Identifying Radiologists who Might Benefit from Interventions

- Criteria based on multiple measures
- Confidence interval approach

Diana L. Miglioretti, PhD

Professor, UC Davis
Investigator, Group Health Research Institute

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

Outline

- Review criteria for minimally acceptable performance based on multiple measures

    - Sensitivity and specificity

    - Recall rate, cancer detection rate, and positive predictive value

- Review method for classifying radiologists based on observed performance

    - Often measured from a small number of mammograms

Criteria for identifying low performance of screening mammography

Measure	Low Performance Range	Percent of BCSC Radiologists in Low Performance Range
Sensitivity	<75%	18%
Specificity	<88% or >95%	48%
Recall Rate	<5% or >12%	49%
PPV1	<3% or >8%	38%
CDR	<2.5/1000	28%

Carney, et al. *Radiology*. 2010;255(2):354-61.

Limitations

- Most radiologists are in the "low" range for at least one measure

- Important to consider measures jointly
  - Low false positive rate OK if sensitivity is high

- Many radiologists interpret few mammograms associated with a cancer diagnosis
  - Observed measures imprecise, esp. sensitivity, cancer detection rate, and PPV
  - What if a radiologist recalled 7 out of 10 cancers?

Methods for Developing Combined Criteria

- 6 expert radiologists met in Seattle

- Started with screening performance criteria from Carney, et al. *Radiology*, 2010

- Considered multiple measures:

  - Sensitivity and specificity

  - Recall rate, cancer detection rate, and PPV
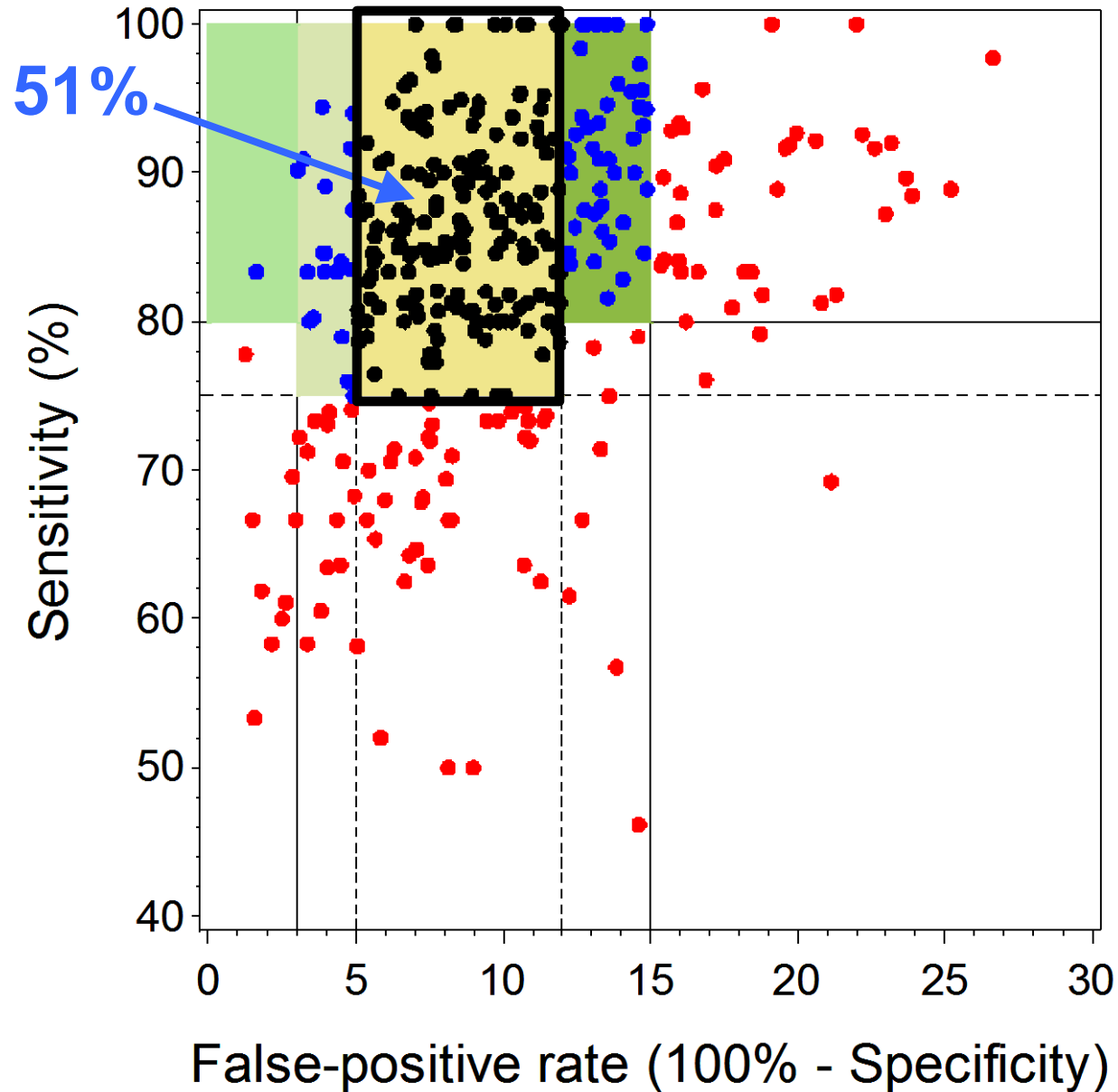
- Used BCSC data as benchmarks

