Two-Stage Biomarker Protocols for Improving the Precision of Early Detection of Prostate Cancer

Christine L. Barnett, PhD, Scott A. Tomlins, MD, PhD, Daniel J. Underwood, PhD, John T. Wei, MD, Todd M. Morgan, MD, James E. Montie, MD, Brian T. Denton, PhD

Background. *New cancer biomarkers are being discovered at a rapid pace; however, these tests vary in their predictive performance characteristics, and it is unclear how best to use them.* ***Methods.*** *We investigated 2-stage biomarker-based screening strategies in the context of prostate cancer using a partially observable Markov model to simulate patients' progression through prostate cancer states to mortality from prostate cancer or other causes. Patients were screened every 2 years from ages 55 to 69. If the patient's serum prostate-specific antigen (PSA) was over a specified threshold in the first stage, a second stage biomarker test was administered. We evaluated design characteristics for these 2-stage strategies using 7 newly discovered biomarkers as examples. Monte Carlo simulation was used to estimate the number of screening biopsies, prostate cancer deaths, and quality-adjusted life-years (QALYs) per 1000 men.* ***Results.*** *The all-cancer biomarkers significantly underperformed the high-grade cancer biomarkers in terms of QALYs. The screening strategy that used a PSA threshold of 2 ng/mL and a second biomarker test with high-grade sensitivity and specificity of 0.86 and 0.62, respectively, maximized QALYs. This strategy resulted in a prostate cancer death rate within 1% of using PSA alone with a threshold of 2 ng/mL, while reducing the number of biopsies by 20%. Sensitivity analysis suggests that the results are robust with respect to variation in model parameters.* ***Conclusions.*** *Two-stage biomarker screening strategies using new biomarkers with risk thresholds optimized for high-grade cancer detection may increase quality-adjusted survival and reduce unnecessary biopsies.* ***Key words:*** *Biomarkers, Prostate Cancer, Markov Model, Simulation.* ***(Med Decis Making XXXX;XX:xx–xx)***

Although prostate cancer is the most common solid tumor in American men, controversy surrounds prostate cancer screening. The American Urological Association (AUA) recommends shared decision making for men ages 55–69 considering prostate-specific antigen (PSA)–based screening and specifies screening intervals of 2 years to preserve the majority of the benefits of screening and reduce overdiagnosis and false positives.[1] However, the United States Preventive Services Task Force recommends against prostate cancer screening with the PSA due to the resulting unnecessary biopsies and overtreatment of low-risk disease.[2] In recent years, many new biomarkers have been discovered for early detection of prostate cancer that may be able to supplement the PSA test to reduce unnecessary biopsies. Patients and their health care providers now have access to these new biomarkers, which could potentially be combined into multi-stage biomarker screening strategies that improve the precision with which screening can be performed. These discoveries have the potential to improve patient survival and lower the burden of screening by better discriminating between patients with and without cancer. However, these tests vary in their predictive characteristics, and the ideal way to use them to achieve optimal long-term health benefits is unclear. In this article we study the question of how to design 2-stage biomarker screening strategies in the context of prostate cancer.

Several new diagnostic prostate cancer biomarkers have recently come to market.[3,4] Some of these biomarkers are PSA derivatives, such as free PSA and [–2]proPSA. Some of the biomarkers are based on combinations of serum markers, such as the prostate health index (phi), which uses a combination of total PSA, free PSA, and [–2]proPSA to

DOI: 10.1177/0272989X17696996

generate a score,[5,6] and the 4Kscore, which uses a panel of total PSA, free PSA, intact PSA, and human kallikrein 2 (hK2) to estimate a patient's risk of high-grade cancer (Gleason score ≥ 7) on biopsy. Other molecular biomarkers include prostate cancer antigen 3 (PCA3) and TMPRSS2:ERG (T2:ERG), which are detectable in post digital rectal exam (DRE) urine.[7–13] The Mi-Prostate Score (MiPS) early detection test combines a patient's serum PSA, urine PCA3 score, and urine T2:ERG score in a multivariate regression model to estimate individualized risk estimates for all prostate cancer and high-grade prostate cancer.[14] These tests vary in the outcome they predict (all cancer v. high-grade cancer) and in their sensitivities and specificities. No study has yet attempted to compare these biomarkers to determine which characteristics achieve optimal long-term health outcomes.

To better understand the optimal design of screening strategies in a multibiomarker setting, we estimated long-term health outcomes using a partially observable Markov model. We validated the model by comparing model-based estimates of health outcomes with independent estimates reported in the literature. We compared each of the biomarkers based on patients who were screened from ages 55 to 69 with a screening interval of 2 years. During each screening period, we used an innovative 2-stage biomarker screening strategy. If the patient's serum PSA was over a specified threshold (2 or 4 ng/mL), a second biomarker test was administered. We estimated the number of prostate cancer deaths and screening biopsies per 1000 men, as well as the gain in quality-adjusted life-years (QALYs) compared with no screening, in order to identify the ideal biomarker characteristics. We drew conclusions about optimal screening strategy design characteristics that may generalize to other disease contexts in which multiple biomarkers can be used to achieve early detection.

METHODS

To evaluate screening strategies that use biomarkers of varying sensitivity and specificity, we developed a partially observable Markov model in which pretreatment states are not directly observable. Biomarker tests give (imperfect) information about the true state of the patient. The partially observable pretreatment states in the model include no prostate cancer, undetected organ-confined prostate cancers based on Gleason score (GS < 7, GS $= 7$, GS > 7), and extraprostatic or lymph node-positive cancer (EPLN). The EPLN state aggregates these 2 conditions into one state because they are similarly associated with decreased survival. The states were selected because they distinguish patients on the basis of likely treatment options, outcomes, and survival.

Model Parameters

Figure 1 displays the health states and possible state transitions for the model. Each year that the screening strategy calls for testing, the following sequence of events in the model occur: The patient receives biomarker tests; the biomarker test results determine whether a biopsy is performed; and the patient transitions to his next health state. As our model focuses on screening of the general population, the screening strategy terminates after initial biopsy and the patient continues to make state transitions in the absence of screening until reaching one of the absorbing states: all-other-cause mortality or prostate cancer mortality. The parameters used to calculate the transition probabilities are described in Table 1, and how these parameters were calculated is described in the supplementary material.

