New Mammography Screening Performance Metrics Based on the Entire Screening Episode

Brian L. Sprague, PhD,[1] Diana L. Miglioretti, PhD,[2] Christoph I. Lee, MD,[3] Hannah Perry, MD,[1] Anna A.N. Tosteson, ScD,[4] Karla Kerlikowske, MD[5]

[1]University of Vermont; [2]University of California, Davis; [3]University of Washington; [4]Geisel School of Medicine at Dartmouth; [5]University of California, San Francisco

Background

- Existing screening mammography performance metrics are based on the American College of Radiology (ACR) definitions
- ACR metrics consider only the "initial" assessment by the radiologist interpreting the initial screening mammogram
- ACR metrics were designed to evaluate radiologist's performance in breast imaging interpretation, yet are also widely used to inform women, healthcare providers, and policymakers regarding the potential benefits, harms, and limitations of mammography screening
- Screening outcomes that inform clinical decision-making are impacted by the interpretative performance **of the entire screening episode**
- The **entire screening episode** includes interpretation of both screening mammography and diagnostic imaging performed to work-up abnormal screens
- No existing metrics describe the performance of the entire screening episode

Purpose

- We calculated mammography screening performance metrics based on the final assessment of the entire screening episode.
- We compared these new metrics to conventional screening performance metrics.

Methods

- Observational data were prospectively collected by active Breast Cancer Surveillance Consortium (BCSC) breast imaging registries:
  - Carolina Mammography Registry
  - Kaiser Permanente Washington Registry
  - New Hampshire Mammography Network
  - Vermont Breast Cancer Surveillance System
  - San Francisco Mammography Registry
  - Metropolitan Chicago Breast Cancer Registry
- Eligible women were aged 40-79 with a screening mammogram during 2005-2017.
- To reflect regular participation in screening, all analyses were limited to women undergoing a screening mammogram within 30 months after a prior screening mammogram.
- Screening performance metrics were defined separately based on:
  1. the initial screening assessment per established ACR BI-RADS definitions; and
  2. the final assessment after diagnostic workup

Results

- Over 2.5 million mammography screening episodes identified among 791,347 individual women, with exams interpreted by 705 radiologists at 146 facilities
- 8.6% of screening episodes had a positive (abnormal) initial assessment. Among the initial positive screens, 64.8% had a negative final assessment, 19.5% had a category 3 short-interval follow-up final assessment, and 15.7% had a category 4/5 biopsy recommendation final assessment.

Table 1. Screening mammography performance metrics based on initial vs. final assessment.

- Cancer detection rate was similar for final (4.1 per 1000; 95% CI: 3.8-4.3) vs. initial assessment (4.1 per 1000; 95% CI: 3.9-4.3).
- Interval cancer rate was 12% higher based on final (0.77 per 1000; 95% CI: 0.71-0.83) vs. initial assessment (0.69 per 1000; 95% CI: 0.64-0.74)
- Modest difference in sensitivity based on final (84.1% [95% CI: 83.0-85.1] vs. initial (85.7% [95% CI: 84.8-86.6%) assessment.

	Initial Assessment		Final Assessment	
Screening Performance Metric	**Estimate**	**95% CI**	**Estimate**	**95% CI**
Standard performance measures				
Cancer detection rate, per 1000	4.1	3.9, 4.3	4.1	3.8, 4.3
Interval cancer rate, per 1000	0.69	0.64, 0.74	0.77	0.71, 0.83
Sensitivity, %	85.7	84.8, 86.6	84.1	83.0, 85.1
Specificity, %	91.8	91.0, 92.4	97.4	97.0, 97.6
Screening benefit				
Stage I or IIa screen-detected invasive cancers, per 1000	2.7	2.5, 2.8	2.6	2.5, 2.8
Screening false-alarms				
False-positive recall for additional imaging, per 1000	82.1	75.3, 89.4	NA	NA
False-positive short interval follow-up recommendation, per 1000	NA	NA	16.5	13.9, 19.5
False-positive biopsy recommendation, per 1000	NA	NA	9.8	8.5, 11.3
Screening failures				
Stage IIb or higher interval invasive cancers, per 1000	0.16	0.15, 0.18	0.18	0.16, 0.20
Stage IIb or higher screen-detected invasive cancers, per 1000	0.29	0.26, 0.33	0.27	0.24, 0.31

CI, confidence interval; NA, not applicable.

Table 2. Screening performance measures based on initial vs. final assessment, according to breast density.

- Absolute differences in performance metrics between final and initial assessment increased with breast density

	BI-RADS Breast Density			
Screening Performance Metric	Almost entirely fat	Scattered fibroglandular density	Heterogeneously dense	Extremely dense
Based on initial assessment		***Estimate (95% CI)***		
Cancer detection rate, per 1000	2.6 (2.3, 2.9)	4.1 (3.9, 4.4)	4.4 (4.1, 4.7)	3.8 (3.4, 4.2)
Interval cancer rate, per 1000	0.21 (0.14, 0.30)	0.44 (0.39, 0.49)	0.93 (0.85, 1.03)	1.42 (1.23, 1.63)
Sensitivity, %	92.6 (89.7, 94.8)	90.4 (89.2, 91.4)	82.4 (80.8, 84.0)	72.6 (69.0, 76.0)
Specificity, %	95.5 (95.0, 95.9)	92.4 (91.6, 93.1)	90.1 (89.1, 91.0)	91.0 (90.0, 91.9)
Based on final assessment		***Estimate (95% CI)***		
Cancer detection rate, per 1000	2.6 (2.3, 2.9)	4.1 (3.8, 4.3)	4.3 (4.0, 4.6)	3.6 (3.3, 4.0)
Interval cancer rate, per 1000	0.23 (0.17, 0.33)	0.49 (0.44, 0.55)	1.03 (0.93, 1.14)	1.57 (1.38, 1.80)
Sensitivity, %	91.7 (88.7, 94.0)	89.3 (87.9, 90.5)	80.5 (78.6, 82.3)	69.6 (66.0, 73.1)
Specificity, %	98.4 (98.1, 98.6)	97.6 (97.3, 97.9)	96.9 (96.5, 97.2)	96.9 (96.5, 97.3)

BI-RADS, Breast Imaging Reporting and Data System; CI, confidence interval

Discussion

- Determination of screening performance metrics based on the final assessment of the screening episode rather than the initial assessment of the screening exam results in re-classification of approximately 1.9% of screen-detected cancers.
- This has a modest influence on the cancer detection rate and screening sensitivity, but corresponds to a 12% increase in the interval cancer rate.
- This phenomenon is largest among women with dense breasts.
- Women, clinicians, policymakers, and researchers should use final assessment performance metrics to support informed screening decisions.

The University of Vermont
LARNER COLLEGE OF MEDICINE

This work was supported by the National Cancer Institute (grant P01CA154292). Data collection for this research was additionally supported by grant U54CA163303 from the National Cancer Institute, grant R01 HS018366-01A1 from the Agency for Healthcare Research and Quality, and award PCS-1504-30370 from the Patient-Centered Outcomes Research Institute (PCORI)

Breast Cancer Surveillance Consortium
BCSC
Working together to advance breast cancer research