Miglioretti, *et al.*; *AJR.* 2015;204(4):W486-91.

Sensitivity and Specificity Criteria

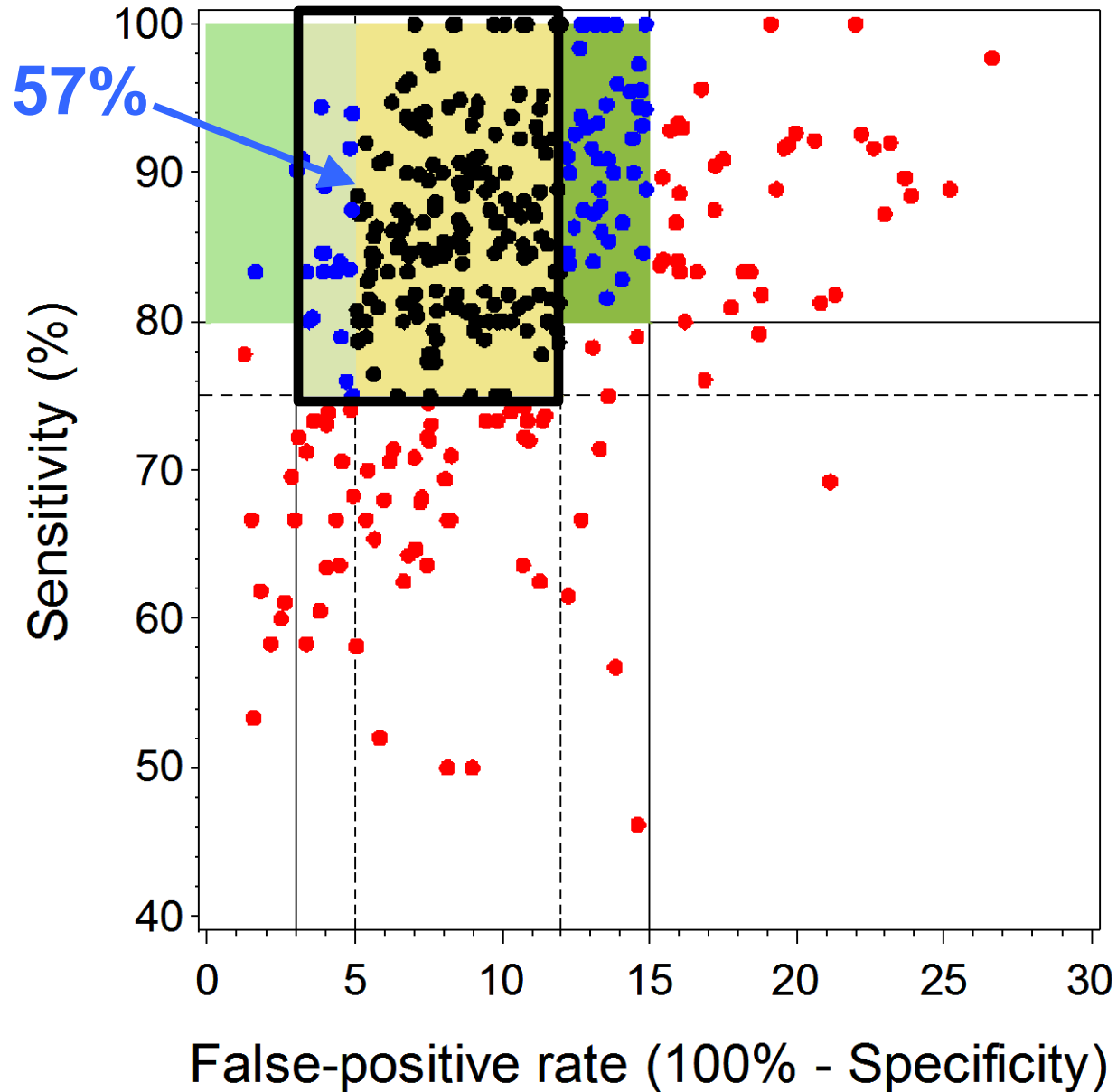*For radiologists who have complete cancer capture for all mammography cases*

