Predictive Analytics for Optimal Detection of Metastatic Prostate Cancer

Brian Denton

Department of Industrial and Operations Engineering
University of Michigan

May 16, 2018

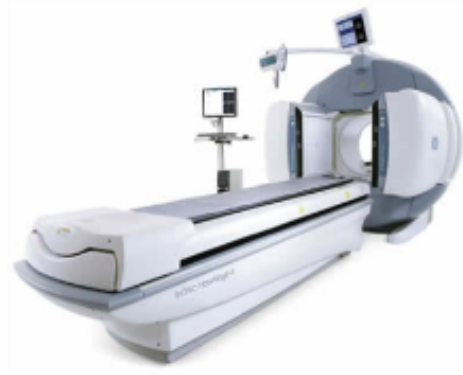
Imaging is used to detect bone and lymph node metastasis

Bone Scan (BS)

- Time-consuming procedure ($3 - 4$ hours)
- Costs $\$600 - \$1,000$

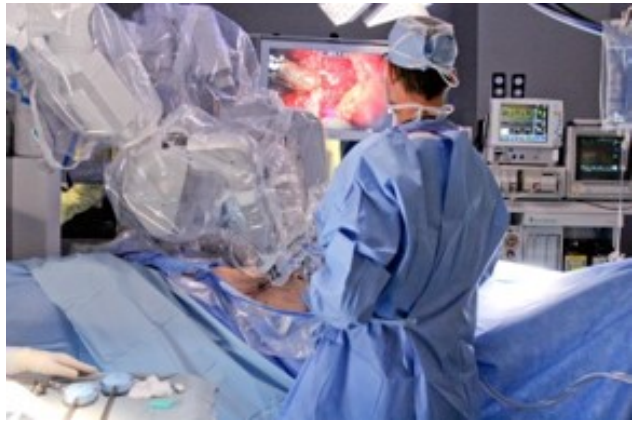
Computed Tomography (CT)

- Exposes patient to 60 times more radiation than an x-ray
- Costs $\$300 - \$1,500$



Significant harms associated with missing a case of metastatic cancer

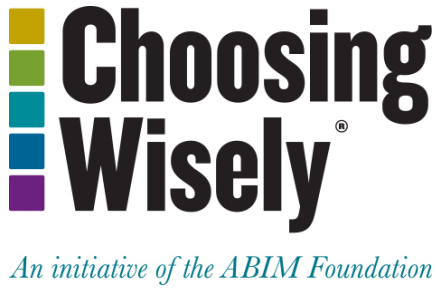
- Negative health outcomes due to delays in chemotherapy



- Missed diagnoses subject patients to unnecessary treatments (e.g., radical prostatectomy) that cause serious side effects

Avoidance of imaging in low-risk PCa top priority for AUA "Choosing Wisely" campaign

- Potentially harmful radiation exposure



- Incidental findings that require painful and risky follow-up procedures (e.g., bone biopsy)
- Blocks access to imaging resources for other patients and unnecessarily increases healthcare costs

Conflicting imaging guidelines for PCa staging

- European Urological Association (EAU)
- American Urological Association (AUA)
- National Comprehensive Cancer Network (NCCN)
- Briganti's classification and regression tree (CART)*

*A. Briganti, N. Passoni et al. "When to perform bone scan in newly diagnosed prostate cancer: External validation of the currently available guidelines and proposal of a novel risk stratification tool" *European Urology*, 57 (4), 2010.

Research objective

To determine which patients should receive a BS and/or a CT scan and which patients can safely avoid imaging on the basis of individual risk factors.

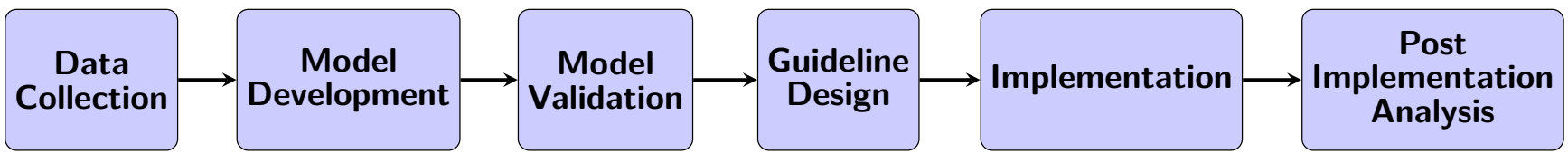
Michigan Urological Surgery Improvement Collaborative

- Physician-led, statewide collaborative
- 43 urology practices from throughout Michigan ($> 90\%$ of urologists in the state)
- Complete preoperative data for men with newly-diagnosed PCa were retrospectively reviewed

Methodology

The methodological approach consists of

- Development and validation of predictive models
- Correction for the bias due to missing data
- Classification modeling for the detection of metastasis



Risk Prediction Models for Metastatic Prostate Cancer

Can we develop predictive models that are-well calibrated to provide reliable predictions for newly-diagnosed MUSIC patients?

Predictive modeling

Multivariate logistic regression determines the probability of a positive BS and CT scan as a function of several covariates:

- Age
- Prostate-specific antigen (PSA) (ng/ml)
- Clinical tumor stage (e.g., T1a/b/c, T2a/b/c and T3/4)
- Gleason score (GS)
- Percentage of biopsy positive cores over total number of cores taken

Development sample characteristics

- For BS:
  - March 2012 - June 2013
  - 416 patients received BS, 48 (11.5%) were positive
- For CT scan:
  - March 2012 - September 2013
  - 643 patients received CT scan, 62 (9.6%) were positive

