

A surveillance model of prostate cancer trends informs PSA screening policies

Ruth Etzioni^{1,2,*}

Roman Gulati¹

Jeffrey Katcher¹

Lurdes Y. T. Inoue^{1,2}

¹Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

²University of Washington, Seattle, Washington, USA

Presented at the International Microsimulation Association conference in Ottawa, Canada, June 8–10, 2009

Abstract

Comparative cost-effectiveness is imperative to fairly balancing the benefits and harms of candidate prostate-specific antigen (PSA) screening policies. We present outcomes-based costs and benefits projected by a microsimulation model of prostate cancer incidence and mortality trends in the US. The model links PSA growth with cancer progression events. Individuals are assigned log-linear PSA growth with a change point at disease onset; intercept and slope parameters are estimated using data from the Prostate Cancer Prevention Trial. Cancer progresses from localized and asymptomatic to metastatic and symptomatic states with progression rates that depend on PSA levels. Progression rates are estimated using a simulated maximum likelihood approach that calibrates model projections of age-, year-, and stage-specific incidence with observed incidence trends from Surveillance, Epidemiology, and End Results registries. Cancer-specific survival is incorporated via a constant and a PSA-based cure rate, each calibrated to reproduce prostate cancer deaths reported in the Prostate, Lung, Colorectal, and Ovarian Screening Trial. Using the calibrated model, we evaluate benefits (early detections, life saved) and costs (overdiagnosis) for each combination of the following policy components: PSA test-positive threshold of 2.5 or 4.0 ng/mL; annual, biennial, or quadrennial PSA screening; screening ages 50–69, 50–74, or 50–79 years. We rank the 18 candidate policies using four outcomes-based cost-effectiveness ratios (years of overdiagnosis to years of life saved and to years of early detection, overdiagnoses to individuals with life savings and to early detections) and identify policies that dominate across performance measures.

Keywords: *Cancer control; cost-effectiveness; prostate cancer; overdiagnosis; PSA screening; simulated maximum likelihood; surveillance model*

*Corresponding author. Address: Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, M2-B230, Seattle WA 98109-1024, USA. Tel: +1.206.667.6561. Fax: +1.206.667.7264. Email: retzioni@fhcrc.org.

1 Introduction

Prostate cancer directly affects many men. In 2008 in the US, 186,320 new diagnoses were added to the more than 2 million men alive on January 1, 2008, who had a history of the disease. Based on the most recent data available, the National Cancer Institute projects that 1 in 6 men born today will be diagnosed with prostate cancer some time in their lifetime (Horner et al., 2009).

It is widely recognized that prostate-specific antigen (PSA) screening is responsible for a significant portion of the high incidence of prostate cancer seen in the US today. PSA screening began in the late 1980s and was rapidly adopted in the early 1990s, coincident with doubling of incidence rates among men over 50 years. Prostate cancer incidence rates have since fallen from their peak but remain above the historic pre-PSA levels.

Yet the benefits of PSA screening remain unclear. While prostate cancer mortality has declined by approximately 4% per year since 1992, continuing debate surrounds its role in this decline. Recently, results from two large-scale randomized trials of PSA screening, the Prostate, Lung, Colorectal, and Ovarian (PLCO) Screening Trial in the US (Andriole et al., 2009) and the European Randomized study of Screening for Prostate Cancer (ERSPC) (Schröder et al., 2009), were released. Briefly, the PLCO trial found no difference in the rates of death from prostate cancer in men who underwent annual PSA screening compared with men who were assigned to usual care. The ERSPC found that PSA screening every 4 years (every 2 years in the Swedish study center) reduced the rate of death from prostate cancer by 20% compared with men who received no screening. The difference between the two findings is likely due to differing levels of screening in the reference groups, test-positive thresholds, biopsy compliance rates, and levels of aggressiveness in looking for prostate cancer prior to the trials.