Prostate Cancer Screening

The structure of the 2-stage biomarker screening strategy is illustrated in Figure 2, in which 2 thresholds divide PSA values into low, intermediate, and

Received 4 February 2016 from Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI (CLB, BTD); Department of Urology, University of Michigan, Ann Arbor, MI (SAT, JTW, TMM, JEM, BTD); Department of Pathology, University of Michigan, Ann Arbor, MI (SAT); and Department of Industrial & Systems Engineering, North Carolina State University, Raleigh, NC (DJU). This work was presented in a poster session at the SMDM 36th Annual Meeting. Financial support for this study was provided in part by grants from the National Science Foundation (CMMI 1536444 to BTD, DGE 1256260 to CLB), a grant from the Early Detection Research Network (7-U01-CA-113913-09 to JTW), and a grant from the University of Michigan MCubed program (to BTD, SAT, and JTW). The funding agreements ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. Revision accepted for publication 27 January 2017.

Supplementary material for this article is available on the *Medical Decision Making* Web site at http://journals.sagepub.com/home/mdm.

Address correspondence to Brian T. Denton, Department of Industrial and Operations Engineering, University of Michigan, 2893 IOE Building, 1205 Beal Ave., Ann Arbor, MI 48109-2117; telephone: (734) 763-2060; fax: (734) 764-3451; e-mail: btdenton@umich.edu.

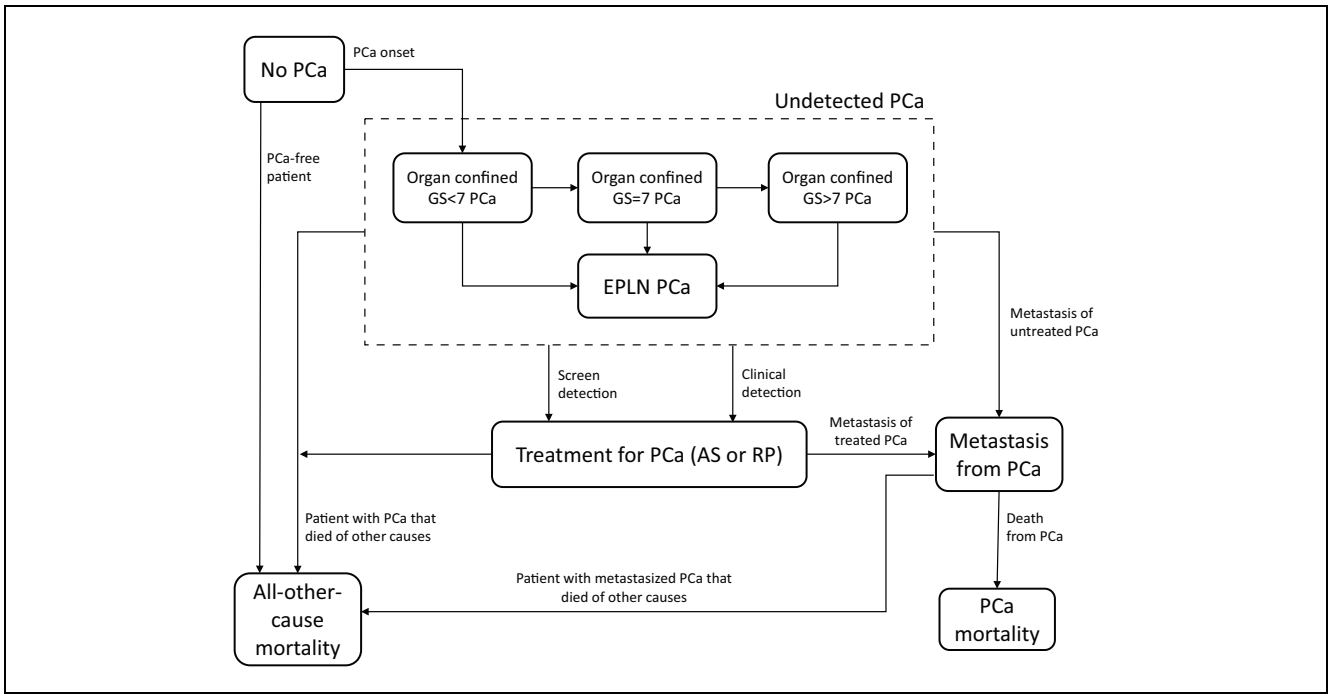

Figure 1 State transition diagram. Health states and progression paths in the Markov model are shown, where transitions between states are represented by arrows. Patients who are detected with prostate cancer (PCa) are treated immediately with radical prostatectomy (RP) or active surveillance (AS). GS, Gleason score; EPLN, extraprostatic or lymph node-positive cancer.

high. A patient receives a biopsy if his PSA value is "high" (>10 ng/mL). If his PSA value is "low" at a given screening age, then no biopsy is recommended. If the PSA is between the "low" and "high" thresholds, then a second biomarker test is used. If the second biomarker test is positive, the patient receives a biopsy; otherwise, the patient does not receive a biopsy and continues to be screened in future years. We evaluated 2 PSA thresholds to trigger a second biomarker test: 2 and 4 ng/mL. We selected these thresholds because it has been reported that phi, 4Kscore, and [–2] proPSA have the ability to select men with PSA values of 2–10 ng/mL for prostate biopsy and because 4 ng/mL is a commonly used biopsy threshold.[15] We chose to use this 2-stage screening strategy for multiple reasons. First, PSA is an established test and many new biomarkers are only approved to be used along with the PSA test. Second, new biomarkers can be expensive, and this approach pragmatically uses the new biomarkers when they will add greatest value and does not use them when they have little value. Additionally, we assumed 100% adherence to the screening strategy and performed sensitivity analysis on the adherence rates.