Predictors of metastatic disease

	BS Univariable logistic regression model		**CT scan Univariable logistic regression model**	
Variables	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Age at diagnosis (year)	$1.04\ (1.01-1.08)$	0.01	$1.02\ (0.99-1.05)$	0.02
ln(PSA+1), ng/mL	$2.25\ (1.76-2.88)$	<0.0001	$2.79\ (2.21-3.54)$	<0.0001
Biopsy Gleason score, No. (%)				
$\leq 3+4$	Reference		Reference	
$4+3$	$5.04\ (0.90-28.31)$	0.07	$15.49\ (1.84-130.48)$	0.01
$8-10$	$16.05\ (3.82-67.45)$	0.0002	$50.69\ (6.96-369.16)$	<0.0001
Clinical T stage, No. (%)				
T1	Reference		Reference	
T2	$2.64\ (1.31-5.33)$	0.007	$2.05\ (1.09-3.86)$	0.03
T3/4	$9.19\ (3.51-24.03)$	<0.0001	$21.05\ (9.52-46.56)$	<0.0001
Positive cores, %	$13.32\ (4.26-41.72)$	<0.0001	$35.08\ (12.06-102.03)$	<0.0001

Statistical validation

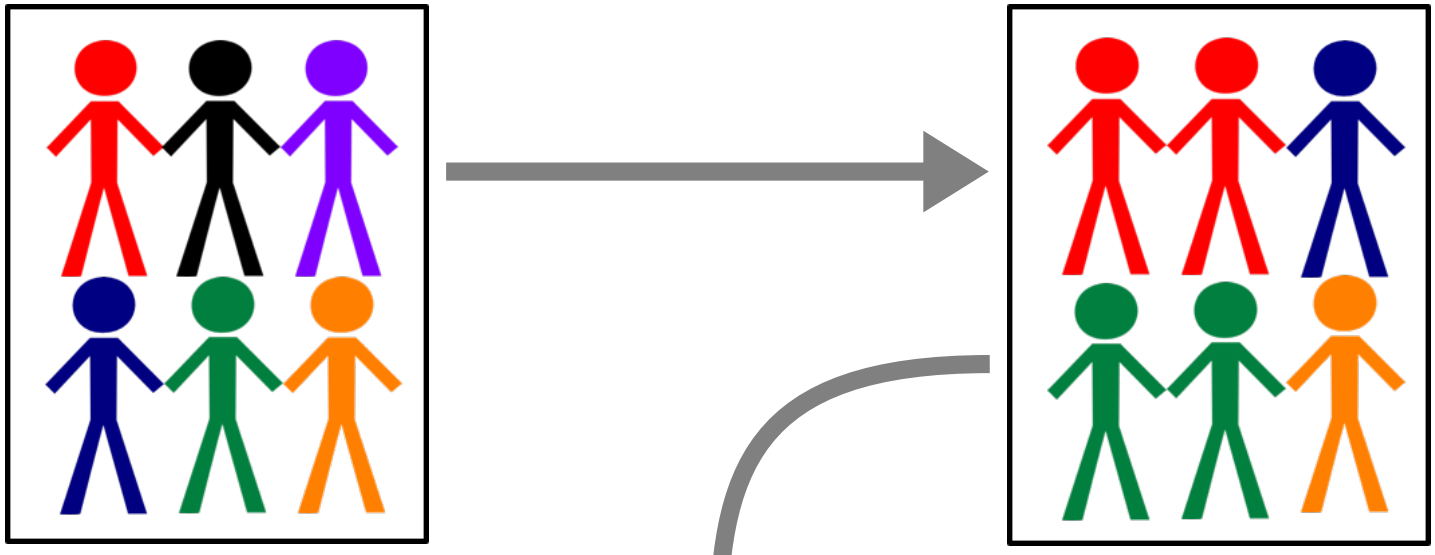
Internal validation

- Boot strapping using the development sample to estimate optimism

External validation

- Independent datasets were used to validate the predictive models

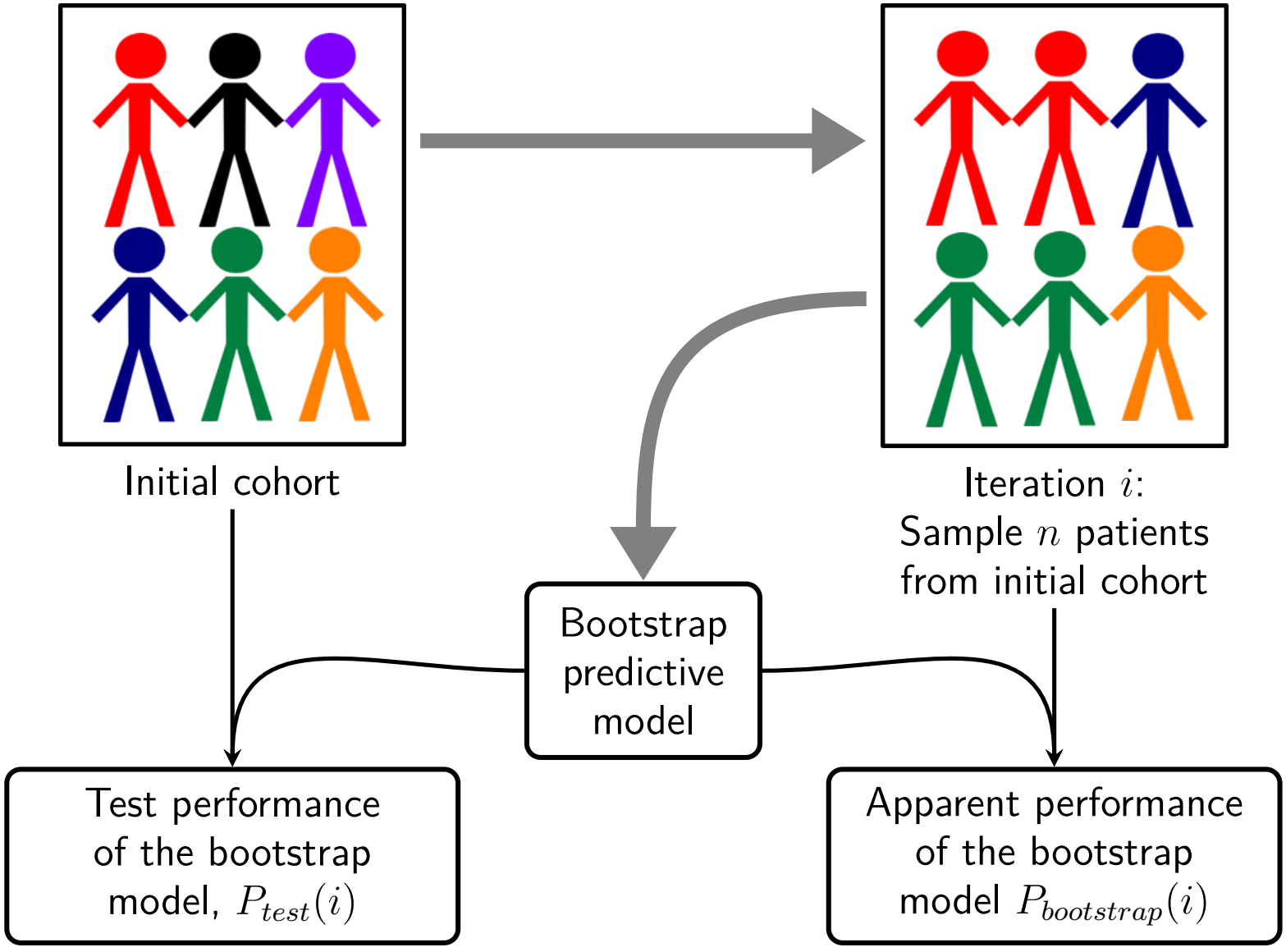
s



Initial cohort

Iteration i:
Sample n patients
from initial cohort

Bootstrap
predictive
model



Estimate the optimism:

$$o(i) = P_{bootstrap}(i) - P_{test}(i)$$

Using bootstrapping to correct for optimism bias

- After m iterations of bootstrapping, we can estimate the expected optimism:

