging or to evaluate a breast concern or problem; Tables E2 and E3 (online) present this information stratified by diagnostic volume. A greater proportion of mammograms were obtained to evaluate a breast concern or problem in women aged at least 40 years (17%, compared with 4% for additional imaging mammography). Only 2% of the 42 526 mammograms obtained for additional imaging that were not missing symptom information revealed a lump compared with 33% of the 65 691 mammograms obtained because of a breast concern or problem that had nonmissing symptom information. Eighteen percent of mammograms obtained in patients who were

not missing self-reported family history indicated a positive prior history.

Table 3 presents unadjusted performance estimates across categorized volume measures. No consistent patterns emerged for additional imaging mammography. For mammography performed because of a breast concern or problem, sensitivity increased as average annual diagnostic volume and diagnostic focus increased. Further, CDR decreased with increased average annual total volume, screening volume, and diagnostic volume.

In adjusted analyses, diagnostic volume was significantly ( $P < .001$ ) associated with sensitivity for additional imaging mammography (Table 4, Fig E4 [online]); the estimated odds ratios were bell shaped, with an apex at 1000 mammograms. When we compared diagnostic volumes of 1500 and 1000 mammograms, the estimated odds ratio was 0.82 (95% CI: 0.69, 0.99); that is, the odds of a true-positive assessment for an additional imaging mammogram interpreted by a radiologist with a diagnostic volume of 1500 mammograms are approximately 18% lower than those for a radiologist with a diagnostic volume of 1000 mammograms. When we compared a diagnostic volume of 2000 mammograms with a diagnostic volume of 1000 mammograms, the estimated odds ratio for sensitivity decreased to 0.43 (95% CI: 0.29, 0.64). When we decreased diagnostic volume to 300 mammograms, the estimated odds ratio for sensitivity compared with 1000 mammograms decreased to 0.65 (95% CI: 0.35, 1.19). There was insufficient evidence of an association between diagnostic volume and FPR ( $P = .20$ ) or CDR ( $P = .061$ ).

Mammography performed to evaluate a breast concern or problem revealed no interaction between diagnostic volume and self-reported lump for sensitivity ( $P = .27$ ) or CDR ( $P = .20$ ); for FPR, the interaction was significant ( $P = .011$ ). On the basis of a main effects-only model, the association between diagnostic volume and sensitivity increased until a diagnostic volume of approximately 1500 mammograms was reached; thereafter, the association

Table 2

### Distribution of Patient and Radiologist Characteristics for Mammograms Obtained for Additional Imaging and for Breast Concern or Problem

Characteristic	Additional Imaging ( <i>n</i> = 46 369)	Breast Concern or Problem ( <i>n</i> = 70 767)
<b>Patient-level characteristics</b>		
Age (y)		
<40	1694 (4)	11 973 (17)
40–49	16 132 (35)	20 219 (29)
50–59	14 243 (31)	17 913 (25)
60–69	7994 (17)	10 749 (15)
70–79	4578 (10)	6769 (10)
≥80	1728 (4)	3144 (4)
Breast density		
Almost entirely fat	1783 (5)	4250 (9)
Scattered fibroglandular tissue	15 534 (40)	18 064 (37)
Heterogeneously dense	19 063 (49)	20 706 (43)
Extremely dense	2403 (6)	5466 (11)
Unknown	7586 (16)	22 281 (31)
First-degree family history		
No	33 892 (82)	51 605 (82)
Yes	7217 (18)	11 247 (18)
Unknown	5260 (11)	7915 (11)
Self-reported lump		
No	41 555 (98)	43 788 (67)
Yes	971 (2)	21 903 (33)
Unknown	3843 (8)	5076 (7)
Time since last mammography (mo)		
No previous mammography	3706 (8)	6307 (10)
<12	1918 (4)	20 283 (32)
12–35	33 201 (75)	31 112 (49)
35–59	3022 (7)	3519 (5)
≥60	2176 (5)	2785 (4)
Unknown	2346 (5)	6761 (10)
<b>Radiologist-level characteristics</b>		
Average annual diagnostic volume		
<100 mammograms	853 (2)	1044 (1)
100–199 mammograms	1219 (3)	1340 (2)
200–299 mammograms	6645 (14)	4814 (7)
300–499 mammograms	14 398 (31)	13 984 (20)
500–999 mammograms	8028 (17)	9230 (13)
≥1000 mammograms	15 226 (33)	40 355 (57)
Average annual screening volume		
<480 mammograms	122 (0)	105 (0)
480–1000 mammograms	5073 (11)	3111 (4)
1000–1499 mammograms	5361 (12)	4154 (6)
1500–1999 mammograms	9153 (20)	9429 (13)
2000–2999 mammograms	8257 (18)	15 447 (22)
≥3000 mammograms	18 403 (40)	38 521 (54)
Average annual total volume		
<480 mammograms	17 (0)	6 (0)
480–1000 mammograms	1890 (4)	1494 (2)
1000–1499 mammograms	6326 (14)	4259 (6)
1500–1999 mammograms	5018 (11)	4135 (6)
2000–2999 mammograms	10 962 (24)	11 241 (16)

