Supplementary Online Content


eTable 1. Subject Characteristics in the NCI-EDRN Urinary PC3 Evaluation Trial Validation Cohort
eFigure 1. STARD Study Flow Diagram for Developmental cohort
eFigure 2. STARD Study Flow Diagram for Validation Cohort
eFigure 3. Cost analysis and model assumption details
eAppendix.

This supplementary material has been provided by the authors to give readers additional information about their work.
**eTable 1: Subject Characteristics in the NCI-EDRN Urinary PC3 Evaluation Trial Validation Cohort**

<table>
<thead>
<tr>
<th>Demographic or Clinical Variable</th>
<th>Diagnosis Based on Prostate Biopsy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Cancer N (%)</td>
<td>Indolent Cancer: Gleason score ≤ 6 N (%)</td>
</tr>
<tr>
<td></td>
<td>297 (52.9%)</td>
<td>116 (20.7%)</td>
</tr>
<tr>
<td>Age in Yrs [Mean (range)]</td>
<td>60 (27-86)</td>
<td>62 (48-82)</td>
</tr>
<tr>
<td>Race</td>
<td>White/Caucasian</td>
<td>235 (79.1)</td>
</tr>
<tr>
<td></td>
<td>Black/African American</td>
<td>39 (13.1)</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
<td>8 (2.7)</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>15 (5.1)</td>
</tr>
<tr>
<td>Hispanic/Latino Ethnicity</td>
<td>Yes</td>
<td>26 (8.9)</td>
</tr>
<tr>
<td></td>
<td>Smoking status (Ever Smoked)</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Family History of Prostate cancer</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>DRE Results</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Enlarged/Benign</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abnormal - Suspicious</td>
</tr>
<tr>
<td></td>
<td>Serum PSA Pre Biopsy in ng/ml</td>
<td>Median (range)</td>
</tr>
<tr>
<td></td>
<td>Urinary PCA3 Score Pre Biopsy</td>
<td>Median (range)</td>
</tr>
<tr>
<td></td>
<td>Urinary T2:ERG Score Pre Biopsy</td>
<td>Median (range)</td>
</tr>
</tbody>
</table>

* = significant difference for Gleason 7+ compared to benign
# = significant difference for Gleason 7+ compared to benign and indolent combined (no aggressive PCa)
^ = significant difference for Gleason 7+ compared to Gleason 6 PCa.
Continuous variables were evaluated by Wilcoxon rank-sum test, and categorical variables by Fisher Exact, or Freeman-Halton tests.
eFigure 1: STARD Study Flow Diagram for Developmental cohort*

*Constitute a new, previously undescribed cohort comprised of subjects not included in prior studies of prostate cancer biomarkers, except 103 patients whose urine specimens had also undergone T2:ERG measurement using an earlier version of the T2:ERG urine assay in a previous study. (15)
eFigure 2. STARD Study Flow Diagram for Validation Cohort

Ineligible after consent, N=18, including:
- History of prostate cancer (3)
- Prior prostate surgery (6)
- Prior saturation biopsy (1)
- Prior prostate biopsy within 6 months (4)
- Otherwise did not meet inclusion criteria (4)

Inadequate urine sample, N=18

Uninformative PCA3, N=9

Biopsy not performed/done, N=13

Extended template biopsy not done, N=5

Biopsy outside window, N=6

T2:ERG Score not calculated due to low PSA, N=1

Repeat Biopsy, N=297

Eligible, N=858

Initial Biopsy, N=561, included as validation cohort

No Cancer (Gleason = NA), N=297

Cancer (Gleason ≥ 6), N=264, including:
- Gleason = 6, (116)
- Gleason ≥ 7, i.e., HG Cancer (148)
eFigure 3: Cost analysis and model assumption details.

We examined costs under three scenarios, as depicted below. The diagrams depict the flow of the 516 study patients based on how they would be treated under three different approaches based on assumptions about the likelihood of developing early and late stage disease. For example, in Scenario 1, we assume that 78 of the 98 patients with Gleason ≤6 score tumors (20%) develop late stage disease. We assume 122 of the 156 patients with Gleason ≥7 tumors (78%) develop late stage disease. In Scenario 2, we assume that biopsy is 100% sensitive for early stage tumors. All 98 and 156 of the men with Gleason ≤6 and Gleason ≥7 tumors are treated for early-stage disease. The numbers reported in the figure reflect the risk of developing late-stage tumors among men ages 55 to 64. We separately analyzed costs among men ages 65 to 74.