$$\text{Optimism} = \frac{\sum_{i=1}^{m} o(i)}{m}$$

- This optimism estimate can update the apparent performance of our model:

$$P_{validated} = P_{apparent} - \text{Optimism}$$

Performance measures

- **Discrimination**
  - How well can the model differentiate between patients with positive and negative imaging results?
- **Calibration**
  - How reliable are the predicted risks?

Calibration

- Calibration slope is equal to one in the development sample
- In an external dataset, the calibration slope ($\beta_{calibration}$) is estimated using a logistic regression model with the linear predictor as the only explanatory variable:

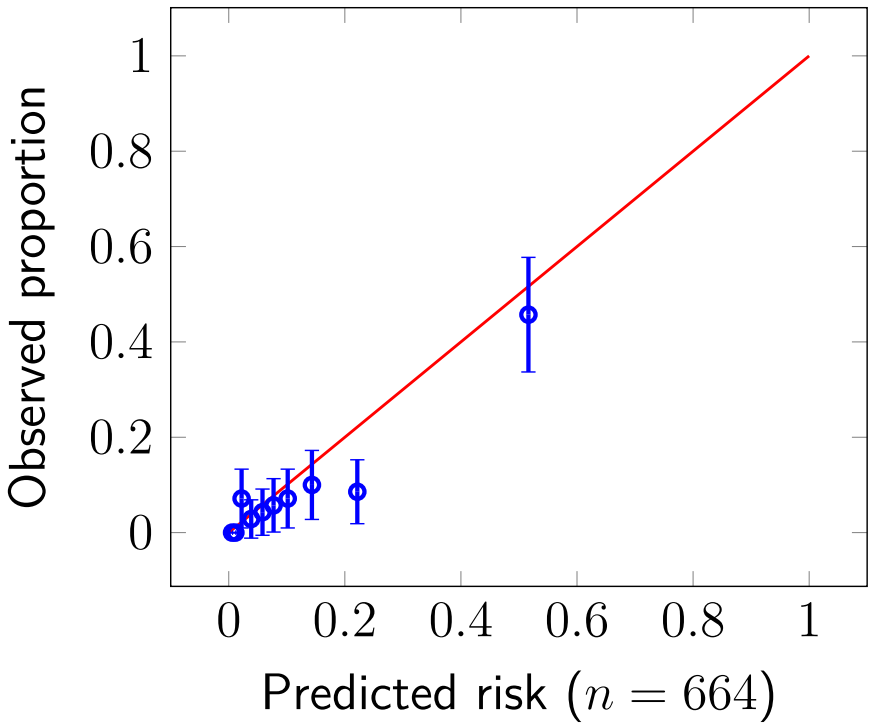
$$\log \frac{\mathbb{P}(\text{Disease present})}{\mathbb{P}(\text{Disease not present})} = \alpha + \beta_{calibration}\, \text{LP}_i$$

Agreement between internal and external validation estimates

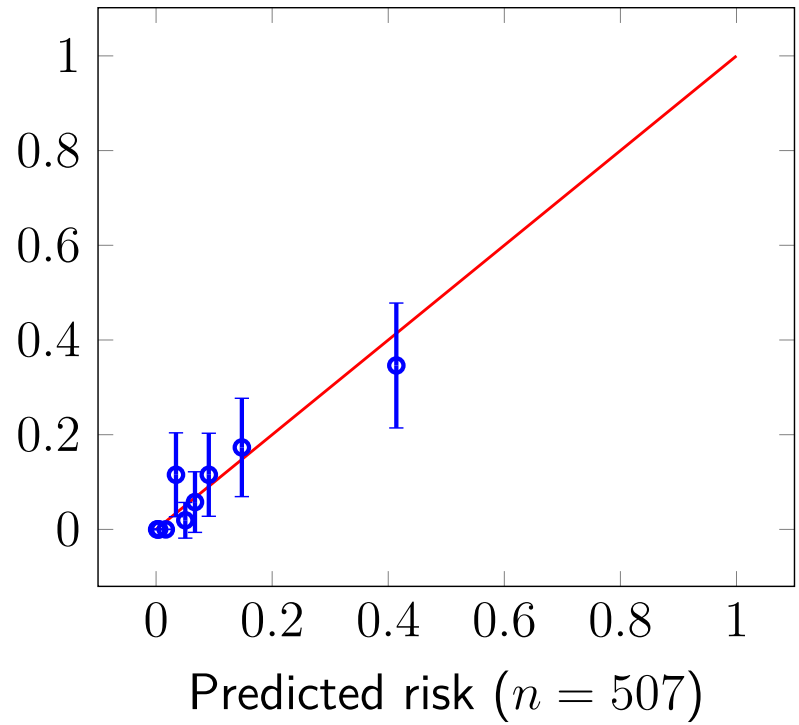
	Internal Validation		**External Validation**	
	Bone scan $(\mathbf{n = 416})$	**CT scan** $(\mathbf{n = 643})$	**Bone scan** $(\mathbf{n = 664})$	**CT scan** $(\mathbf{n = 507})$
ROC area	0.82	0.87	0.81	0.86
Brier score	0.080	0.060	0.068	0.061
Calibration slope	0.86	0.90	0.99	0.94

Performance measures were found by applying the predictive models fit in the development samples to the external validation samples.