We sampled PSA scores using a random effects model that includes the patient's current age and his age at onset of a preclinical tumor.[16] For the sensitivity and specificity of the second biomarker test, we used values reported in the literature. We performed a systematic review of the literature and chose the sensitivities and specificities that were nondominated (i.e., biomarkers such that no other biomarker had both a higher sensitivity and a higher specificity). Table 2 shows sensitivity and specificity values that we used for all cancer and high-grade cancer (Gleason score $\geq$7). These 14 tests were evaluated for 2 PSA thresholds (2 ng/mL and 4 ng/mL), resulting in 28 screening strategies. Biopsy results were randomly sampled as either positive or negative, assuming a sensitivity of 0.8.[30] If the biopsy result was positive, we estimated the probability that the biopsy provides an incorrect grading at diagnosis based on data reported by Epstein and others.[17]

Clinical Detection of Prostate Cancer

Patients were diagnosed with prostate cancer in 1 of 2 ways: by routine screening (i.e., an elevated biomarker score that leads to a positive biopsy) or by

Table 1 Parameters, Their Sources, and the Specific Values Used in Our Base Case and Sensitivity Analysis

Parameter	Symbol	Low Value(s)	Base Case Value(s)	High Value(s)	Source (Ref. No.)
Annual transition rate from no PCa to GS < 7	w_t	Lower bound of 95% CI	0.004–0.069	Upper bound of 95% CI	28
Annual other-cause mortality rate	d_t	–20%	0.002–0.347	+20%	29
Annual metastasis rate for patients with undiagnosed PCa	e_t	–10%	0.002–0.035	+10%	Calibrated
Annual PCa-specific mortality rate given metastasized PCa	z_t	–10%	0.181–0.204	+10%	
Sensitivity of prostate biopsy procedure	f	–10%	0.8	+10%	30
Annual transition rate from GS < 7 to GS = 7	$o1o2$	–10%	0.101	+10%	28
Annual transition rate from GS = 7 to GS > 7	$o2o3$	–10%	0.087	+10%	31
Annual transition rate from GS < 7 to EPLN	$o1e$	–10%	0.029	+10%	31
Annual transition rate from GS = 7 to EPLN	$o2e$	–10%	0.081	+10%	31
Annual transition rate from GS > 7 to EPLN	$o3e$	–10%	0.097	+10%	31
Probability of no possible recurrence following definitive treatment in state EPLN	pnc	–10%	0.468	+10%	31
Proportion of patients detected with GS < 7 who undergo active surveillance	s	–10%	0.485	+10%	32
Annual metastasis rate for patients with possible recurrence after definitive treatment in EPLN	g	–10%	0.006	+10%	19
Instantaneous QALY disutility for screening	δ_{Scr}	0.0	0.00019	0.00019	Mayo Clinic Radical Prostatectomy Registry
Instantaneous QALY disutility for a prostate biopsy	δ_{Biop}	0.00346	0.00577	0.00750	33
Instantaneous QALY disutility for PCa diagnosis	δ_{Dia}	0.0125	0.01667	0.0208	33
Instantaneous QALY disutility for radical prostatectomy	δ_{Tre}	0.0917	0.24667	0.323	33
Annual QALY disutility for 9-year post–radical prostatectomy recovery period	δ_{Rec}	0.0	0.05	0.07	33
Annual QALY disutility for active surveillance	δ_{AS}	0.0	0.03	0.15	33
Annual QALY disutility for metastasis	δ_{Met}	0.14	0.4	0.76	33

Note: CI, confidence interval; EPLN, extraprostatic or lymph node-positive cancer; GS, Gleason score; PCa, prostate cancer; QALY, quality-adjusted life-year.

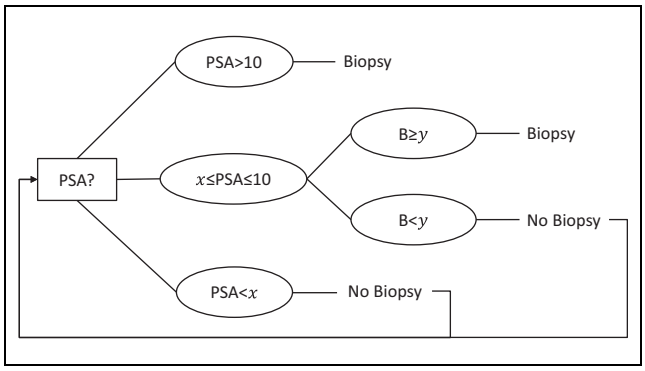

Figure 2 Two-stage biomarker screening strategy where the result of the prostate-specific antigen (PSA) test determines whether a second biomarker is used. If a patient's PSA score is greater than 10 ng/mL, he will automatically receive a biopsy. B *represents the observed second biomarker result for the patient,* x *is the PSA threshold to trigger a second biomarker test, and* y *is the threshold for the second biomarker to trigger biopsy.*

clinical detection (i.e., prostate cancer that develops symptoms). We assumed that the "lead time clock" for clinical detection starts once a patient has both prostate cancer and a PSA score $\geq$3 ng/mL. Savage and others[18] developed a distribution of lead times from an elevated PSA measurement of $\geq$3 ng/mL to clinical diagnosis of prostate cancer. For each patient, we randomly sampled a lead time from this distribution. If a patient's lead time is x years, after the patient has had prostate cancer and a PSA score $\geq$3 ng/mL for x years, if the patient is alive and has neither been diagnosed nor treated for prostate cancer, then the patient is assumed to be clinically detected.[18]

Prostate Cancer Treatment

Following diagnosis, patients received watchful waiting, active surveillance, or radical prostatectomy. We assumed that patients with Gleason score $\geq$7 received radical prostatectomy. Patients diagnosed with Gleason score <7 were assumed to be treated via active surveillance or radical prostatectomy. Based on practice patterns reported by Liu and others,[19] we assumed that 48.5% of patients diagnosed with Gleason score <7 received active surveillance, while the other 51.5% received radical prostatectomy. Given the lack of consensus in published guidelines for active surveillance, we assumed that patients received a biopsy 1 year after diagnosis, followed by a biopsy every 2 years for 10 years following diagnosis.[20] Patients over age 80 were assumed to receive watchful waiting.