Table 2 (continues)

Table 2 (continued)

**Distribution of Patient and Radiologist Characteristics for Mammograms Obtained for Additional Imaging and for Breast Concern or Problem**

Characteristic	Additional Imaging ( <i>n</i> = 46 369)	Breast Concern or Problem ( <i>n</i> = 70 767)
3000–4999 mammograms	9976 (22)	12 141 (17)
≥5000 mammograms	12 180 (26)	37 491 (53)
Average annual diagnostic focus (%)		
≤10	2287 (5)	3694 (5)
11–15	6350 (14)	6862 (10)
16–20	18 307 (39)	15 960 (23)
21–25	11 947 (26)	10 497 (15)
>25	7478 (16)	33 754 (48)

Note.—Data are numbers of mammograms, and data in parentheses are percentages. Unknown percentages are based on all mammograms; remaining percentages are based on mammograms with the characteristic known.

reached a plateau ( $P = .049$ ) (Table 4, Figure E4 [online]). Given the significant interaction, results for FPR have been stratified according to self-reported lump. For women who did not report a lump, diagnostic volume was associated with FPR for mammography performed to evaluate a breast concern or problem ( $P = .004$ ). The estimated association was bell shaped, increased to an apex of approximately 1100 mammograms, and then decreased. In women who reported a lump, we detected no association ( $P = .58$ ). CDR displayed no evidence of an association with diagnostic volume for mammography performed because of a breast concern or problem ( $P = .24$ ).

For additional imaging mammography, diagnostic focus exhibited a marginal association with sensitivity ( $P = .055$ ) and significant associations with FPR ( $P = .024$ ) and CDR ( $P = .007$ ) (Table 5, Fig E5 [online]). For both sensitivity and FPR, the estimated odds ratio curve is bell shaped, increasing to apexes of 18% and 23%, respectively, and then decreasing. When we compared diagnostic focuses of 10% and 20%, the estimated adjusted odds ratio for FPR was 0.85 (95% CI: 0.75, 0.97). That means the odds of a false-positive assessment for an additional imaging mammogram interpreted by a radiologist with a diagnostic focus of 10% were approximately 15% lower than the odds of a false-positive assessment for

an additional imaging mammogram interpreted by a radiologist with 20% diagnostic focus. When we compared two additional imaging mammograms interpreted by radiologists with diagnostic foci of 5% and 20%, the odds ratio for FPR decreased to 0.75 (95% CI: 0.60, 0.94). For CDR, the estimated odds ratio curve increased linearly from 0.83 to 1.33, respectively, when we compared diagnostic foci of 5% and 40% with a focus of 20%

For all three performance measures, the interaction between diagnostic focus and self-reported lump in mammography performed to address a breast concern or problem was significant ( $P = .020$ ,  $P < .001$ , and  $P < .027$  for sensitivity, FPR, and CDR, respectively). Among women who did not report a lump, the association between diagnostic focus and both sensitivity and FPR was bell shaped (Table 5, Fig E5 [online]); only the association with FPR was significant ( $P < .001$ ). For CDR, the estimated odds ratio curve linearly decreased from 1.27 to 0.87, respectively, when we compared diagnostic foci of 5% and 40% with a focus of 20%; however, this difference was not significant ( $P = .293$ ). Among women who reported a lump, diagnostic focus was marginally associated with sensitivity ( $P = .051$ ); when compared with a focus of 20%, increasing the diagnostic focus from 5% to 30% corresponded to an estimated odds ratio increase (from

0.58 [95% CI: 0.31, 1.09] to 1.47 [95% CI: 0.79, 2.75]).