Predictive models are well-calibrated to external data sets



BS model



CT scan model

Bias-corrected Performance of Imaging Guidelines

How can we account for the systematic bias as not all men with newly-diagnosed prostate cancer received imaging?

Verification bias

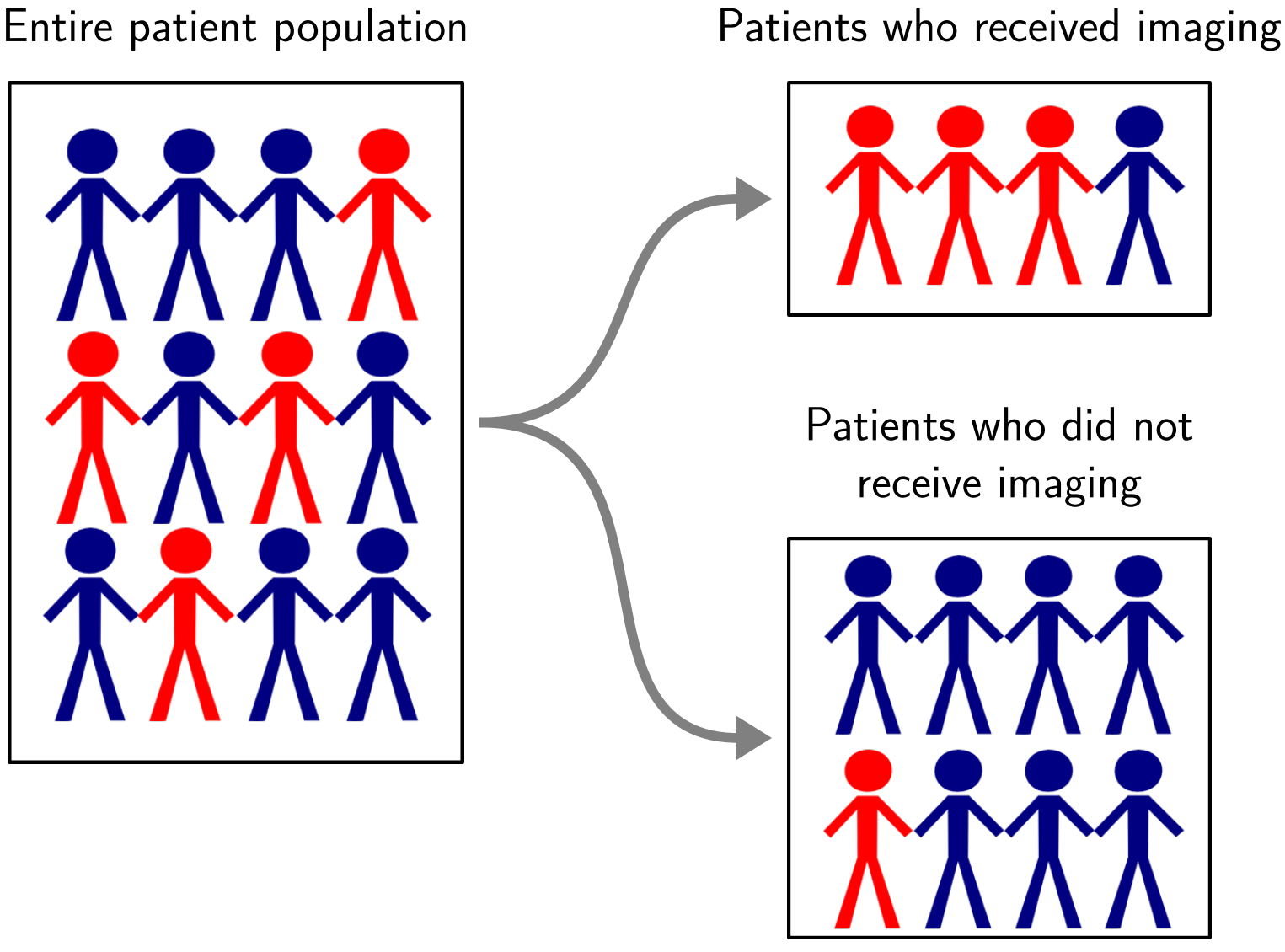
- G+ and G- indicate whether a patient is recommended to receive imaging or not based on guideline G

- Unadjusted sensitivity and specificity are estimated based only on patients who received imaging tests:

$$\begin{aligned} \text{Sensitivity} &= \mathbb{P}(\text{G}+ \mid \text{Disease present}) \\ \text{Specificity} &= \mathbb{P}(\text{G}- \mid \text{Disease not present}) \end{aligned}$$

- Not all patients in our cohort received imaging, which leads to verification bias

Verification bias



Verification bias correction

We used our logistic regression model to estimate sensitivity and specificity based on the entire patient population:

$$\mathbb{P}(\text{G}+ \mid \text{Disease present}) = \frac{\mathbb{P}(\text{Disease present} \mid \text{G}+)\mathbb{P}(\text{G}+)}{\mathbb{P}(\text{Disease present})}$$

$$\mathbb{P}(\text{G}- \mid \text{Disease not present}) = \frac{\mathbb{P}(\text{Disease not present} \mid \text{G}-)\mathbb{P}(\text{G}-)}{\mathbb{P}(\text{Disease not present})}$$

Main assumption: factors considered by the guideline are the only factors that influence imaging decisions

Begg, C. B., Greenes, R. A. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics*, 39:207, 1983.