Importantly, both studies note the high rates of overdiagnosis associated with PSA screening. Overdiagnosis (i.e., PSA detection of prostate cancer that would not present clinically before death from other causes) represents the primary driver of costs of PSA screening strategies. Unfortunately, overdiagnosis rates cannot be estimated directly since once a man is detected through PSA screening we cannot observe whether or not he would have presented clinically in his lifetime. Using common data on US population incidence trends, three modeling groups estimated overdiagnosis rates to be between 23% and 42% of PSA detections (Draisma et al., 2009). Similarly, early detections of prostate cancer (i.e., PSA detection of prostate cancer in a localized stage that would have presented clinically in an advanced stage) and associated years of life saved, primary drivers of benefits associated with screening strategies, cannot be observed directly. Projection of these outcomes associated with observed PSA screening practices is an ongoing area of research for modeling groups.

Estimating outcomes such as overdiagnosis, early detections, and life saved is challenging since it requires a model of underlying cancer progression that captures the heterogeneity known to exist in the population. While the trial results provide essential information for calibrating and validating such a model, it should be emphasized that they do not necessarily extend to the population, do not directly provide information about alternative PSA screening policies, and do not permit inference concerning cost-effectiveness of competing PSA screening strategies. *Yet knowledge about population-level cost-effectiveness measures of candidate screening policies is essential for making informed cancer control policy decisions.* Even population-representative trials cannot provide this knowledge for two main reasons. First, even if additional trials were undertaken to consider alternative PSA screening policies, there are more distinct policies in terms of screening frequencies, screening ages, and test-positive criteria than can be practically studied. Second, trials are fundamentally limited as a

method for estimating unobservable costs and benefits associated with PSA screening. In contrast, population surveillance modeling can permit evaluation of many PSA screening policies and project associated unobservable costs and benefits provided it (a) links PSA growth with the underlying cancer progression and (b) accurately represents the disease process in the population.

With this motivation, our primary aim in this paper is to demonstrate the value of a microsimulation model that links PSA growth with prostate cancer progression to assess outcomes-based cost-effectiveness measures among competing PSA screening policies. We adapt a model previously developed (Inoue et al., 2008) and calibrated to the US population (Gulati et al., 2009) to project overdiagnosis, early detections, and years of life saved for men aged 50–84 across 18 PSA screening policies. We also examine robustness of policy rankings and discuss policy implications.

2 Methods

To draw inference about outcomes associated with competing PSA screening policies, we rely on a microsimulation model of men aged 50–84 in the US population. Each man is assigned log-linear PSA growth over time, and transitions from one disease state to the next depend on his PSA trajectory. This model was previously fit to the Baltimore Longitudinal Study of Aging (Inoue et al., 2008) and extended to the US population (Gulati et al., 2009). Here we briefly review the model and describe extensions to project cancer-specific survival and derive outcomes-based cost-effectiveness measures. Additional model details are provided in the PSAPC model profile available at <http://cisnet.cancer.gov/profiles>.

2.1 Natural history model

2.1.1 PSA growth

We assume PSA growth is log-linear in age with a change point at disease onset. Distributions on the intercept and slope parameters capture between-individual variability. We formalize this for the j th man as:

$$\log(y_j(t)) = \beta_{0j} + \beta_{1j}t + \beta_{2j}(t - t_{oj})^+ + \varepsilon \quad (1)$$

where $y_j(t)$ is PSA at age t , t_{oj} is age at disease onset, $\beta_{0j} \sim N(\mu_0, \sigma_0^2)$ is the log PSA intercept at age 35, $\beta_{1j} \sim N^+(\mu_1, \sigma_1^2)$ is the log PSA slope before disease onset, $\beta_{2j} \sim N^+(\mu_2, \sigma_2^2)$ is the increment to the log PSA slope after disease onset, and $\varepsilon \sim N(0, \tau^2)$ reflects unexplained variation in log PSA measurements. (Here the superscript $+$ denotes truncation at zero.) Parameters were estimated using longitudinal PSA test results from the control arm of the Prostate Cancer Prevention Trial (PCPT), which screened 18,882 men for up to seven years. Two parameters—pre-onset log PSA intercept (at age 35) and slope—were further tuned using PSA test positivity and cancer detection rates from the initial screening round of the PLCO trial to overcome identifiability challenges to estimation.