Patients receiving active surveillance continue to progress through the natural history of the disease until they have a biopsy result of Gleason score $\geq$7. We made the same assumptions about surveillance biopsies as described above. If a patient's Gleason score was upgraded as a result of a surveillance biopsy, he was assumed to have a radical prostatectomy. However, if he was never detected to have higher risk disease, he had the survival of an untreated patient. Survival following radical prostatectomy depends on the stage of the disease at treatment. There are 2 posttreatment states patients can transition to following treatment: no recurrence following treatment (NRFT), and possible recurrence following treatment (PRFT). If a patient has organ-confined disease at surgery, he transitions directly to NRFT. If a patient has extraprostatic or lymph node-positive disease at treatment, he transitions to NRFT with probability 0.468 (defined as *pnc* in Table 1), and he transitions to PRFT with probability 0.532. The annual metastasis rate for patients in PRFT is 0.006 based on the Mayo Clinic Radical Prostatectomy Registry (defined as g in Table 1). From the postdiagnosis states, patients eventually transition to metastasis and/or death from prostate cancer or other causes.

Model Validation

To perform model validation, we compared estimates of clinical statistics from our model with literature estimates. The model estimates were based on the assumption that all men were screened annually from age 50 to 75 with a PSA threshold of 4 ng/mL, because that was a common strategy at the time on which the literature estimates are based.[21,22] We compared our model results with independent estimates from the literature for age-dependent risks of prostate cancer death, expected lifespan for a 40-year-old man, age-dependent risks of prostate cancer diagnosis, Gleason score distribution at diagnosis, and biopsy-detectable prostate cancer prevalence rates by age.

Simulation Parameters

The AUA recommends shared decision making for men considering PSA-based screening from ages 55 to 69 and recommends a screening interval of 2 years. Based on this recommendation, patients were

Table 2 Biomarker Sensitivities and Specificities for All Cancer and High-Grade Cancer (Gleason Score ≥7) Reported in the Literature

Biomarker Test	Threshold	Sensitivity	Specificity	Source (Ref. No.)
All cancer				
% p2PS	≥1.7	0.70	0.70	35
% p2PS	≥2.5	0.38	0.90	35
phi	≥38.7	0.85	0.61	35
PCA3	—	0.93	0.37	36
T2:ERG	—	0.67	0.87	36
T2:ERG	—	0.37	0.93	37
High-grade cancer				
4Kscore	≥9%	0.90	0.52	34
4Kscore	≥12%	0.86	0.62	4
4Kscore	≥15%	0.79	0.70	34
All-cancer MiPS	≥25%	0.94	0.41	14
All-cancer MiPS	≥52%	0.68	0.78	14
High-grade MiPS	≥10%	0.95	0.36	14
High-grade MiPS	≥15%	0.88	0.55	14
High-grade MiPS	≥26%	0.70	0.76	14

Note: The sensitivities and specificities for 4Kscore and the MiPS tests were calculated using data presented by Parekh and others[34] and Tomlins and others,[14] respectively. These 14 tests were evaluated for 2 PSA thresholds (2 ng/mL and 4 ng/mL), resulting in 28 screening strategies. Blank entries for thresholds indicate no threshold given in the source.

PSA-screened every 2 years from ages 55 to 69.[1] Each patient simulation began at age 40. The model was used to evaluate 28 different prostate cancer screening strategies based on published estimates of sensitivity and specificity for biomarkers reported in the literature. Table 2 shows the sensitivity and specificity values for all cancer and high-grade cancer (Gleason score ≥7). We compared these values with using PSA alone and to hypothetical perfect biomarkers that have a sensitivity and specificity of 1.0 for either all cancer or high-grade cancer. We also investigated the tradeoff of sensitivity and specificity by evaluating long-term health outcomes for patients under 30 different thresholds for the high-grade MiPS test. To perform this analysis, we used a large dataset of PSA, PCA3, and T2:ERG scores from a presumed cancer-free population of patients undergoing diagnostic prostate biopsy to estimate the high-grade sensitivity and specificity of the high-grade MiPS test under each threshold.[14]

For each strategy evaluated, we estimated the mean number of screening biopsies and prostate cancer deaths per 1000 men and the mean QALYs gained per 1000 men relative to no screening. Our QALY measurements account for disutilities of screening, biopsy, diagnosis, active surveillance, radical prostatectomy, recovery from radical prostatectomy, and metastasis; the values of the disutilities with their sources are shown in Table 1. The reward update function for QALYs was

$$r_t(s_t,a_t)=1-\delta_{Scr}-\delta_{Biop}-\delta_{Dia}-\delta_{Tre}-\delta_{Rec}-\delta_{AS}-\delta_{Met},$$

where $r_t(s_t,a_t)$, is the reward a patient receives at age t, which is 1 minus the disutilities associated with screening, biopsy, diagnosis, treatment and the presence of metastatic cancer, as defined in Table 1. The arguments for the reward are the health state s_t that defines the cancer status of the patient and the action, a_t, that defines whether a screening test or biopsy was performed. The total expected QALYs a patient receives in his lifetime is

$$R=E^{\pi}\left[\sum_{t=40}^{T} r_t(s_t,a_t)\right],$$

where T denotes maximum lifespan and the expectation is with respect to the stochastic process induced by the screening strategy π that defines the frequency of testing and the thresholds at which to perform biomarker tests and/or biopsies. Since we are not analyzing costs, we did not use a discount factor. This amounts to assuming a risk-neutral decision maker (e.g., the patient).