Verification bias greatly impacts apparent performance of imaging guidelines

	Uncorrected		Bias-corrected	
Clinical guidelines	**Sensitivity**	**Specificity**	**Sensitivity**	**Specificity**
Bone scan				
EAU	97.9	33.4	84.5	75.7
AUA	**97.9**	**43.5**	**81.2**	**82.0**
NCCN	97.9	40.8	82.3	80.9
Briganti's CART	89.6	45.4	79.3	83.3
CT scan				
EAU	98.4	36.5	89.9	74.4
AUA	**96.8**	**49.2**	**87.2**	**82.5**

The numbers are the percentages. EAU: European Urological Association; AUA: American Urological Association; NCCN: National Comprehensive Cancer Network; CART: classification and regression tree.

Classification Modeling for Metastatic Cancer Detection

Can we design imaging guidelines using machine learning methods that can outperform the published guidelines?

Classification models

- Two important challenges:
  - Learning from unlabeled data
    - In practice not all patients receive imaging at diagnosis
  - Learning from imbalanced data
    - A minority of patients has metastatic cancer
- To address these challenges, two machine learning paradigms are combined:
  - Semi-supervised learning
  - Cost-sensitive learning

Cost-sensitive Laplacian Kernel Logistic Regression (Cos-LapKLR)

$$f^* = \underset{f \in \mathcal{H}}{\operatorname{argmin}} \frac{1}{l} \sum_{i=1}^{l} \delta \mathbb{1}_{\{y_i=1\}} \log\left(1 + e^{-f(\mathbf{x}_i)}\right) + (1-\delta)\mathbb{1}_{\{y_i=-1\}} \log\left(1 + e^{f(\mathbf{x}_i)}\right)$$

$$+ \quad \gamma_{\mathcal{H}} \|f\|^2_{\mathcal{H}} \quad + \quad \gamma_{\mathcal{M}} \mathbf{f^T L f}$$

where f is the decision function, $f^*(\mathbf{x}) = \sum_{i=1}^{l+u} \alpha_i^* \mathbf{K}(\mathbf{x}_i, \mathbf{x})$, u the number of unimaged patients, $\mathbf{K}$ the positive definite kernel function and $\mathbf{L}$ the Laplacian matrix.

Cost-sensitive Laplacian Kernel Logistic Regression (Cos-LapKLR)

Higher cost on missing metastatic cancers

$$f^* = \underset{f \in \mathcal{H}}{\operatorname{argmin}} \frac{1}{l} \sum_{i=1}^{l} \delta \mathbb{1}_{\{y_i=1\}} \log\left(1 + e^{-f(\mathbf{x}_i)}\right) + (1-\delta)\mathbb{1}_{\{y_i=-1\}} \log\left(1 + e^{f(\mathbf{x}_i)}\right)$$

$$+ \quad \gamma_{\mathcal{H}} \|f\|^2_{\mathcal{H}} \quad + \quad \gamma_{\mathcal{M}} \mathbf{f^T L f}$$

where f is the decision function, $f^*(\mathbf{x}) = \sum_{i=1}^{l+u} \alpha_i^* \mathbf{K}(\mathbf{x}_i, \mathbf{x})$, u the number of unimaged patients, $\mathbf{K}$ the positive definite kernel function and $\mathbf{L}$ the Laplacian matrix.

Cost-sensitive Laplacian Kernel Logistic Regression (Cos-LapKLR)

Higher cost on missing metastatic cancers

$$f^* = \underset{f \in \mathcal{H}}{\operatorname{argmin}} \frac{1}{l} \sum_{i=1}^{l} \delta \mathbb{1}_{\{y_i=1\}} \log\left(1 + e^{-f(\mathbf{x}_i)}\right) + (1-\delta)\mathbb{1}_{\{y_i=-1\}} \log\left(1 + e^{f(\mathbf{x}_i)}\right)$$

$$+ \gamma_{\mathcal{H}} \|f\|^2_{\mathcal{H}} + \gamma_{\mathcal{M}} \mathbf{f^T L f}$$

Avoid overfitting

where f is the decision function, $f^*(\mathbf{x}) = \sum_{i=1}^{l+u} \alpha_i^* \mathbf{K}(\mathbf{x}_i, \mathbf{x})$, u the number of unimaged patients, $\mathbf{K}$ the positive definite kernel function and $\mathbf{L}$ the Laplacian matrix.

Cost-sensitive Laplacian Kernel Logistic Regression (Cos-LapKLR)

Higher cost on missing metastatic cancers

$$f^* = \underset{f \in \mathcal{H}}{\operatorname{argmin}} \frac{1}{l} \sum_{i=1}^{l} \delta \mathbb{1}_{\{y_i=1\}} \log\left(1 + e^{-f(\mathbf{x}_i)}\right) + (1-\delta)\mathbb{1}_{\{y_i=-1\}} \log\left(1 + e^{f(\mathbf{x}_i)}\right)$$

$$+ \gamma_{\mathcal{H}} \|f\|^2_{\mathcal{H}} + \gamma_{\mathcal{M}} \mathbf{f^T L f}$$