Assumptions:

a) Sensitivity and specificity of TMPRSS2, PCA3 and Serum PSA: Applying the TMPRSS2, PCA3 and Serum PSA testing algorithm described in the paper to the study cohort, we calculated that 51% of men without cancer would test positive, 87% of men with Gleason 6 tumors would test positive, and 95% of men with Gleason ≥7 tumors would test positive.


c) Lifetime costs of treatment: Cost estimates for patients diagnosed with early and late stage disease are from a study that analyzed SEER-Medicare data (Stokes et al. 2011). The study sample included Medicare beneficiaries ≥65 diagnosed with prostate cancer in 1991-2002. The costs represent the net lifetime stage-specific prostate cancer costs. The authors compared costs among cancer cases to costs incurred by matched non-cancer controls.

Stokes et al. included patients in the sample regardless of treatment approach (the proportion of men undergoing common treatment approaches are provided in Stokes et al.’s Table 2). Thus, cost estimates reflect the treatments available and treatment shares among Medicare beneficiaries during the study period. Likewise, long-term cost estimates reflect the survival experience of Medicare beneficiaries diagnosed during the study period. The sample includes patients who were treated and cured and patients who were treated but progressed to late-stage disease.
We re-stated costs in 2013 dollars using the all-item Consumer Price Index for urban consumers (Crawford and Church 2014). We discounted costs for late stage disease by 3% per year assuming they occur 5 years in the future. Crawford M, Church J. CPI Detailed Report Data for May 2014. Stokes ME, Ishak J, Proskorovsky I, Black LK, Huang Y. Lifetime economic burden of prostate cancer. BMC Health Serv Res. 2011;11:349.

d) Likelihood of developing late stage disease if early stage disease is undetected: We based assumptions about the age-specific lifetime risk of developing late-stage prostate cancer on data reported in Albertson et al. We assume the probability of developing a late-stage tumor among men ages 55 to 64 with Gleason ≤6 tumors is 20%. We assume the probability of developing a late-stage tumor among men with Gleason ≥7 tumors is 78%. The corresponding figures for men ages 65 to 74 are 22% and 47%.

Consistent with the view that we are modeling the effects of a one-time screen, and for the sake of simplicity, we do not consider the possibility that patients with cancer who test negative on the first test continue to undergo repeat testing.


e) Cost of enzalutamide: The data used to measure the costs incurred by patients with late-stage tumors (Stokes et al.) pre-date the release of enzalutamide, radium Ra 223 dichloride, and sipuleucel-T. These therapies are costly, and may have a substantial impact on the net cost associated with early detection of prostate cancer. We assumed the cost of enzalutamide is $65,619 per patient (Howard et al.). We simulated costs assuming 20%, 50%, and 80% of patients diagnosed with late-stage tumors receive enzalutamide.

Howard DH, Bach PB, Bernt ER, Conti R. Pricing in the market for anticancer drugs. J Econ Perspect. In press. Because the cost of the T2:ERG and PCA3 testing strategy has not been well-established, we did not include the cost in our analysis. Reported costs do not include the direct costs associated with T2:ERG and PCA3 testing.
Scenario 1: No biopsy

Abnormal PSA/DRE (N = 1000)

- No cancer (508)
  - No treatment
    - Cost: $0

- Gleason: \( \leq 6 \) (190)
  - Death before cancer
    - (151; 80%)
    - Cost: $0
  - Late stage disease
    - (39; 20%)
    - Cost: $30,000

- Gleason: \( \geq 7 \) (302)
  - Death before cancer
    - (66; 22%)
    - Cost: $0
  - Late stage disease
    - (236; 78%)
    - Cost: $30,000
Scenario 2: PSA-driven biopsy

Abnormal PSA/DRE (N = 1000)

- No cancer (508)
  - Biopsy Cost: $2,300
  - No treatment Cost: $0

- Gleason: ≤6 (190)
  - Biopsy Cost: $2,300
  - Early stage disease Cost: $39,000

- Gleason: ≥7 (302)
  - Biopsy Cost: $2,300
  - Early stage disease Cost: $39,000
A. POWER ANALYSIS SIMULATION

We note that our analysis was based on a total sample size of 561 subjects: non cases are those with Gleason score <7, i.e. 413 men are non cases. Cases are those that exhibit Gleason score ≥ 7, i.e. we have 148 men are cases.