Table 3 Best Performing Strategies in Terms of QALYs Gained per 1000 Men Compared with No Screening

Second Biomarker				Expected QALYs Gained per 1000 Men	Number of Screening Biopsies per 1000 Men	Number of PCa Deaths per 1000 Men
Test	Threshold	Sensitivity	Specificity			
Perfect: HG[a]	—	1	1	21.04	128.2	27.5
4Kscore[a]	≥12%	0.86	0.62	18.59	211.9	27.7
4Kscore[a]	≥15%	0.79	0.70	18.52	200.2	27.8
4Kscore[a]	≥9%	0.9	0.52	18.51	222.6	27.6
HG MiPS[a]	≥15%	0.88	0.55	18.48	219.6	27.7
MiPS[a]	≥25%	0.94	0.41	18.38	231.7	27.6
HG MiPS[a]	≥10%	0.95	0.36	18.30	235.0	27.6
Perfect: all	—	1	1	18.01	146.5	27.1
HG MiPS[a]	≥26%	0.70	0.76	17.93	188.4	27.9
MiPS[a]	≥52%	0.68	0.78	17.79	184.2	28.0
PSA alone	—	—	—	17.75	251.7	27.5
PCA3	—	0.93	0.37	17.65	236.9	27.6
phi	≥38.7%	0.85	0.61	17.46	218.7	27.6

Note: Each strategy has a PSA threshold of 2 ng/mL to trigger a second biomarker test and assumes a biopsy will automatically be performed on any patient with a PSA ≥10 ng/mL. PCa, prostate cancer; QALY, quality-adjusted life-year.
a. Sensitivity and specificity to high-grade (HG) prostate cancer (Gleason score ≥7).

Simulation was performed to generate sample paths and obtain statistical estimates of expected rewards for each strategy. This simulation model was implemented in C/C++. We ran each strategy for 30,000,000 sample paths, which took less than 12.5 minutes to run using 3.40 GHz with 16 GB of RAM. The largest 95% confidence interval reflecting Monte Carlo error was less than 1% of the corresponding sample-mean point estimate.

Sensitivity Analysis

We performed 1-way sensitivity analysis on all of the model parameters. These parameters were varied from their base case values to high and low values, as defined in Table 1. We also performed probabilistic sensitivity analysis, during which we varied each model parameter according to a uniform distribution between the low and high values reported in Table 1. During the probabilistic sensitivity analysis, we performed 30 experiments with 30,000,000 sample paths for each experiment. We additionally looked at the impact of varying screening participation and adherence rates. We looked at the effect of varying these parameters on the expected number of prostate cancer deaths per 1000 men and the increase in QALYs per 1000 men relative to no screening. To perform this sensitivity analysis, we used the strategy with a PSA threshold of 2 ng/mL and a second biomarker test with a high-grade sensitivity and specificity that maximized QALYs.

RESULTS

Model Validation

Table A.1 in the supplementary material compares estimates of clinical statistics from our model with literature estimates. Overall, our estimates from the model compare well with estimates from the literature. Any variations are most likely due to our assumption that patients have perfect adherence to the screening strategy.

Base Case Analysis

We estimated the expected number of QALYs gained per 1000 men relative to no screening for each of the biomarkers in Table 2 as well as 2 hypothetical perfect biomarkers. Ten of the new biomarkers maximized expected QALY gains with overlapping confidence intervals. The performance outcomes for these 10 biomarkers are shown in Table 3 along with the results for the hypothetical perfect biomarkers. While there was no statistically significant difference between these 10 tests in the QALYs gained per 1000 men, the number of biopsies per 1000 men varied from 184 to 237. These 10 tests also performed significantly better than using PSA alone with a threshold of 4 ng/mL, achieving between 55% and 65% more QALYs per 1000 men. In terms of the initial PSA threshold to trigger a second biomarker test, a PSA threshold of 2 ng/mL

performed significantly better than 4 ng/mL in all 2-stage strategies, where using an initial PSA threshold of 2 ng/mL achieved between 55% and 65% more QALYs gained per 1000 men than using an initial PSA threshold of 4 ng/mL.

Figure 3 provides results for the number of screening biopsies and prostate cancer deaths per 1000 men. The figure displays tests from the literature that were on the efficient frontier (i.e., any strategy that resulted in more biopsies and more prostate cancer deaths than another strategy was removed), in addition to the perfect biomarkers and using PSA alone. Figure 3 shows the tradeoff that occurs between minimizing prostate cancer deaths and minimizing the number of screening biopsies. Although a PSA threshold of 4 ng/mL resulted in fewer biopsies, it also resulted in more prostate cancer deaths. For example, consider the strategy with a PSA threshold of 4 ng/mL and a second biomarker test with a sensitivity and specificity of 0.67 and 0.87 compared with the same strategy with a PSA threshold of 2 ng/mL. The latter strategy is more aggressive and thus reduces prostate cancer deaths by 6% compared with the former strategy; however, the latter strategy increases the number of screening biopsies being performed by 49%. As expected, screening strategies with higher sensitivity resulted in fewer prostate cancer deaths and more biopsies, while strategies with higher specificity resulted in fewer biopsies and more prostate cancer deaths. Only 2 tests maximized QALYs and also appeared on the efficient frontier of Figure 3: using PSA alone with a threshold of 2 ng/mL and using the phi test with a threshold of 38.7. Intuitively, using PSA alone with a threshold of 2 ng/mL minimized prostate cancer deaths. The screening strategy that used a PSA threshold of 2 ng/mL and a second biomarker test with high-grade sensitivity and specificity of 0.86 and 0.62, respectively, maximized QALYs and resulted in a prostate cancer death rate within 1% of using PSA alone with a threshold of 2 ng/mL while reducing the number of biopsies by 20%.

In addition to the efficient frontier of tests, Figure 3 also shows the results for using PSA alone and for hypothetical biomarkers with perfect sensitivity and specificity to all cancer and to high-grade cancer. There exists a 2-stage biomarker strategy that can simultaneously reduce the number of prostate cancer deaths and the number of screening biopsies compared with using PSA alone with a threshold of 4 ng/mL. In particular, using a PSA threshold of 2 ng/mL followed by a test with sensitivity and specificity of 0.37 and 0.93, respectively, can reduce the number of prostate cancer deaths by 2% and the number of screening biopsies by 7% compared with using PSA alone with a threshold of 4 ng/mL. For both PSA thresholds, the test with perfect sensitivity and specificity to high-grade cancer resulted in more prostate cancer deaths but fewer biopsies compared with the test with perfect sensitivity and specificity to all cancer. This further highlights the tradeoff between these 2 competing objectives.

To further investigate the relationship between possible biomarker thresholds, the subsequent sensitivities and specificities that they imply, and long-term health outcomes, we evaluated 30 different thresholds for the high-grade MiPS test using the logistic regression model described by Tomlins and others[14]; the thresholds we considered ranged from 6% to 35% risk of high-grade cancer on biopsy.

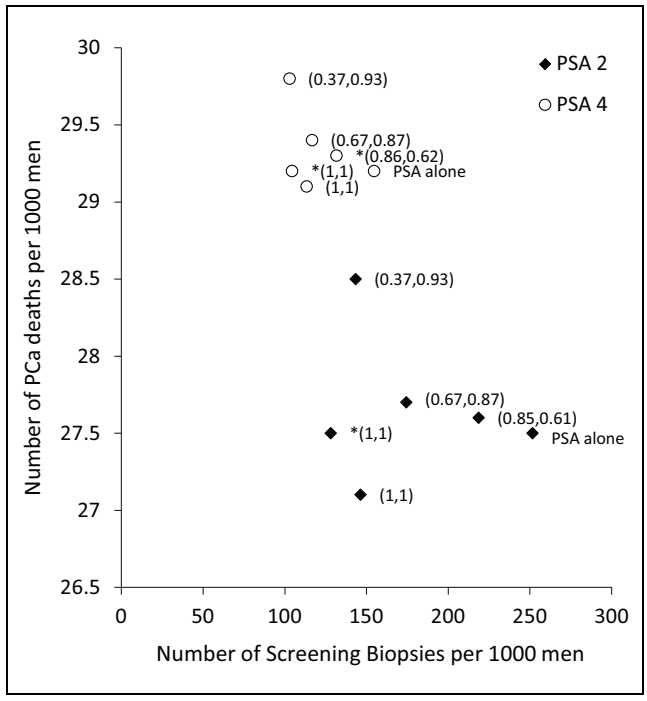

Figure 3 Estimated number of prostate cancer (PCa) deaths and screening biopsies per 1000 men from modeled screening strategies. Each point on the graph represents a different screening strategy and is labeled with the sensitivity and specificity of the second biomarker. An asterisk indicates that the sensitivity and specificity are for high-grade prostate cancer (Gleason score ≥7). This graph displays only the nondominated strategies of each strategy type: that is, strategies such that no other strategy resulted in both a lower number of screening biopsies and a lower number of PCa deaths per 1000 men screened (with the exception of the hypothetical perfect biomarkers and PSA alone, which have been shown for reference).

Figure A.1 in the supplementary material shows the relationship between the 30 MiPS thresholds, the resulting sensitivity and specificity to high-grade disease, and the expected increase in QALYs per 1000 men compared with no screening. Figure A.1 demonstrates that as specificity is increased and sensitivity is decreased, the expected number of QALYs decreases, which indicates that it is important to maximize sensitivity to high-grade disease in order to maximize expected QALYs. We found that although several thresholds perform equally well in terms of QALYs, we can distinguish between these strategies by looking at performance outcomes in addition to QALYs.

Sensitivity Analysis

We performed 1-way and probabilistic sensitivity analysis on the screening strategy that maximized expected QALYs, which has a PSA threshold of 2 ng/mL, and a second biomarker test with a high-grade sensitivity and specificity of 0.86 and 0.62, respectively. Using the base case parameter values, this high-grade MiPS strategy resulted in 27.7 prostate cancer deaths, 212 screening biopsies, and a gain of 19 QALYs per 1000 men. The 1-way sensitivity analysis results are shown in Figures A.2 and A.3 in the supplementary material, which are tornado diagrams that display the effect each parameter has on the expected increase in QALYs and the expected number of prostate cancer deaths per 1000 men, respectively. The parameter that had the greatest effect on both expected gain in QALYs and expected number of prostate cancer deaths was d_t, the annual other-cause mortality rate. When the low and high values of the annual other-cause mortality rate are used, the expected increase in QALYs per 1000 men ranged from 8 to 35 relative to the base case value of 19 QALYs, and the expected number of prostate cancer deaths per 1000 men ranged from 22.4 to 35.5 relative to the base case value of 27.7.

The probabilistic sensitivity analysis results are presented in Figure A.4 of the supplementary material, which shows the number of screening biopsies versus the number of prostate cancer deaths per 1000 men from 30 experiments. The number of prostate cancer deaths ranged from 19.2 to 33.8, while the number of screening biopsies ranged from 196 to 215 per 1000 men. Sensitivity analyses related to screening participation and adherence rates are presented in the supplementary material. We found that nonparticipation had a significantly larger impact on patient outcomes than less than perfect adherence.

DISCUSSION

We developed and validated a new, partially observable Markov model that considers prostate cancer screening and treatment decisions for a cohort of men, starting at age 40, through to the end of life. We used this model to examine alternative choices of 2-stage biomarker-based screening strategies based on newly discovered biomarkers. The screening strategy with a high sensitivity PSA threshold of 2 ng/mL and a second biomarker with high-grade sensitivity and specificity of 0.86 and 0.62, respectively, increased the number of QALYs per 1000 men by 19 QALYs compared with no screening and by 7 QALYs compared with using the PSA test alone with a threshold of 4 ng/mL. Our model predicts 1 prostate cancer death averted per 200 men screened, assuming men were screened annually from age 50 to 75 with a PSA threshold of 4 ng/mL. Gulati and others[23] reported similar findings with a number needed to screen between 186 and 220.

Two recent modeling studies also examined the use of new biomarkers for prostate cancer screening. Birnbaum and others[24] and Heijnsdijk and others[25] evaluated the use of PCA3 and phi, respectively. We build on this previous work by evaluating many new biomarkers head-to-head in the same model, providing useful information when choosing between the many new biomarkers available. Another key difference from both of these studies is that we evaluated how the tradeoff in sensitivity and specificity affects performance of new biomarkers, including hypothetical perfect biomarkers that provide an upper bound on the potential benefits of new biomarkers. Finally, we evaluated the biomarkers in the context of QALYs as well as prostate cancer deaths and number of biopsies per 1000 men.