While not significant, there was borderline evidence of an increasing association between screening volume and FPR for both additional imaging mammography ( $P = .098$ ) and breast concern or problem mammography ( $P = .076$ ) (Table E5, Fig E6 [online]). There were nonlinear associations between screening volume and sensitivity among women who underwent mammography to address a breast concern or problem. In women who did not report a lump, the estimated association was bell-shaped, with an apex at approximately 3000 mammograms ( $P = .003$ ); in women who did report a lump, the estimated association increased until approximately 4000 mammograms and then plateaued ( $P = .148$ ). For CDR, we found no evidence of an association with screening volume. Neither indication had consistent evidence of a relationship between total volume and diagnostic performance (Table E6 [online]).

The clinical interpretation of results did not change after we completed the sensitivity analyses described previously.

## Discussion

We found that diagnostic volume and diagnostic focus were most consistently associated with performance; neither screening volume nor total volume was consistently associated with performance. For diagnostic volume and diagnostic focus, the associations were approximately quadratic: Radiologists with either low volume/focus or high volume/focus had the lowest FPR; radiologists with an annual diagnostic volume of approximately 1000 mammograms or a diagnostic focus of approximately 20% had the highest FPR. In parallel, radiologists with either low volume/focus or high volume/focus had the lowest sensitivity, indicating an important tradeoff between the two performance measures. These results generally held for both additional imaging and breast concern or problem diagnostic mammography. In contrast, CDR for additional imaging mammography increased linearly as