2.1.2 Cancer progression

Cancer progression occurs based on hazard functions that increase with PSA levels. Hazards of metastasis and clinical diagnosis are proportional to noise-free PSA for the j th man, denoted $\tilde{y}_j(\cdot)$. To reflect increased risk of

symptoms as disease advances, the hazard rate for clinical diagnosis may increase at metastasis. In addition, a hazard of disease onset is proportional to age. These hazard functions are formalized as follows:

$$\lambda_{oj}(t) = \gamma_o t \quad (2)$$

$$\lambda_{mj}(t) = \gamma_m \tilde{y}_j(t) \quad (3)$$

$$\lambda_{cj}(t) = \begin{cases} \gamma_c \tilde{y}_j(t) & t \leq t_{mj} \\ \theta_c \gamma_c \tilde{y}_j(t) & t > t_{mj} \end{cases} \quad (4)$$

where t_{mj} represents age at transition from localized to metastatic disease and θ_c denotes the multiplier in the hazard of clinical diagnosis following transition to metastatic disease. Parameters were estimated using a microsimulation framework (described next) that calibrates model-projected age-, year-, and stage-specific incidence with observed data from the core registries of the Surveillance, Epidemiology, and End Results (SEER) program. We develop a Poisson likelihood for observed incidence and estimate the parameters using a Nelder-Mead simplex algorithm (Nelder and Mead, 1965) adapted to a stochastic likelihood setting (Spall, 2003). Gulati et al. (2009) provide additional details.

2.2 Microsimulation framework

The conceptual model outlined above is implemented in a microsimulation framework as follows. A simulated population of individuals is generated to match observed counts of men in the SEER registries; this step involves generating dates of birth and death that replicate age cohorts in the target population using US life tables. Simulated individuals are then randomly assigned PSA growth curves from the PSA growth model. Given individual PSA growth curves, we randomly assign ages at natural/clinical history events using the probability integral transform corresponding to the hazard functions in (2) to (4). These individual natural/clinical histories represent the model-projected collection of courses of cancer progression in the absence of PSA screening.

Over this simulated population we first superimpose observed PSA screening patterns. Since PSA screening practices were not tracked in real time, we rely on a model of testing patterns constructed using data from the linked SEER-Medicare database to model ages at first tests and from the National Health Interview Survey to model intervals between tests (Mariotto et al., 2007). This model of PSA screening generates ages at PSA tests, and test positivity occurs when a simulated individual’s PSA growth curve exceeds 4.0 ng/mL at a PSA test (a standard test-positive threshold in the 1990s). When an individual tests positive, he is referred to biopsy. Biopsy compliance is determined by a randomly generated indicator based on age- and PSA-specific biopsy compliance rates observed in the PLCO trial. Finally, if an individual undergoes biopsy and has already had onset of prostate cancer, whether the cancer is detected is determined by another randomly generated indicator based on patterns of increasing numbers of biopsy cores over calendar years.

Combining clinical and screen detections under observed PSA screening and biopsy practices produces total model-projected incidence counts by age, year, and disease stage. These model-projected incidence counts are compared with corresponding observed SEER counts via a simulated Poisson likelihood, and a variant of the Nelder-Mead simplex algorithm (Nelder and Mead, 1965; Spall, 2003) is employed to estimate the cancer progression parameters $(\lambda_o, \lambda_m, \lambda_c, \theta_c)$ conditional on the PSA growth model.

Following calibration of the cancer progression model to the general population, we replace observed PSA screening practices with alternative PSA screening patterns. Since we now have a model of cancer progression

that satisfactorily reflects heterogeneity of the disease observed in the population and since we know the PSA level of each simulated individual at any point in time, we can consider hypothetical screening frequencies, age ranges, and test-positive criteria.

2.3 Incorporating cancer-specific survival

We extend the incidence model by adding cancer-specific survival to project measures of life saved under different screening protocols. To do so, it is necessary to incorporate both Gleason grade and treatment, since these represent important prognostic variables associated with cancer-specific survival. We randomly assign high (SEER poorly differentiated or undifferentiated) or low (SEER well-differentiated or moderately differentiated) grade at disease onset based on the observed SEER distribution in the PSA era, and we randomly assign treatment as one of conservative management, radiation therapy, or radical prostatectomy based on the unconditional treatment distribution in 1999. Then, conditional on age, stage, grade, and treatment, cancer-specific survival following clinical diagnosis is randomly assigned to individuals based on SEER survival among cases diagnosed just prior to the PSA era (i.e., cases diagnosed in 1983–1986). While we recognize that survival based on 1983–1986 data may not accurately reflect the survival that would be expected in the absence of screening in later years, these data are the best available source of information on survival that is not contaminated by the lead time and overdiagnosis associated with screening. Consequently we consider a modified version of these data that allows improvements in survival among clinically detected cases over time due to advances in diagnostics, monitoring of diagnosed disease, and treatment. Then, individuals assigned curative treatment enjoy a survival benefit (with hazard ratio 0.56) relative to untreated cases (Bill-Axelsson et al., 2005).