Avoid overfitting

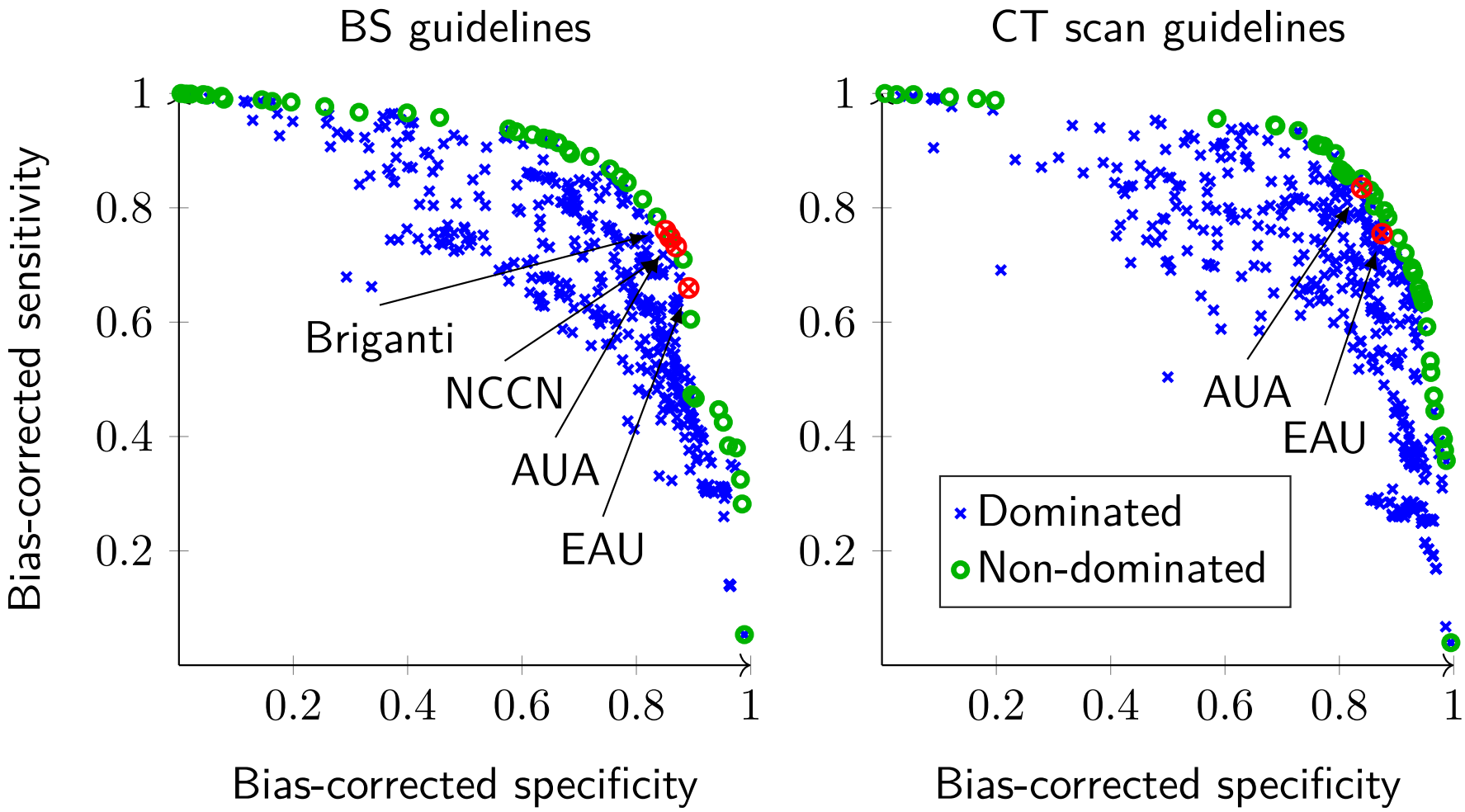
Extract information from unimaged patients

where f is the decision function, $f^*(\mathbf{x}) = \sum_{i=1}^{l+u} \alpha_i^* \mathbf{K}(\mathbf{x}_i, \mathbf{x})$, u the number of unimaged patients, $\mathbf{K}$ the positive definite kernel function and $\mathbf{L}$ the Laplacian matrix.