© 2017 American Medical Association. All rights reserved.
In order to explore the attained power we need to conduct a simulation study of the bootstrap scheme we employed to derive the significance related to the comparison of PSA alone VS our three marker rule. To achieve that, we need to generate data from a mechanism that yields similar score distributions to all three biomarkers, i.e. PSA, PCA3, and T2ERG. Normality of these biomarkers is severely violated. Thus, for our simulation purposes we transform all markers by taking the natural logarithm to which we then fit separately three normal distributions. These distributions serve as data generators to scores which we will back-transform by the taking the exponential function. This way the generated data will approximate our actual data and hence we can be safe for the power results that we attain here. We force the sample size of our simulated data to be the same as with our actual sample size (148 and 413, i.e. total 561). We conduct two simulations: (1) One in which the ROC of the PSA is empirically estimated and in every bootstrap iteration we match its sensitivity to the binary rule based score. We use 1,000 monte carlo iterations with 1,000 bootstrap samples each. (2) One in which the ROC of PSA is estimated by a kernel based estimate in which the bandwidth used is optimal in terms of asymptotic mean integrated squared error (AMISE) (see Silverman (1998)). The attained power in simulations (1) and (2) are 83.5% and 86% respectively.

B. RELATIONSHIP BETWEEN URINARY PCA3/T2ERG AND PSA

Post-hoc, cross-tables were constructed to explore the relationship between the binary PCA3 and T2:ERG status and PSA: Significant correlation was observed between the urinary PCA3/T2Erg and dichotomized PSA (less than or greater than 10 ng/ml) in the Developmental Cohort (Fisher’s exact test p = 0.008) though this was not significant in the validation cohort (p=0.09)

Developmental cohort 2X2 table:

<table>
<thead>
<tr>
<th>PCA3/T2Erg</th>
<th>PSA ≤ 10</th>
<th>PSA &gt; 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>146 (7, 139)</td>
<td>337 (127,210)</td>
</tr>
<tr>
<td>Y</td>
<td>3 (0,3)</td>
<td>30 (22,8)</td>
</tr>
</tbody>
</table>

*Within the bracket, (high-grade cancer, no cancer/ low grade cancer)

<Testing for independence between PCA3/T2Erg rule and dichotomized PSA>

⇒ Chi-square p-value = 0.009 / Fisher’s exact p = 0.008

Validation cohort 2X2 table:

<table>
<thead>
<tr>
<th>PCA3/T2Erg</th>
<th>PSA &lt;= 10</th>
<th>PSA &gt; 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>146 (11, 135)</td>
<td>344 (94, 250)</td>
</tr>
<tr>
<td>Y</td>
<td>14 (3, 11)</td>
<td>57 (40, 17)</td>
</tr>
</tbody>
</table>

© 2017 American Medical Association. All rights reserved.
*Within the bracket, the 1st number : high-grade cancer, and the 2nd number : no cancer or low grade cancer

<Testing for independence between PCA3/T2Erg rule and dichotomized PSA>

Chi-square $p = 0.08$ / Fisher’s exact $p = 0.09$

C. CALCUTATING TWO-SIDED P-VALUE AS A MEASURE OF SIGNIFICANCE OF THE DIFFERENCE IN SPECIFICITY FOR DETECTING AGGRESSIVE PROSTATE CANCER

Post-hoc, we used a Kernel based approach to determine p-value for one-sided and two-sided difference between specificity in detecting aggressive prostate cancer. We used normal kernel smoothers with optimal bandwidth in terms of mean integrated squared error in order to obtain the ROC estimator of the PSA, and this yielded a two-tailed p-value that corresponds to the difference in specificity = 0.0148 and one tailed p-value = 0.0074. Under this approach we employed a normal kernel density estimate to obtain a smooth ROC curve for the PSA. More specifically the algorithm employed was as follows:

**Step 1:** Sample with replacement separately from the healthy and the diseased.

**Step 2:** Based on these bootstrap samples build the ROC of the PSA and use a normal kernel to smooth it, by separately smoothing out the underlying cdfs. Then calculate the sensitivity and false positive rate (FPR) of the new test.

**Step 3:** For the current bootstrap sample find the FPR(PSA) on the kernel based ROC curve that corresponds to sensitivity of the new test.

**Step 4:** For the current bootstrap iteration calculate $d = \text{FPR(PSA)} - \text{FPR(new test)}$ for the current bootstrap sample.

**Step 5:** Repeat 10000 times and obtain the histogram of $d$ as well as the percentage of those values smaller than zero ($d < 0$).

The results obtained by the above algorithm yield a one sided p-value of 0.0074 and a two-tailed p-value of 0.0148.
For the kernel based ROC of the PSA along with the underlying bootstrap distribution of the differences see Fig 1a and 1b respectively. The fact that both the one and two sided p-value are <0.05 with this approach is not surprising since our simulation studies have shown a similar behavior of the kernels compared to the empirical based ROC curves (see Bantis and Feng (2016)). These findings are in line -and in fact theoretically justified- by the paper of Lloyd and Yong (1999) entitled as "Kernel estimators of the ROC curve are better than empirical". We note that we employ a bandwidth that corresponds to optimality in terms of asymptotic integrated mean squared error for normal kernels (see Silverman (1998)).

References: