Predictive Analytics for Optimal Detection of Metastatic Prostate Cancer

Brian Denton

Department of Industrial and Operations Engineering  
University of Michigan

Feb 21, 2019
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)*

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

<table>
<thead>
<tr>
<th>Variables</th>
<th>BS Univariable logistic regression model</th>
<th>CT scan Univariable logistic regression model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at diagnosis (year)</td>
<td>1.04 (1.01 – 1.08) p = 0.01</td>
<td>1.02 (0.99 – 1.05) p = 0.02</td>
</tr>
<tr>
<td>ln(PSA+1), ng/mL</td>
<td>2.25 (1.76 – 2.88) p &lt; 0.0001</td>
<td>2.79 (2.21 – 3.54) p &lt; 0.0001</td>
</tr>
<tr>
<td>Biopsy Gleason score, No. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 3 + 4</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>4 + 3</td>
<td>5.04 (0.90 – 28.31) p = 0.07</td>
<td>15.49 (1.84 – 130.48) p = 0.01</td>
</tr>
<tr>
<td>8 – 10</td>
<td>16.05 (3.82 – 67.45) p = 0.0002</td>
<td>50.69 (6.96 – 369.16) p &lt; 0.0001</td>
</tr>
<tr>
<td>Clinical T stage, No. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>T2</td>
<td>2.64 (1.31 – 5.33) p = 0.007</td>
<td>2.05 (1.09 – 3.86) p = 0.03</td>
</tr>
<tr>
<td>T3/4</td>
<td>9.19 (3.51 – 24.03) p &lt; 0.0001</td>
<td>21.05 (9.52 – 46.56) p &lt; 0.0001</td>
</tr>
<tr>
<td>Positive cores, %</td>
<td>13.32 (4.26 – 41.72) p &lt; 0.0001</td>
<td>35.08 (12.06 – 102.03) p &lt; 0.0001</td>
</tr>
</tbody>
</table>
Statistical validation

**Internal validation**

- Boot strapping using the development sample to estimate optimism

**External validation**

- Independent datasets were used to validate the predictive models
Initial cohort

Iteration $i$:
Sample $n$ patients from initial cohort

Bootstrap predictive model

Estimate the optimism:
$o(i) = P_{\text{bootstrap}}(i) - P_{\text{test}}(i)$
Initial cohort

Iteration $i$: Sample $n$ patients from initial cohort

Bootstrap predictive model

Test performance of the bootstrap model, $P_{test}(i)$

Apparent performance of the bootstrap model $P_{bootstrap}(i)$

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:

$$P_{\text{validated}} = P_{\text{apparent}} - \text{Optimism}$$
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 ($\beta_{\text{calibration}}$) is estimated using a logistic regression model with the linear predictor as the only explanatory variable:

\[
\log \left( \frac{P(\text{Disease present})}{P(\text{Disease not present})} \right) = \alpha + \beta_{\text{calibration}} \text{LP}_i
\]
Agreement between internal and external validation estimates

<table>
<thead>
<tr>
<th></th>
<th>Internal Validation</th>
<th></th>
<th>External Validation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bone scan</td>
<td>CT scan</td>
<td>Bone scan</td>
<td>CT scan</td>
</tr>
<tr>
<td>(n = 416)</td>
<td>(n = 643)</td>
<td></td>
<td>(n = 664)</td>
<td>(n = 507)</td>
</tr>
<tr>
<td>ROC area</td>
<td>0.82</td>
<td>0.87</td>
<td>0.81</td>
<td>0.86</td>
</tr>
<tr>
<td>Brier score</td>
<td>0.080</td>
<td>0.060</td>
<td>0.068</td>
<td>0.061</td>
</tr>
<tr>
<td>Calibration slope</td>
<td>0.86</td>
<td>0.90</td>
<td>0.99</td>
<td>0.94</td>
</tr>
</tbody>
</table>

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:

\[
\text{Sensitivity} = \mathbb{P}(G+ \mid \text{Disease present})
\]
\[
\text{Specificity} = \mathbb{P}(G- \mid \text{Disease not present})
\]

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

Entire patient population

Patients who received imaging

Patients who did not receive imaging
Verification bias correction

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

\[
P(G+ \mid \text{Disease present}) = \frac{P(\text{Disease present} \mid G+)P(G+)}{P(\text{Disease present})}
\]

\[
P(G- \mid \text{Disease not present}) = \frac{P(\text{Disease not present} \mid G-)P(G-)}{P(\text{Disease not present})}
\]

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

Verification bias greatly impacts apparent performance of imaging guidelines

<table>
<thead>
<tr>
<th>Clinical guidelines</th>
<th>Uncorrected</th>
<th>Bias-corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td><strong>Bone scan</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EAU</td>
<td>97.9</td>
<td>33.4</td>
</tr>
<tr>
<td>AUA</td>
<td><strong>97.9</strong></td>
<td><strong>43.5</strong></td>
</tr>
<tr>
<td>NCCN</td>
<td>97.9</td>
<td>40.8</td>
</tr>
<tr>
<td>Briganti’s CART</td>
<td>89.6</td>
<td>45.4</td>
</tr>
<tr>
<td><strong>CT scan</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EAU</td>
<td>98.4</td>
<td>36.5</td>
</tr>
<tr>
<td>AUA</td>
<td><strong>96.8</strong></td>
<td><strong>49.2</strong></td>
</tr>
</tbody>
</table>