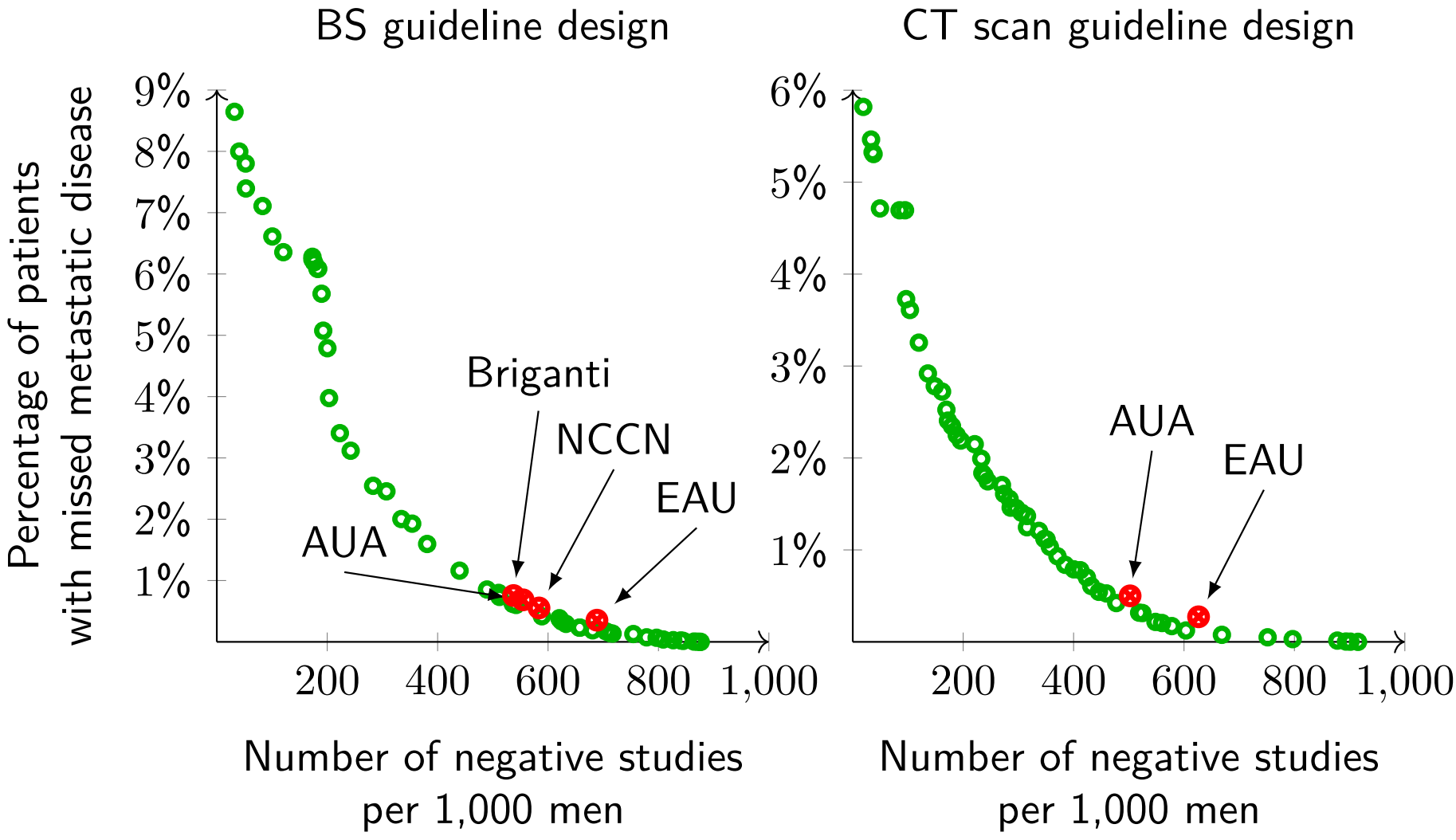
Alternative classification models

- Several other classification models adapted for imbalanced data learning were implemented:
  - Cost-sensitive logistic regression and support vector machines
  - Random forests and AdaBoost combined with advanced resampling techniques
- The diagnostic accuracy of guidelines developed from classification models was corrected for verification bias

Published guidelines are near-Pareto optimal



Published guidelines are close to the efficient frontier with missed metastasis rate $< 1\%$



Impact of recommendations if implemented

Bone scan PSA > 20 or GS > 7	CT scan PSA > 20 or GS > 7 or clinical T stage $\geq$ T3
■ $\mathbf{20.7\%}$ (prev $\mathbf{27\%}$) of patients would be scanned – Of those, $\mathbf{17.0\%}$ (prev $\mathbf{12\%}$) would be positive ■ Estimated $\mathbf{0.8\%}$ of patients have missed metastatic disease ■ $\mathbf{38\%}$ negative scans would be avoided	■ $\mathbf{22.6\%}$ (prev $\mathbf{27\%}$) of patients would be scanned – Of those, $\mathbf{14.3\%}$ (prev $\mathbf{10\%}$) would be positive ■ Estimated $\mathbf{0.4\%}$ of patients have missed metastatic disease ■ $\mathbf{44\%}$ negative scans would be avoided

MUSIC Imaging Appropriateness Criteria instituted across Michigan

- Statewide goal of performing imaging in $\geq 95\%$ of patients that meet the criteria and in $< 10\%$ of those that do not
- MUSIC members were provided with a toolkit including placards with the criteria and explanations for patients

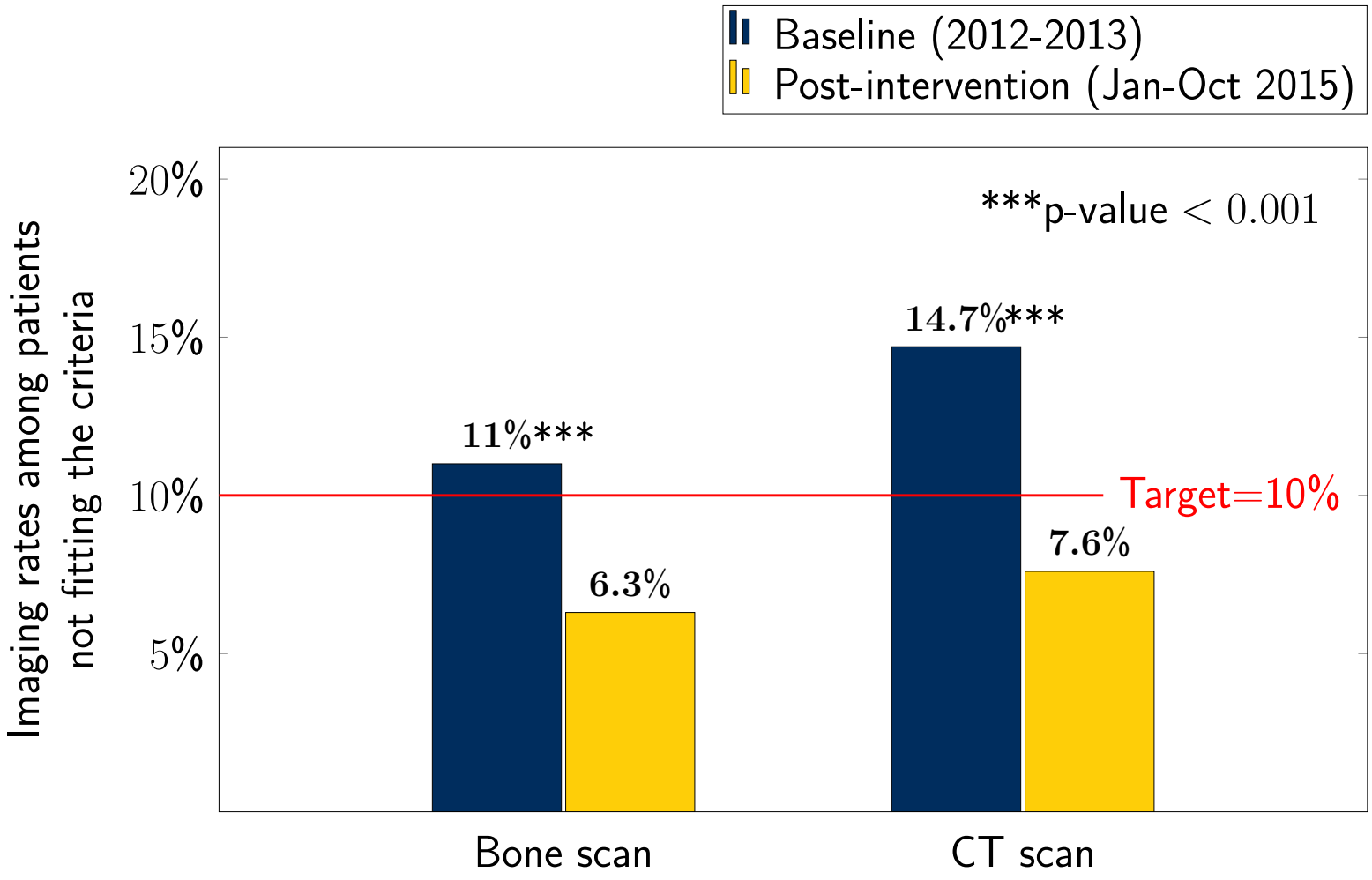
MUSIC Imaging Appropriateness Criteria

	Bone Scan	CT Scan
	Order Bone Scan If:	Order Bone Scan If:
PSA	> 20	> 20
	OR	*OR*
Gleason	≥ 8	≥ 8
		OR
Clinical T Stage		≥ cT3