Original Criteria

Miglioretti, *et al.*; *AJR.* 2015;204(4):W486-91.

Low FP Rate but Acceptable Sensitivity

Miglioretti, *et al.*; *AJR.* 2015;204(4):W486-91.

High Sensitivity (>80%)

69%

Sensitivity (%)

False-positive rate (100% - Specificity)

Miglioretti, *et al.*; *AJR.* 2015;204(4):W486-91.

Minimally-acceptable performance criteria for radiologists with complete cancer capture

Criteria	Sensitivity		Specificity	% of BCSC Radiologists who Met Criteria
Original	≥75%		88-95%	51%
Updated Criteria 1	≥80%	and	≥85%	62%
Updated Criteria 2	75-79%	and	88-97%	7%

Miglioretti, *et al.*; *AJR.* 2015;204(4):W486-91.

CDR, Recall, and PPV Criteria

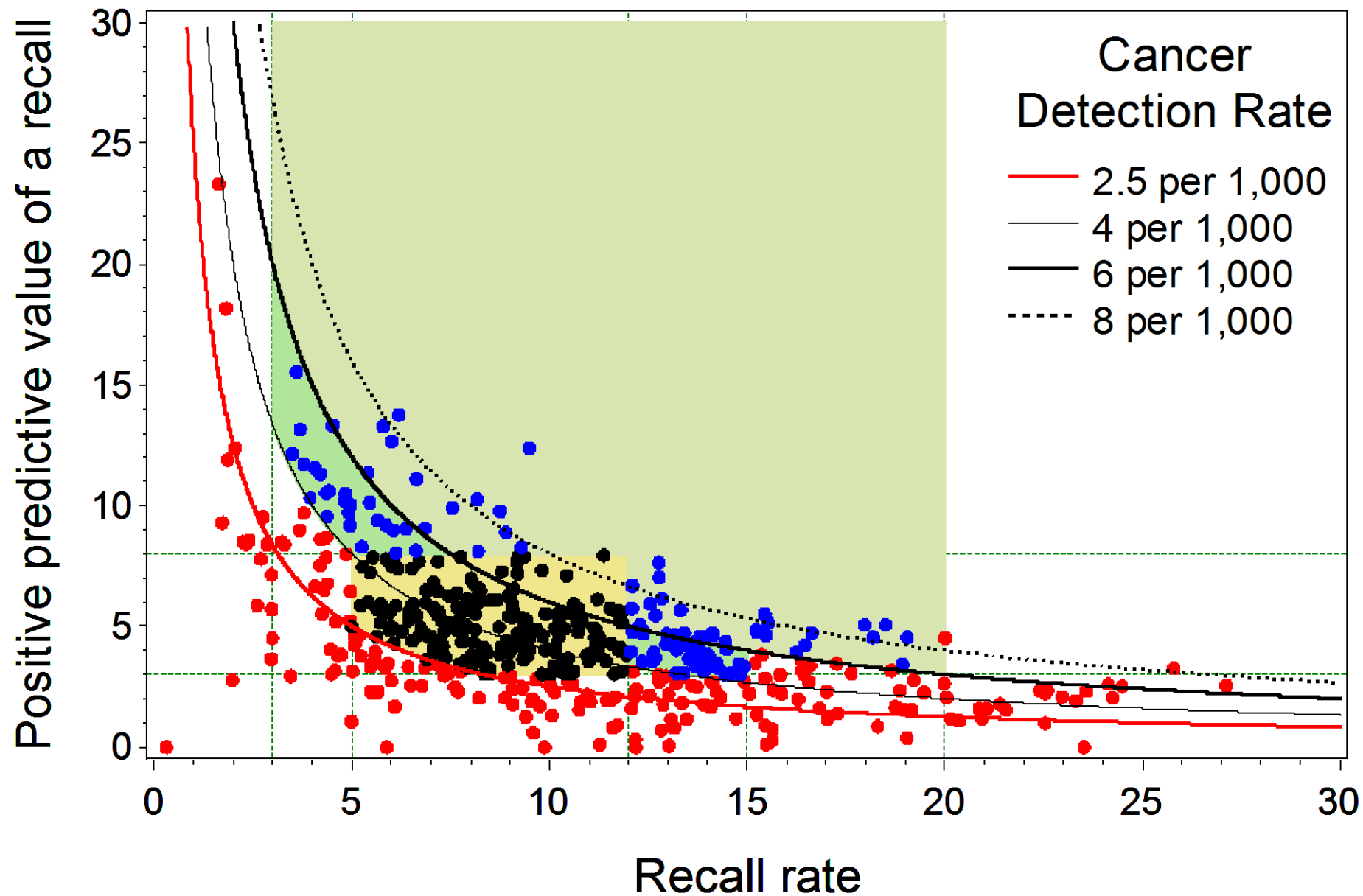*For radiologists who only have complete cancer capture for positive mammograms*