Table 3

## Unadjusted Diagnostic Performance Measure Estimates across Categorized Volume Measures, Stratified by Indication

Volume and Focus	Additional Imaging			Evaluation of a Breast Concern or Problem		
	Sensitivity	FPR	CDR	Sensitivity	FPR	CDR
<b>Average annual diagnostic volume</b>						
<100 mammograms	85.6 (77.3, 91.8)	9.4 (8.2, 10.6)	32.5 (25.8, 40.1)	83.8 (76.2, 89.8)	7.6 (6.6, 8.7)	36.7 (29.8, 44.5)
100–199 mammograms	81.0 (61.1, 93.7)	11.4 (8.6, 14.6)	36.9 (22.2, 56.7)	83.3 (69.1, 93.0)	6.9 (5.0, 9.3)	52.7 (36.3, 73.1)
200–299 mammograms	85.4 (80.1, 89.8)	12.1 (11.2, 13.1)	35.6 (30.6, 41.0)	81.0 (75.1, 86.1)	7.9 (7.1, 8.7)	36.3 (31.0, 42.1)
300–499 mammograms	91.3 (88.4, 93.8)	13.0 (12.3, 13.7)	39.0 (35.3, 42.9)	85.7 (81.7, 89.1)	8.5 (7.9, 9.2)	41.7 (37.2, 46.6)
500–999 mammograms	92.9 (89.5, 95.4)	10.1 (9.3, 10.9)	43.3 (38.5, 48.6)	85.8 (81.5, 89.5)	7.5 (6.9, 8.2)	36.4 (32.2, 41.1)
1000–1499 mammograms	96.0 (93.2, 98.0)	11.3 (10.5, 12.2)	40.2 (35.4, 45.3)	90.0 (85.6, 93.4)	8.5 (7.8, 9.2)	35.0 (30.5, 39.9)
1500–1999 mammograms	86.3 (78.7, 92.0)	13.8 (12.3, 15.5)	48.5 (39.2, 59.0)	91.7 (86.5, 95.5)	9.3 (8.5, 10.3)	31.3 (26.4, 36.8)
≥2000 mammograms	89.1 (81.7, 94.4)	19.7 (17.4, 22.1)	70.1 (56.4, 85.7)	88.3 (82.0, 93.1)	7.9 (7.0, 8.8)	29.5 (24.4, 35.2)
<b>Average annual screening volume</b>						
480–1000 mammograms	90.7 (86.8, 93.8)	13.6 (12.7, 14.6)	47.7 (42.1, 53.8)	84.4 (79.7, 88.4)	7.4 (6.8, 8.1)	36.3 (31.8, 41.1)
1000–1499 mammograms	85.3 (79.1, 90.2)	10.9 (10.0, 11.9)	31.9 (26.9, 37.6)	88.2 (82.9, 92.4)	7.8 (7.0, 8.7)	39.5 (33.7, 45.8)
1500–1999 mammograms	88.0 (82.4, 92.3)	10.5 (9.6, 11.5)	33.0 (28.0, 38.6)	82.2 (75.9, 87.6)	6.2 (5.4, 7.0)	37.0 (31.2, 43.5)
2000–2999 mammograms	93.8 (91.2, 95.9)	13.7 (12.9, 14.4)	46.4 (42.0, 51.1)	84.2 (80.6, 87.4)	8.2 (7.6, 8.7)	34.6 (31.2, 38.2)
3000–3999 mammograms	90.1 (86.5, 93.1)	13.9 (13.0, 14.9)	50.0 (44.5, 55.9)	89.8 (86.4, 92.7)	9.2 (8.6, 9.8)	34.0 (30.4, 37.8)
≥4000 mammograms	92.7 (88.9, 95.5)	9.8 (9.1, 10.6)	37.9 (33.2, 42.9)	87.9 (82.5, 92.2)	6.5 (5.9, 7.2)	29.3 (24.9, 34.1)
<b>Average annual total volume</b>						
480–1000 mammograms	87.6 (82.5, 91.8)	12.0 (11.1, 13.0)	36.1 (31.0, 41.7)	84.4 (79.0, 89.0)	9.9 (8.9, 11.0)	51.4 (44.2, 59.4)
1000–1499 mammograms	85.5 (76.5, 92.2)	9.0 (7.8, 10.2)	28.3 (22.0, 35.6)	89.0 (81.1, 94.6)	8.2 (6.9, 9.5)	40.7 (32.2, 50.5)
1500–1999 mammograms	88.0 (83.0, 92.0)	12.5 (11.6, 13.5)	37.4 (32.2, 43.0)	87.8 (82.5, 92.0)	7.4 (6.6, 8.2)	39.9 (34.1, 46.3)
2000–2999 mammograms	93.4 (90.2, 95.8)	12.4 (11.6, 13.2)	41.4 (36.9, 46.4)	82.8 (77.9, 87.0)	8.1 (7.4, 8.8)	35.0 (30.7, 39.7)
3000–4999 mammograms	91.2 (88.3, 93.7)	13.2 (12.4, 14.0)	50.3 (45.5, 55.5)	84.9 (81.4, 88.1)	7.0 (6.6, 7.5)	30.7 (27.8, 33.9)
≥5000 mammograms	92.4 (89.5, 94.8)	12.1 (11.4, 12.8)	44.7 (40.4, 49.4)	89.2 (86.0, 91.9)	7.9 (7.4, 8.4)	30.2 (27.2, 33.3)
<b>Average annual diagnostic focus (%)</b>						
≤5	88.6 (82.5, 93.3)	11.8 (10.7, 13.0)	37.8 (31.4, 44.9)	80.1 (73.0, 86.2)	8.6 (7.5, 9.7)	41.9 (34.8, 49.9)
6–10	85.2 (68.9, 95.1)	13.3 (10.6, 16.2)	39.3 (25.5, 57.1)	82.5 (68.9, 92.1)	7.6 (6.1, 9.2)	28.4 (19.8, 38.9)
11–15	89.6 (86.1, 92.6)	12.2 (11.5, 13.0)	37.7 (33.7, 42.0)	83.5 (79.2, 87.4)	6.8 (6.2, 7.4)	37.7 (33.4, 42.3)
16–20	91.1 (88.0, 93.6)	11.9 (11.3, 12.6)	38.0 (34.3, 42.0)	85.0 (80.7, 88.8)	7.9 (7.3, 8.5)	32.3 (28.5, 36.4)
21–25	91.6 (88.3, 94.3)	9.4 (8.7, 10.0)	39.2 (35.0, 43.8)	93.2 (89.6, 95.9)	8.8 (7.9, 9.6)	49.9 (43.9, 56.4)
26–30	92.9 (87.8, 96.4)	16.0 (14.5, 17.6)	59.4 (50.0, 69.8)	87.8 (83.3, 91.6)	10.0 (9.2, 10.8)	36.1 (31.5, 41.2)
>30	87.5 (79.5, 93.3)	17.1 (15.1, 19.1)	52.3 (41.7, 64.5)	90.0 (84.1, 94.4)	8.9 (8.1, 9.8)	27.6 (22.9, 32.8)

Note.—Data are estimates, and data in parentheses are 95% CIs. Estimates were obtained by collapsing across all mammograms within the volume stratum, and 95% CIs were calculated by inverting a likelihood ratio test. FPR is rate per 100 mammograms, and CDR is rate per 1000 mammograms.

both diagnostic volume and diagnostic focus increased; no relationship between volume and CDR was found for breast concern or problem mammography.