To project cancer-specific survival following a screen detection, we consider two cure rate assumptions. Under the first assumption, a constant cure rate p applies to all individuals who would have died of their disease in the absence of screening, and age at cancer-specific death following screen detection is randomly assigned (with probability p) to be the age at other-case death. Under the second assumption, this random assignment depends on individual PSA at screen detection relative to PSA at clinical diagnosis and to the difference between PSA at onset and PSA at clinical diagnosis. This assumption allows for a higher probability of cure among men who are detected earlier or who have slower cancer progression. This is operationalized for the j th individual as:

$$p_j = \left(\frac{\Delta_L \tilde{y}_j}{\Delta_S \tilde{y}_j} \right)^\alpha, \quad (5)$$

where $\Delta_L \tilde{y}_j = \tilde{y}_j(t_{c_j}) - \tilde{y}_j(t_{s_j})$ denotes the difference in noise-free PSAs at clinical diagnosis and at screen detection (i.e., the change in PSA over the lead time) and $\Delta_S \tilde{y}_j = \tilde{y}_j(t_{c_j}) - \tilde{y}_j(t_{o_j})$ denotes the difference in noise-free PSAs at clinical diagnosis and at onset (i.e., the change in PSA over the sojourn time). Hence this ratio reflects how early screen detection occurs (on the PSA scale) relative to PSA growth in the latent disease period. While this approach is qualitatively intuitive, it may not be quantitatively correct. Consequently, we calibrate a tuning parameter $0 \leq \alpha \leq 1$ using a least squares criterion comparing model-projected prostate cancer deaths with observed prostate cancer deaths reported from the screening arm of the PLCO trial (Andriole et al., 2009). The criterion is minimized using the `optimize` function in the R statistical computing environment (R Development Core Team, 2009), a modified golden section search (MGSS) algorithm. The constraint that $0 \leq \alpha \leq 1$ forces a rapid decline in the cure rate as PSA at screen detection approaches PSA at clinical diagnosis (or, equivalently, as lead time approaches zero). While the constant cure rate is simpler, the PSA-based cure

rate allows for greater life saved associated with early detection.

2.4 Outcomes-based cost-effectiveness measures

To demonstrate comparison of alternative PSA screening policies, we project counts and years of overdiagnoses, early detections, and life savings for hypothetical cohorts of men in the US population tracked for 30 years. These outcomes are examined individually and combined in the following four ratios:

- O/ED: Counts of overdiagnoses to counts of early detections
- O/LS: Counts of overdiagnoses to counts of men with any life savings
- YO/YED: Years of overdiagnosis to years of early detections
- YO/YLS: Years of overdiagnosis to years of life saved

Consideration of both ratios based on counts and ratios based on durations of events is potentially important for understanding both how many individuals are impacted by a given policy and the sizes of the impact.

3 Results

We obtained preliminary estimates of PSA growth parameters using PCPT data from 1,022 cancer cases and 7,058 non-cases determined by end-of-study biopsy. Linear mixed-effects models were fit to the last available three to four tests for cases and to all tests for non-cases. When we combined resulting PSA growth estimates with preliminary cancer progression estimates, we found that projected age-adjusted test-positivity and cancer detection rates (25% and 31%) did not match values reported in the initial screening round of the PLCO (8% and 40%) (Andriole et al., 2005). Investigating this lack-of-fit we found that several combinations of μ_0 and μ_1 led to similar incidence projections. Our final selected values of these parameters preserve a close agreement with observed incidence while projecting PSA test-positivity and cancer detection rates of 11% and 57%. Note that we expect our projections to be slightly higher than trial values since we are simulating de novo screening while many trial participants had undergone previous screens (Andriole et al., 2005).