Criteria for identifying low performance

Measure	Low Performance Range	% of BCSC Radiologists in Low Performance Range	
Sensitivity	<75%	18%	
Specificity	<88% or >95%	46%	
Recall Rate	<5% or >12%	49%	
PPV1	<3% or >8%	38%	**60%**
CDR	<2.5/1000	28%	

Carney, et al. *Radiology*. 2010;255(2):354-61.

Cancer Detection Rate

— 2.5 per 1,000
— 4 per 1,000
— 6 per 1,000
---- 8 per 1,000

Miglioretti, *et al.*; *AJR.* 2015;204(4):W486-91.

Minimally acceptable criteria for radiologists with complete cancer capture of positive exams only

Criteria	CDR		Recall Rate		PPV 1		% of BCSC Radiologists who Met Criteria
Original	≥2.5/1000		5-12%		3-8		40%
Updated Criteria 1	≥6/1000	&	3-20%				13%
Updated Criteria 2	≥4-<6/1000	&	3-15%	&	≥3		31%
Updated Criteria 3	2.5-<4/1000	&	5-12%	&	3-8		18%

62%

Miglioretti, *et al.*; *AJR.* 2015;204(4):W486-91.

Are Radiologists Meeting Targets?

Challenge: Observed performance is often based on small sample of rare events.