The mechanisms by which radiologists acquire experience and translate it into improved diagnostic mammography performance are likely complex. Particularly challenging is the fact that both accrual of experience, for which we view radiologist interpretive volume as a measurable surrogate, and diagnostic performance are multifaceted. That both annual diagnostic volume and diagnostic focus exhibited the strongest evidence of an association suggests that

a key determinant in developing thresholds for considering diagnostic examination findings abnormal is the accrual of diagnostic mammography experience, as opposed to experience in the interpretation of screening mammograms.

Both diagnostic volume and diagnostic focus had strong nonlinear relationships with performance: radiologists at the extremes of the volume range had worse performance for sensitivity and better performance for FPR. This may reflect changing mechanisms by which increasing diagnostic experience may influence thresholds for recommending biopsy of a suspicious lesion; as diag-

nostic volume or diagnostic focus increases, radiologists are exposed to more mammograms obtained in women who ultimately receive a breast cancer diagnosis. One hypothesis is that radiologists benefit from this because they learn to recognize different abnormalities as benign or malignant and then, because they have experience interpreting a broader range of lesions, sensitivity increases and FPR decreases. However, the incremental learning benefit may diminish with volume, leading to a plateau effect, or it may ultimately be detrimental as workload increases to very high levels.



Table 4

## Estimated Adjusted Odds Ratio Associations between Annual Diagnostic Volume and Diagnostic Performance

Annual Diagnostic Volume	Sensitivity		FPR		CDR	
	Odds Ratio	P Value*	Odds Ratio	P Value*	Odds Ratio	P Value*
Additional imaging	...	<.001	...	.198	...	.061
100 mammograms	0.53 (0.23, 1.21)	...	0.83 (0.64, 1.06)	...	0.83 (0.64, 1.07)	...
200 mammograms	0.59 (0.29, 1.20)	...	0.85 (0.68, 1.05)	...	0.85 (0.68, 1.06)	...
300 mammograms	0.65 (0.35, 1.19)	...	0.87 (0.72, 1.04)	...	0.87 (0.72, 1.04)	...
400 mammograms	0.71 (0.43, 1.17)	...	0.89 (0.76, 1.03)	...	0.89 (0.76, 1.03)	...
500 mammograms	0.78 (0.52, 1.16)	...	0.91 (0.81, 1.02)	...	0.90 (0.80, 1.02)	...
1000 mammograms	1.00 (NA)	...	1.00 (NA)	...	1.00 (NA)	...
1500 mammograms	0.82 (0.69, 0.99)	...	1.05 (0.99, 1.12)	...	1.09 (1.01, 1.18)	...
2000 mammograms	0.43 (0.29, 0.64)	...	1.06 (0.90, 1.25)	...	1.19 (0.97, 1.46)	...
2500 mammograms	0.18 (0.08, 0.39)	...	1.04 (0.76, 1.43)	...	1.28 (0.86, 1.89)	...
Breast concern or problem						
No self-reported lump	...	.049†	...	.004	...	.244†
100 mammograms	0.50 (0.28, 0.87)	...	0.63 (0.47, 0.84)	...	0.84 (0.65, 1.10)	...
200 mammograms	0.54 (0.34, 0.88)	...	0.67 (0.53, 0.86)	...	0.86 (0.69, 1.08)	...
300 mammograms	0.60 (0.39, 0.90)	...	0.72 (0.59, 0.89)	...	0.88 (0.73, 1.06)	...
400 mammograms	0.65 (0.46, 0.92)	...	0.77 (0.65, 0.91)	...	0.90 (0.77, 1.05)	...
500 mammograms	0.71 (0.53, 0.93)	...	0.82 (0.72, 0.94)	...	0.92 (0.82, 1.03)	...
1000 mammograms	1.00 (NA)	...	1.00 (NA)	...	1.00 (NA)	...
1500 mammograms	1.19 (0.96, 1.47)	...	0.93 (0.82, 1.07)	...	1.04 (0.94, 1.15)	...
2000 mammograms	1.18 (0.71, 1.95)	...	0.66 (0.45, 0.99)	...	1.04 (0.75, 1.45)	...
2500 mammograms	1.07 (0.44, 2.58)	...	0.41 (0.19, 0.88)	...	1.01 (0.53, 1.93)	...
Self-reported lump	...	NA	...	.582	...	NA
100 mammograms	NA	...	0.85 (0.63, 1.16)	...	NA	...
200 mammograms	NA	...	0.87 (0.67, 1.14)	...	NA	...
300 mammograms	NA	...	0.89 (0.71, 1.12)	...	NA	...
400 mammograms	NA	...	0.91 (0.76, 1.09)	...	NA	...
500 mammograms	NA	...	0.93 (0.80, 1.07)	...	NA	...
1000 mammograms	NA	...	1.00 (NA)	...	NA	...
1500 mammograms	NA	...	1.03 (0.95, 1.11)	...	NA	...
2000 mammograms	NA	...	1.00 (0.81, 1.24)	...	NA	...
2500 mammograms	NA	...	0.95 (0.62, 1.46)	...	NA	...