We found that using an initial PSA threshold with a high sensitivity (2 ng/mL) and a second biomarker that has a high sensitivity (between 0.68 and 0.95) and low to moderate specificity (between 0.36 and 0.78) to high-grade disease appears to maximize expected QALYs. Interestingly, high specificity in the second biomarker test, which is concomitant with low sensitivity, results in significant reduction in QALYs. In our model, there are 2 populations of prostate cancer patients: (1) patients with low-grade disease (Gleason score $\leq$6), and (2) patients with high-grade disease (Gleason score $\geq$7). Patients

with low-grade disease are unlikely to die from prostate cancer and, therefore, are unlikely to benefit from screening. Patients with high-grade cancer are more likely to develop metastatic disease, which causes a prostate cancer death. Thus, biomarker tests for high-grade cancer outperform all-cancer biomarkers for 2 reasons: 1) They are more likely to detect high-grade disease and prevent a prostate cancer death, and 2) these high-grade biomarkers reduce the number of biopsies for patients with low-grade disease reducing the burden of screening on patients that are unlikely to benefit.

In our 1-way sensitivity analysis, we found that other-cause mortality has the greatest impact on the expected increase in QALYs relative to no screening, suggesting that the presence of comorbidity is an important consideration when determining the optimal prostate cancer screening strategy. We found that the results were most sensitive to variation in the QALY disutilities and the metastasis rate for patients with undiagnosed prostate cancer and least sensitive to variation in transition probabilities. In our probabilistic sensitivity analysis, the prostate cancer mortality rate was more sensitive to variation in model parameters than the mean number of biopsies.

Many different screening strategies performed equally well in terms of QALYs; however, we have found that it is possible to distinguish these "equal" screening strategies by looking at additional performance measures that may better account for patient preferences. For example, some strategies that achieved similar QALYs varied significantly in rates of biopsy and prostate cancer deaths, with reductions in prostate cancer deaths coming at the expense of a greater biopsy rate. This tradeoff emphasizes the importance of a shared decision-making approach to account for patient preferences regarding risk of prostate cancer mortality and harms from biopsy.

The hypothetical biomarkers that perfectly detect all cancer and high-grade cancer performed significantly better than screening strategies based on sensitivities and specificities reported in the literature. This suggests there is potential for additional gains from new biomarker discoveries. Interestingly, the high-grade hypothetical perfect biomarker achieved similar rates of prostate cancer mortality when compared with the perfect all-cancer biomarker, while reducing the number of screening biopsies to which patients are subjected. These data suggest that screening biomarkers with an ability to detect high-grade cancers may reduce unnecessary biopsies.

Our study has limitations based on assumptions used in the modeling process. First, estimates of sensitivity and specificity for biomarkers can be dataset dependent, as the estimates come from different datasets and, therefore, may have different biases; however, our analysis still provides useful insights into how the sensitivity and specificity of biomarkers affect long-term health outcomes. Second, we are not aware of any longitudinal studies of long-term health outcomes associated with these new biomarkers. In the absence of data to support correlations between disease status, risk of preclinical progression and recurrence, PSA levels, and new biomarkers operating characteristics, we have assumed no explicit correlations. If correlations exist, this could lead to biased results and conclusions. Third, we assumed that each patient receives at most 1 screening biopsy in his life. About 7%–12% of men undergoing biopsy have had a previous negative biopsy[26,27]; however, the majority of patients receive a single biopsy, and cancers detected on second biopsy are typically less clinically significant. Since our intent is to measure the public health impact of biomarker screening, we do not believe that this assumption significantly influenced our results.

These limitations notwithstanding, a number of conclusions can be drawn from this study. Identifying biomarkers and risk thresholds optimized for identification of high-grade cancers has the greatest impact on measures of performance in the screening setting. Combining new biomarkers with PSA has the potential to reduce the number of screening biopsies (thus decreasing overdiagnosis) and decrease the rate of prostate cancer mortality. The sensitivity analysis suggests that our conclusions are robust with respect to plausible variation in model parameters. New biomarkers with risk thresholds optimized for identification of high-grade cancer can reduce the number of prostate cancer deaths compared with PSA alone while also increasing quality-adjusted survival. These results support prospective clinical-validation trials using rationally selected thresholds in order to design more efficient strategies for the early detection of prostate cancer. We have shown that 2-stage biomarker screening strategies can be beneficial for the early detection of prostate cancer and have provided a foundation for how this approach could potentially be adapted for other types of cancer screening.

ACKNOWLEDGMENTS

This material is based on work supported in part by the National Science Foundation through Grant No. CMMI

1536444 and by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1256260. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors thank Andrew Vickers for providing the data for the lead time distributions reported by Savage and others.[18]

REFERENCES

1. Carter HB, Albertsen PC, Barry MJ, et al. Early detection of prostate cancer: AUA guideline. J Urol. 2013;190(2):419–26.

2. Moyer A. Screening for prostate cancer: US Preventive Services Task Force recommendation statement. Ann Intern Med. 2012;157(2):120–34.

3. Makarov V, Loeb S, Getzenberg RH, Alan PW. Biomarkers for prostate cancer. Annu Rev Med. 2009;60:139–51.

4. Tosoian JJ, Ross AE, Sokoll LJ, Partin AW, Pavlovich CP. Urinary biomarkers for prostate cancer. Urol Clin North Am. 2016;43(1):17–38.

5. Bryant RJ, Lilja H. Emerging PSA-based tests to improve screening. Urol Clin North Am. 2014;41:267–76.

6. Catalona J, Partin W, Sanda G, et al. A multicenter study of [–2] pro-prostate specific antigen combined with prostate specific antigen and free prostate specific antigen for prostate cancer detection in the 2.0 to 10.0 ng/ml prostate specific antigen range. J Urol. 2011;185(5):1650–5.

7. Bussemakers MJ, van Bokhoven A, Verhaegh GW, et al. DD3: A new prostate-specific gene, highly overexpressed in prostate cancer. Cancer Res. 1999;59(23):5975–9.