Confidence Interval Approach

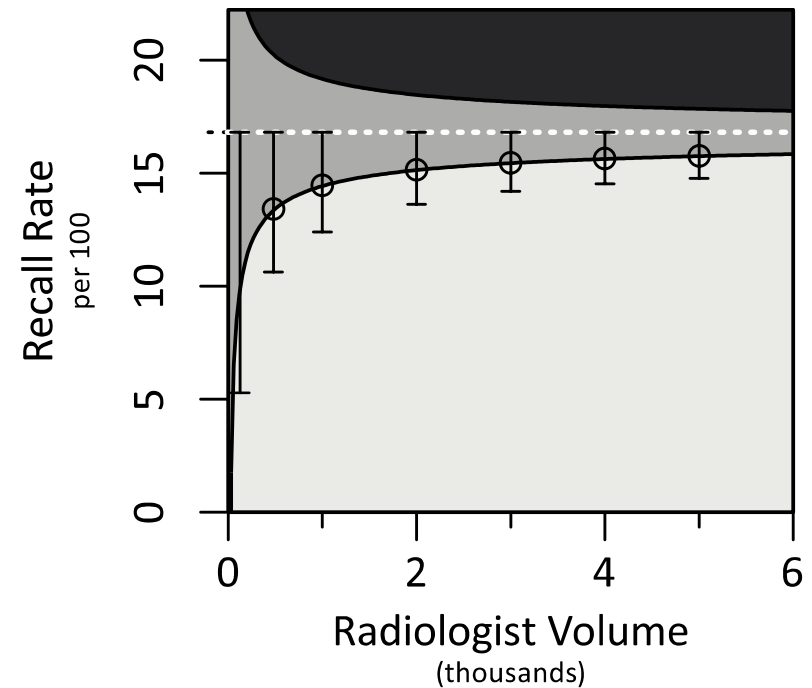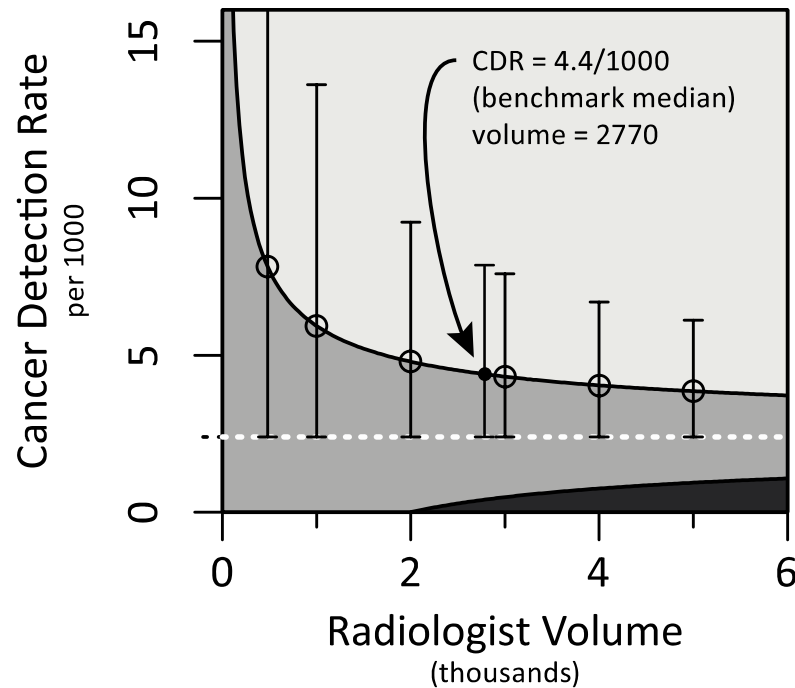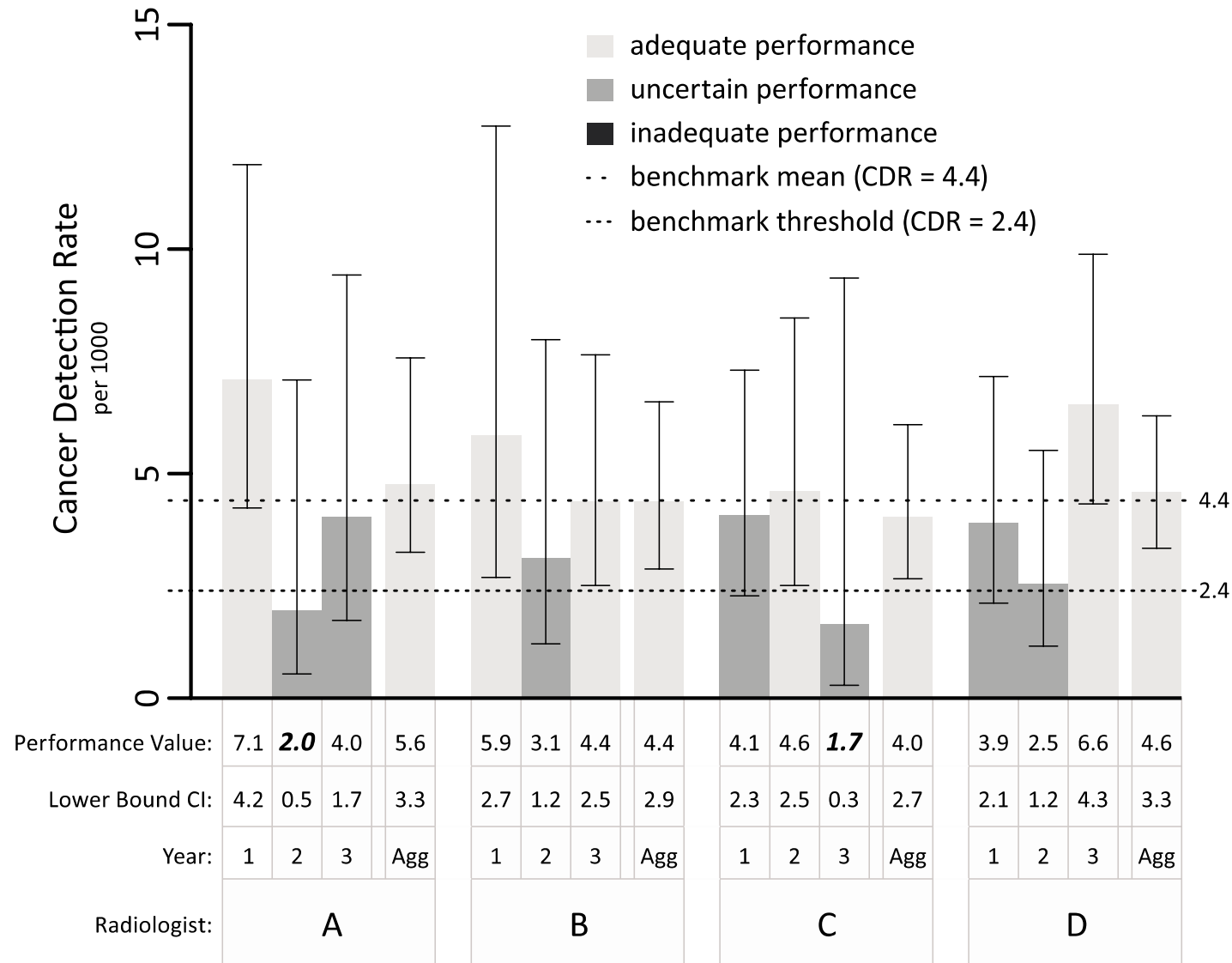- **Compute confidence interval around observed performance**