Note.—Unless otherwise indicated, data are odds ratios, and data in parentheses are 95% CIs. FPR is rate per 100 mammograms, and CDR is rate per 1000 mammograms. For mammography performed to evaluate a breast concern or problem, an interaction with self-reported presence of a lump was evaluated. When the interaction was not significant, a main-effects only model was fit, and lump stratum-specific estimates were the same. When the interaction was significant, lump stratum-specific estimates differed. Results correspond to Figure E4 (online). Sensitivity and CDR data were not obtained for women with a self-reported lump because interaction with lump was not significant (see results for No self-reported lump). NA = not applicable.

\* Two-sided omnibus P value for overall association based on natural cubic spline model.

† P value for main effects—only model when an interaction between volume and presence of self-reported lump was not significant.

We found evidence of significant interaction between the presence of a lump and both diagnostic volume and diagnostic focus for FPR in mammography performed because of a breast concern or problem. Before the study, we hypothesized that the skill required to examine a lump may be different than that required for other concerns and that the accrual of experience may be more beneficial to the interpretation of mammograms obtained in women who

did not have self-reported lumps (the latter generally are more ambiguous).

Buist et al (18) recently reported on a comprehensive study of the association between interpretive volume and screening performance. Radiologists with higher annual volume had clinically and significantly lower FPR with similar sensitivity when compared with radiologists with lower volume. In addition, radiologists with a greater screening focus had lower sensitivity, FPR, and CDR. Food

and Drug Administration regulations state that U.S. physicians who interpret mammograms must have interpreted 960 mammograms within the previous 24 months. The Food and Drug Administration does not distinguish between screening and diagnostic mammography for volume requirements. As we have found, the relationships between continuing experience and diagnostic performance are complex, confounding regulations based on the simple use of