Imaging Goals

- Perform imaging in ≥ 95% of patients meeting criteria
- Perform imaging in < 10% of patient NOT meeting criteria

MUSIC achieved state-wide decrease in utilization of unnecessary imaging tests



Project outcomes

- Reduction in harm to patient health from reduced radiation exposure, fewer unnecessary follow-up procedures, and decreased patient anxiety
- MUSIC collaborative saved more than $262,000 in 2015 through reducing unnecessary imaging tests
- AskMUSIC web tool for predictive models built from MUSIC data

Project outcomes

Selin Merdan, Christine Barnett, David C. Miller, James E. Montie, Brian T. Denton. Data Analytics for Optimal Detection of Metastatic Prostate Cancer, (preprint available at `http://www-personal.umich.edu/~smerdan/`).

Selin Merdan, Paul R. Womble, David C. Miller, Christine Barnett, Zhu Ye, Susan M. Linsell, James E. Montie, Brian T. Denton. Toward better use of bone scans among men with early-stage prostate cancer. *Urology*.

Rachel Risko, **Selin Merdan**, Paul R. Womble, Christine Barnett, Zhu Ye, Susan M. Linsell, James E. Montie, David C. Miller, Brian T. Denton. Clinical predictors and recommendations for staging computed tomography scan among men with prostate cancer. *Urology*.

Acknowledgments

Selin Merdan, PhD

Christine Barnett, PhD

Susan Linsell, MSHA

Karandeep Singh, MD

James E. Montie, MD

David C. Miller, MD

Rachel Risko, BSE

Michigan Urological Surgery Improvement Collaborative

Thank you.

Brian Denton
Department of Industrial and Operations Engineering and Department of Urology
University of Michigan, Ann Arbor, MI
E-mail: btdenton@umich.edu
Website: https://btdenton.engin.umich.edu/

Sensitivity and specificity of imaging tests

- Bone scan
  - Sensitivity of 86%
  - Specificity of 81%
- CT scan
  - Sensitivity of 42%
  - Specificity of 82%

State-wide increase in the utilization of imaging tests for high-risk patients

- An increase in the use of imaging tests in patients that meet the criteria from
  - 82% to 84% for BS
  - 74% to 77% for CT scan
- The MUSIC consortium has made measurable improvements in a short period of time and additional increases are anticipated

Compliance with MUSIC Imaging Criteria

- Our results were presented at collaborative-wide meetings with clinical champions who returned to their practices to present the results to their own practice group
- MUSIC members were provided with a toolkit including placards with the criteria and explanations for patients
- Members received comparative performance feedback that detailed how well their practice patterns correlated with the MUSIC Imaging Criteria

Bias correction

$$\mathbb{P}(\mathrm{G}+ \mid \text{Disease present}) = \frac{\mathbb{P}(\text{Disease present} \mid \mathrm{G}+)\mathbb{P}(\mathrm{G}+)}{\mathbb{P}(\text{Disease present})}$$

$$\mathbb{P}(\mathrm{G}- \mid \text{Disease not present}) = \frac{\mathbb{P}(\text{Disease not present} \mid \mathrm{G}-)\mathbb{P}(\mathrm{G}-)}{\mathbb{P}(\text{Disease not present})}$$

Bias correction

$$\mathbb{P}(\text{G}+ \mid \text{Disease present}) = \frac{\mathbb{P}(\text{Disease present} \mid \text{G}+)\mathbb{P}(\text{G}+)}{\mathbb{P}(\text{Disease present})}$$

$$\mathbb{P}(\text{G}- \mid \text{Disease not present}) = \frac{\mathbb{P}(\text{Disease not present} \mid \text{G}-)\mathbb{P}(\text{G}-)}{\mathbb{P}(\text{Disease not present})}$$

- Separate the entire population into G+ and G−
- Develop logistic regression model among patients who received imaging to estimate the probability of metastatic disease for every patient

Bias correction

$$\mathbb{P}(\text{G}+ \mid \text{Disease present}) = \frac{\mathbb{P}(\text{Disease present} \mid \text{G}+)\mathbb{P}(\text{G}+)}{\mathbb{P}(\text{Disease present})}$$

$$\mathbb{P}(\text{G}- \mid \text{Disease not present}) = \frac{\mathbb{P}(\text{Disease not present} \mid \text{G}-)\mathbb{P}(\text{G}-)}{\mathbb{P}(\text{Disease not present})}$$

Bias correction

$$\mathbb{P}(\text{G}+ \mid \text{Disease present}) = \frac{\mathbb{P}(\text{Disease present} \mid \text{G}+)\mathbb{P}(\text{G}+)}{\mathbb{P}(\text{Disease present})}$$

$$\mathbb{P}(\text{G}- \mid \text{Disease not present}) = \frac{\mathbb{P}(\text{Disease not present} \mid \text{G}-)\mathbb{P}(\text{G}-)}{\mathbb{P}(\text{Disease not present})}$$

$$\begin{aligned}\mathbb{P}(\text{Disease present}) &= \mathbb{P}(\text{Disease present}|\text{G}+)\mathbb{P}(\text{G}+)+ \\ &\quad \mathbb{P}(\text{Disease present}|\text{G}-)\mathbb{P}(\text{G}-)\end{aligned}$$

$$\begin{aligned}\mathbb{P}(\text{Disease not present}) &= \mathbb{P}(\text{Disease not present}|\text{G}+)\mathbb{P}(\text{G}+)+ \\ &\quad \mathbb{P}(\text{Disease not present}|\text{G}-)\mathbb{P}(\text{G}-)\end{aligned}$$