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)

\[ f^* = \arg\min_{f \in \mathcal{H}} \frac{1}{l} \sum_{i=1}^{l} \delta \mathbb{1}_{\{y_i=1\}} \log \left( 1 + e^{-f(x_i)} \right) + (1 - \delta) \mathbb{1}_{\{y_i=-1\}} \log \left( 1 + e^{f(x_i)} \right) + \gamma \mathcal{H} \|f\|_{\mathcal{H}}^2 + \gamma \mathcal{M} f^T L f \]

where \( f \) is the decision function, \( f^*(x) = \sum_{i=1}^{l+u} \alpha^*_i K(x_i, x) \), \( u \) the number of unimaged patients, \( K \) the positive definite kernel function and \( L \) the Laplacian matrix.
Cost-sensitive Laplacian Kernel Logistic Regression (Cos-LapKLR)

\[
f^* = \arg\min_{f \in \mathcal{H}} \frac{1}{l} \sum_{i=1}^{l} \delta \mathbb{1}_{\{y_i = 1\}} \log \left(1 + e^{-f(x_i)}\right) + (1 - \delta) \mathbb{1}_{\{y_i = -1\}} \log \left(1 + e^{f(x_i)}\right) + \gamma_{\mathcal{H}} \|f\|_{\mathcal{H}}^2 + \gamma_{\mathcal{M}} f^T L f \]

where \(f\) is the decision function, \(f^*(x) = \sum_{i=1}^{l+u} \alpha_i^* K(x_i, x)\), \(u\) the number of unimaged patients, \(K\) the positive definite kernel function and \(L\) the Laplacian matrix.
Cost-sensitive Laplacian Kernel Logistic Regression (Cos-LapKLR)

\[ f^* = \arg\min_{f \in \mathcal{H}} \frac{1}{l} \sum_{i=1}^{l} \delta \mathbb{1}_{\{y_i=1\}} \log \left( 1 + e^{-f(x_i)} \right) + (1 - \delta) \mathbb{1}_{\{y_i=-1\}} \log \left( 1 + e^{f(x_i)} \right) \]

\[ + \gamma \mathcal{H} \| f \|_{\mathcal{H}}^2 + \gamma \mathcal{M} f^T L f \]

Higher cost on missing metastatic cancers

Avoid overfitting

where \( f \) is the decision function, \( f^*(x) = \sum_{i=1}^{l+u} \alpha_i^* K(x_i, x) \), \( u \) the number of unimaged patients, \( K \) the positive definite kernel function and \( L \) the Laplacian matrix.
Cost-sensitive Laplacian Kernel Logistic Regression (Cos-LapKLR)

\[
f^* = \arg\min_{f \in \mathcal{H}} \frac{1}{l} \sum_{i=1}^{l} \delta \mathbb{1}_{\{y_i=1\}} \log \left(1 + e^{-f(x_i)}\right) + (1 - \delta) \mathbb{1}_{\{y_i=-1\}} \log \left(1 + e^{f(x_i)}\right) + \gamma_{\mathcal{H}} \|f\|_{\mathcal{H}}^2 + \gamma_{\mathcal{M}} f^T L f
\]

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

Higher cost on missing metastatic cancers
Avoid overfitting
Extract information from unimaged patients
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

<table>
<thead>
<tr>
<th>Bone scan</th>
<th>CT scan</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSA &gt; 20 or GS &gt; 7</td>
<td>PSA &gt; 20 or GS &gt; 7 or clinical T stage ≥ T3</td>
</tr>
</tbody>
</table>

- **20.7%** (prev 27%) of patients would be scanned
  - Of those, **17.0%** (prev 12%) would be positive
- Estimated **0.8%** of patients have missed metastatic disease
- **38%** negative scans would be avoided

- **22.6%** (prev 27%) of patients would be scanned
  - Of those, **14.3%** (prev 10%) would be positive
- Estimated **0.4%** of patients have missed metastatic disease
- **44%** negative scans would be avoided
MUSIC Imaging Appropriateness Criteria instituted across Michigan

- Statewide goal of performing imaging in ≥ 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

<table>
<thead>
<tr>
<th>Bone Scan</th>
<th>CT Scan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order Bone Scan If:</td>
<td>Order Bone Scan If:</td>
</tr>
<tr>
<td>PSA</td>
<td>PSA</td>
</tr>
<tr>
<td>&gt; 20</td>
<td>&gt; 20</td>
</tr>
<tr>
<td>OR</td>
<td>OR</td>
</tr>
<tr>
<td>≥ 8</td>
<td>≥ 8</td>
</tr>
<tr>
<td>OR</td>
<td>OR</td>
</tr>
<tr>
<td>Clinical T Stage</td>
<td>Clinical T Stage</td>
</tr>
<tr>
<td>≥ cT3</td>
<td>≥ cT3</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Imaging rates among patients not fitting the criteria</th>
<th>Baseline (2012-2013)</th>
<th>Post-intervention (Jan-Oct 2015)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone scan</td>
<td>11%***</td>
<td>6.3%</td>
</tr>
<tr>
<td>CT scan</td>
<td>14.7%***</td>
<td>7.6%</td>
</tr>
</tbody>
</table>

*Target = 10%*

***p-value < 0.001
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


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/