Table 5

## Estimated Adjusted Odds Ratio Associations and Annual Diagnostic Focus and Diagnostic Performance

Annual Diagnostic Focus	Sensitivity		FPR		CDR	
	Odds Ratio	PValue*	Odds Ratio	P*	Odds Ratio	P*
Additional imaging (%)	...	.055	...	.024	...	.007
0	0.37 (0.12, 1.13)	...	0.65 (0.46, 0.90)	...	0.78 (0.52, 1.17)	...
5	0.55 (0.27, 1.14)	...	0.75 (0.60, 0.94)	...	0.83 (0.63, 1.09)	...
10	0.78 (0.53, 1.15)	...	0.85 (0.75, 0.97)	...	0.88 (0.76, 1.02)	...
20	1.00 (NA)	...	1.00 (NA)	...	1.00 (NA)	...
30	0.58 (0.36, 0.93)	...	0.97 (0.89, 1.05)	...	1.15 (1.03, 1.28)	...
40	0.20 (0.05, 0.75)	...	0.84 (0.68, 1.03)	...	1.33 (0.98, 1.82)	...
Breast concern or problem (%)						
No lump	...	.661	...	<.001	...	.293
0	0.54 (0.15, 2.02)	...	0.60 (0.37, 0.97)	...	1.40 (0.72, 2.72)	...
5	0.67 (0.28, 1.58)	...	0.74 (0.53, 1.03)	...	1.27 (0.81, 2.00)	...
10	0.81 (0.51, 1.28)	...	0.88 (0.72, 1.08)	...	1.16 (0.90, 1.51)	...
20	1.00 (NA)	...	1.00 (NA)	...	1.00 (NA)	...
30	0.91 (0.56, 1.47)	...	0.74 (0.65, 0.85)	...	0.92 (0.81, 1.04)	...
40	0.69 (0.18, 2.68)	...	0.43 (0.32, 0.57)	...	0.87 (0.62, 1.22)	...
Lump (%)	...	.051	...	.807	...	.338
0	0.48 (0.18, 1.29)	...	0.96 (0.57, 1.61)	...	1.08 (0.67, 1.72)	...
5	0.58 (0.31, 1.09)	...	0.98 (0.69, 1.39)	...	1.03 (0.76, 1.40)	...
10	0.69 (0.49, 0.98)	...	0.99 (0.81, 1.21)	...	1.00 (0.85, 1.17)	...
20	1.00 (NA)	...	1.00 (NA)	...	1.00 (NA)	...
30	1.47 (0.79, 2.75)	...	0.96 (0.86, 1.08)	...	1.13 (0.95, 1.35)	...
40	2.19 (0.42, 11.37)	...	0.91 (0.67, 1.22)	...	1.38 (0.83, 2.29)	...

Note.—Unless otherwise indicated, data are odds ratios, and data in parentheses are 95% CIs. FPR is rate per 100 mammograms, and CDR is rate per 1000 mammograms. For breast concern or problem mammography, an interaction with self-reported presence of a lump was evaluated and found to be significant; hence, lump stratum-specific results are presented. Results correspond to Figure E5 (online).

\* Two-sided omnibus P value for overall association based on natural cubic spline model.

a lower limit for continuing experience. Further, only 33% of diagnostic mammograms for which the radiologist-given indication was breast concern or problem were obtained in women who had reported a breast lump. This suggests some examinations performed in asymptomatic or quasisymptomatic women are routinely being classified as diagnostic examinations instead of as screening examinations. Regulatory thresholds for the number of diagnostic mammograms could therefore be easily circumvented by lack of adherence to strict definitions of what constitutes a screening examination, rendering the threshold ineffective.

This study had some limitations. These are observational data, and we cannot rule out the possibility of residual confounding or differences in case mix across radiologists and time. Further, because of sparse data, we restricted our analyses

to prespecified volume ranges. Although the observed associations may be validly interpreted within these ranges, the results may not be generalized to radiologists with very high volumes.

This study also had a number of important strengths. Detailed information on a large sample of mammograms from the BCSC permitted comprehensive characterization of volume and its association with diagnostic interpretive performance in community-based settings. Our volume data were verified by radiology facility records; therefore, we did not need to rely on self-reported estimates, and we were able to control for a number of well-known correlates of diagnostic performance. The prospective linkage of radiologist-specific volume measures with performance in the subsequent year strengthens the interpretation of our results. Finally, the coverage of the BCSC across the

United States ensures that our results are representative of the associations in community-based settings (15).

This study suggests that diagnostic mammography experience, more immediate feedback generated during the diagnostic process, or both may be beneficial and should be encouraged and shared among radiologists. Because of the multifaceted nature of interpretive volume and diagnostic performance, it is a challenge to characterize their relationship and doing so limits the practicality of setting regulatory thresholds for diagnostic mammographic interpretation.

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

**Disclosures of Potential Conflicts of Interest:**

**S.H.** No potential conflicts of interest to disclose. **D.S.M.B.** No potential conflicts of interest to disclose. **D.L.M.** No potential conflicts of interest to disclose. **M.L.A.** No potential conflicts of interest to disclose. **P.A.C.** No potential conflicts of interest to disclose. **T.O.** No potential conflicts of interest to disclose. **B.M.G.** No potential conflicts of interest to disclose. **K.K.** No potential conflicts of interest to disclose. **R.D.R.** No potential conflicts of interest to disclose. **B.C.Y.** No potential conflicts of interest to disclose. **J.G.E.** No potential conflicts of interest to disclose. **S.H.T.** No potential conflicts of interest to disclose. **R.A.S.** No potential conflicts of interest to disclose. **E.A.S.** No potential conflicts of interest to disclose.