- **Classify radiologists into three *zones***

  - **<u>Adequate</u>: If CI lies completely within acceptable zone**

  - **<u>Inadequate</u>: If CI lies completely outside acceptable zone**

  - **<u>Uncertain</u>: If CI is both within and outside acceptable zone**

Regions of *adequate*, *inadequate*, and *uncertain* performance by volume

CDR = 4.4/1000
(benchmark median)
volume = 2770

Burnside ES, *et al*. PLoS One. 2014;9(2):e89418.

Regions of *adequate*, *inadequate*, and *uncertain* performance by volume

uncertain

uncertain

adequate

Burnside ES, *et al*. PLoS One. 2014;9(2):e89418.

Annual observed performance values as compared to aggregated data

Burnside ES, *et al*. PLoS One. 2014;9(2):e89418.

Discussion

- Combined criteria overcome some limitations of prior criteria

- Large volumes are needed to confidently assess cancer detection rate

- CI approach is a simple approach

  - Could be extended to adjust for case-mix and consider combined criteria

  - Can adjust confidence level

  - May be conservative - confidence intervals are often wide due to small numbers or rare events

References

- Carney PA, Sickles EA, Monsees BS, Bassett LW, Brenner RJ, Feig SA, Smith RA, Rosenberg RD, Bogart TA, Browning S, Barry JW, Kelly MM, Tran KA, Miglioretti DL. Identifying minimally acceptable interpretive performance criteria for screening mammography. *Radiology*. 2010;255(2):354-61.

- Miglioretti DL, Ichikawa L, Smith RA, Bassett LW, Feig SA, Monsees B, Parikh JR, Rosenberg RD, Sickles EA, Carney PA. Criteria for identifying radiologists with acceptable screening mammography interpretive performance on basis of multiple performance measures. *AJR Am J Roentgenol.* 2015;204(4):W486-91.

- Burnside ES, Lin Y, Munoz del Rio A, Pickhardt PJ, Wu Y, Strigel RM, Elezaby MA, Kerr EA, Miglioretti DL. Addressing the challenge of assessing physician-level screening performance: mammography as an example*. PLoS One.* 2014;9(2):e89418.