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, MD5
1University of Vermont; 2University of California, Davis; 3University of Washington; 4Geisel School of Medicine at Dartmouth; 5University 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>
<thead>
<tr>
<th>Table 1. Screening mammography performance metrics based on initial vs. final assessment.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer detection rate for final (4.1 per 1000; 95% CI: 3.8-4.3) vs. initial (4.1 per 1000; 95% CI: 3.9-4.3).</td>
</tr>
<tr>
<td>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).</td>
</tr>
<tr>
<td>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.</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Table 2. Screening performance measures based on initial vs. final assessment, according to breast density.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute differences in performance metrics between final and initial assessment increased with breast density.</td>
</tr>
</tbody>
</table>

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.