**References**

- Sickles EA, Miglioretti DL, Ballard-Barbash R, et al. Performance benchmarks for diagnostic mammography. *Radiology* 2005; 235(3):775–790.
- Nass S, Ball J, Committee on Improving Mammography Quality Standards. Improving breast imaging quality standards. Washington, DC: National Academies Press, 2005.
- Barlow WE, Lehman CD, Zheng Y, et al. Performance of diagnostic mammography for women with signs or symptoms of breast cancer. *J Natl Cancer Inst* 2002;94(15): 1151–1159.
- 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(5):1173–1177.
- Flobbe K, van der Linden ES, Kessels AG, van Engelshoven JM. Diagnostic value of radiological breast imaging in a non-screening population. *Int J Cancer* 2001;92(4):616–618.
- Houssami N, Irwig L, Simpson JM, McKessar M, Blome S, Noakes J. The contribution of work-up or additional views to the accuracy of diagnostic mammography. *Breast* 2003;12(4):270–275.
- Houssami N, Irwig L, Simpson JM, McKessar M, Blome S, Noakes J. The influence of clinical information on the accuracy of diagnostic mammography. *Breast Cancer Res Treat* 2004;85(3):223–228.
- Yankaskas BC, Haneuse S, Kapp JM, et al. Performance of first mammography examination in women younger than 40 years. *J Natl Cancer Inst* 2010;102(10):692–701.
- Miglioretti DL, Smith-Bindman R, Abraham L, et al. Radiologist characteristics associated with interpretive performance of diagnostic mammography. *J Natl Cancer Inst* 2007; 99(24):1854–1863.
- Sickles EA, Wolverton DE, Dee KE. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology* 2002;224(3): 861–869.
- Goldman LE, Walker R, Miglioretti DL, Smith-Bindman R, Kerlikowske K; National Cancer Institute Breast Cancer Surveillance Consortium. Accuracy of diagnostic mammography at facilities serving vulnerable women. *Med Care* 2011;49(1):67–75.
- Jackson SL, Taplin SH, Sickles EA, et al. Variability of interpretive accuracy among diagnostic mammography facilities. *J Natl Cancer Inst* 2009;101(11):814–827.
- Jensen A, Vejborg I, Severinsen N, et al. Performance of clinical mammography: a nationwide study from Denmark. *Int J Cancer* 2006;119(1):183–191.
- 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(4):1001–1008.
- National Cancer Institute. Breast Cancer Surveillance Consortium. U.S. National Institutes of Health Web site. <http://breastscreening.cancer.gov/>. Updated February 24, 2011. Accessed April 29, 2011.
- Carney PA, Goodrich ME, Mackenzie T, et al. Utilization of screening mammography in New Hampshire: a population-based assessment. *Cancer* 2005;104(8):1726–1732.
- Elmore JG, Jackson SL, Abraham L, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology* 2009;253(3):641–651.
- Buist DS, Anderson ML, Haneuse SJ, et al. Influence of annual interpretive volume on screening mammography performance in the United States. *Radiology* 2011;259(1): 72–84.
- American College of Radiology. ACR BI-RADS–Mammography. In: ACR Breast Imaging and Reporting and Data System, Breast Imaging Atlas. 4th ed. Reston, Va: American College of Radiology, 2003.
- Breast Cancer Surveillance Consortium. BCSC Glossary of Terms. National Cancer Institute Web site. [http://breastscreening.cancer.gov/data/bcsc\\_data\\_definitions.pdf](http://breastscreening.cancer.gov/data/bcsc_data_definitions.pdf). Updated September 16, 2009. Accessed May 4, 2011.
- Lehmann EL, Romano JP. Testing statistical hypotheses. New York, NY: Springer, 2005.
- Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. New York, NY: Springer, 2009.
- Liang KY, Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73(1):13–22.
- Miglioretti DL, Haneuse SJ, Anderson ML. Statistical approaches for modeling radiologists' interpretive performance. *Acad Radiol* 2009;16(2):227–238.
- Team RDC. R: A language and environment for statistical computing. R Foundation for Statistical Computing Web site. <http://cran.r-project.org/>. Accessed July 21, 2011.