8. Salagierski M, Schalken JA. Molecular diagnosis of prostate cancer: PCA3 and TMPRSS2: ERG gene fusion. J Urol. 2012; 187(3):795–801.

9. Truong M, Yang B, Jarrard DF. Toward the detection of prostate cancer in urine: a critical analysis. J Urol. 2013;189(2):422–9.

10. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science. 2005;310(5748):644–8.

11. Brenner JC, Chinnaiyan AM, Tomlins SA. ETS fusion genes in prostate cancer. In: Tindall DJ, ed. Prostate Cancer. New York: Springer; 2013.

12. Pettersson A, Graff RE, Bauer SR, et al. The TMPRSS2: ERG rearrangement, ERG expression, and prostate cancer outcomes: a cohort study and meta-analysis. Cancer Epidemiol Biomarkers Prev. 2012;21(9):1497–509.

13. Young A, Palanisamy N, Siddiqui J, et al. Correlation of urine TMPRSS2: ERG and PCA3 to ERG+ and total prostate cancer burden. Am J Clin Pathol. 2012;138(5):685–96.

14. Tomlins A, Day JR, Lonigro J, et al. Urine TMPRSS2:ERG plus PCA3 for individualized prostate cancer risk assessment. Eur Urol. 2016;70(1):45–53.

15. Bratt O, Lilja H. Serum markers in prostate cancer detection. Curr Opin Urol. 2015;25(1):59–64.

16. Gulati R, Inoue L, Katcher J, Hazelton W, Etzioni R. Calibrating disease progression models using population data: a critical precursor to policy development in cancer control. Biostatistics. 2010;11(4):707–19.

17. Epstein I, Feng Z, Trock J, Pierorazio M. Upgrading and downgrading of prostate cancer from biopsy to radical prostatectomy: incidence and predictive factors using the modified Gleason grading system and factoring in tertiary grades. Eur Urol. 2012;61(5):1019–24.

18. Savage CJ, Lilja H, Cronin AM, Ulmert D, Vickers AJ. Empirical estimates of the lead time distribution for prostate cancer based on two independent representative cohorts of men not subject to prostate-specific antigen screening. Cancer Epidemiol Biomarkers Prev. 2010;19(5):1201–7.

19. Liu J, Womble PR, Merdan S, Miller DC, Montie JE, Denton BT. Factors influencing selection of active surveillance for localized prostate cancer. Urology. 2015;86(5):901–5.

20. Cooperberg R, Cowan E, Hilton F, et al. Outcomes of active surveillance for men with intermediate-risk prostate cancer. J Clin Oncol. 2011;29(2):228–34.

21. Ross KS, Carter HB, Pearson JD, Guess HA. Comparative efficiency of prostate-specific antigen screening strategies for prostate cancer detection. JAMA. 2000;284(11):1399–405.

22. Andriole GL, Crawford ED, Grubb III RL, et al. Mortality results from a randomized prostate-cancer screening trial. N Engl J Med. 2009;360(13):1310–9.

23. Gulati R, Mariotto AB, Chen S, Gore JL, Etzioni R. Long-term projections of the harm-benefit trade-off in prostate cancer screening are more favorable than previous short-term estimates. J Clin Epidemiol. 2011;64(12):1412–7.

24. Birnbaum JK, Feng Z, Gulati R, et al. Projecting benefits and harms of novel cancer screening biomarkers: a study of PCA3 and prostate cancer. Cancer Epidemiol Biomarkers Prev. 2015;24(4): 677–82.

25. Heijnsdijk EA, Denham D, de Koning HJ. The cost-effectiveness of prostate cancer detection with the use of prostate health index. Value Health. 2015;24(4):677–82.

26. Nguyen CT, Yu C, Moussa A, Kattan MW, Jones JS. Performance of Prostate Cancer Prevention Trial risk calculator in a contemporary cohort screened for prostate cancer and diagnosed by extended prostate biopsy. J Urol. 2010;183(2):529–33.

27. Thompson M, Ankerst P, Chi C, et al. Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. J Natl Cancer Inst. 2006;98(8):529–34.

28. Haas GP, Delongchamps NB, Jones RF, et al. Needle biopsies on autopsy prostates: sensitivity of cancer detection based on true prevalence. J Natl Cancer Inst. 2007;99(19):1484–9.

29. Arias E. United States life tables, 2006. Natl Vital Stat Rep. 2010;58(21):1–40.

30. Ries LAG, Young JL, Keel GE, Eisner MP, Lin YD, Horner MJ. Cancer Survival among Adults: US SEER Program, 1988–2001: Patient and Tumor Characteristics. Bethesda (MD): US Department of Health and Human Services, National Institutes of Health, National Cancer Institute; 2007.

31. Draisma G, Boer R, Otto SJ, et al. Lead times and overdetection due to prostate-specific antigen screening: estimates from

the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst. 2003;95(12):868–78.

32. Roehl KK, Han M, Ramos CG, Antenor JAV, Catalona WJ. Cancer progression and survival rates following anatomical radical retropubic prostatectomy in 3,478 consecutive patients: long-term results. J Urol. 2004;172(3):910–4.

33. Heijnsdijk EA, Wever EM, Auvinen A, et al. Quality-of-life effects of prostate-specific antigen screening. N Engl J Med. 2012; 367(7):595–605.

34. Parekh J, Punnen S, Sjoberg D, et al. A multi-institutional prospective trial in the USA confirms that the 4Kscore accurately identifies men with high-grade prostate cancer. Eur Urol. 2015; 68(3):464–70.

35. Ferro M, Bruzzese D, Perdonà S, et al. Predicting prostate biopsy outcome: prostate health index (phi) and prostate cancer antigen 3 (PCA3) are useful biomarkers. Clin Chim Acta. 2012; 413(15):1274–8.

36. Salami SS, Schmidt F, Laxman B, et al. Combining urinary detection of TMPRSS2:ERG and PCA3 with serum PSA to predict diagnosis of prostate cancer. Urol Oncol. 2013;31(5):566–71.

37. Sartori DA, Chan W. Biomarkers in prostate cancer: what's new? Curr Opin Oncol. 2014;26(3):259.