Table 1 presents both preliminary and final PSA growth estimates. The final pre-onset slope μ_1 corresponds to a mean annual percent change in PSA of approximately 2% in the absence of cancer, which is consistent with prior studies (e.g., Oesterling et al., 1993; Ellis et al., 2001). The post-onset slope $\mu_1 + \mu_2$ corresponds to a mean annual percent change in PSA for prostate cancer cases of approximately 14%. This is lower than the annual percent change in PSA among cases in previously published stored serum studies (Whittemore et al., 1995; Inoue et al., 2004), but these studies included many pre-PSA-era patients; in contrast, PCPT participants underwent screening and end-of-study biopsies with the majority of PCPT cases diagnosed in the absence of elevated PSA levels. Between-subject variability is considerably lower than within-subject variability, which is higher for cancer cases than non-cases.

Cancer progression parameter estimates are also reported in Table 1. To reflect uncertainty in the estimates due to the simulation framework, we re-estimated these parameters across 20 random seeds. The greatest uncertainty is associated with the multiplier for the clinical detection hazard for distant stage cancers (θ_c). This parameter is difficult to estimate precisely since a small increase when θ_c is large advances clinical diagnosis of a distant stage cancer a matter of days and so has only a small impact on the likelihood.

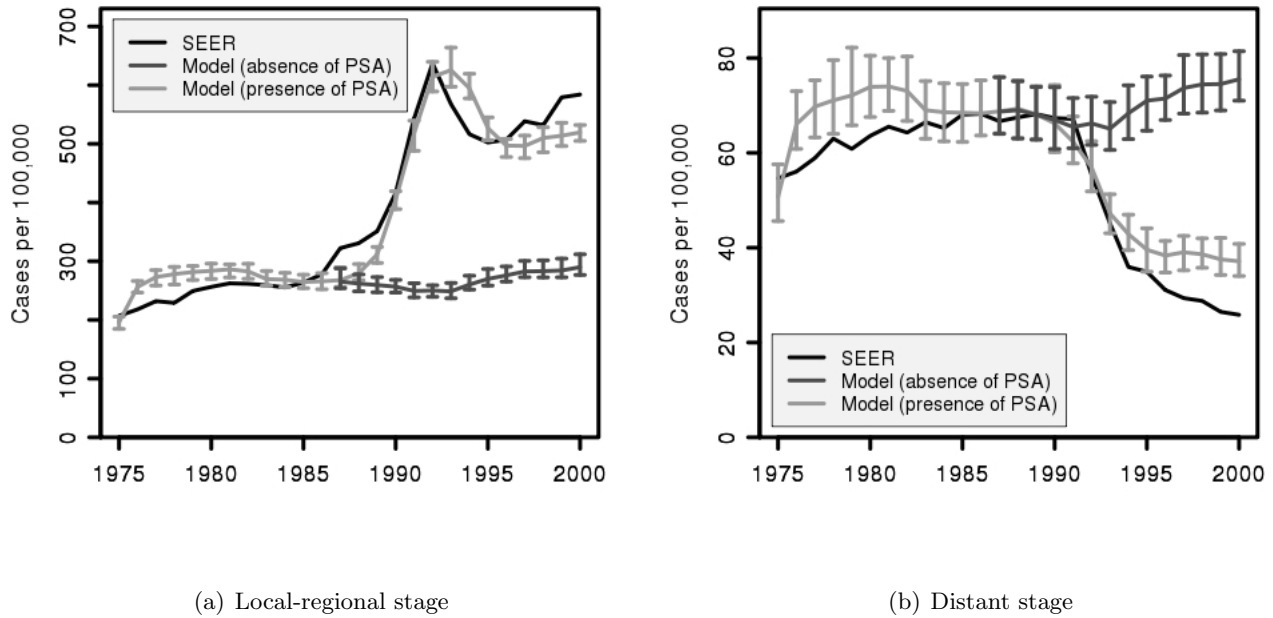
Table 1: Log PSA growth and cancer progression model parameter estimates

Log PSA growth parameters			
Parameter	Description	Posterior Mean	Posterior 95% CI
μ_0	Pre-onset intercept mean (preliminary)	-1.2720	(0.8047, 1.7393)
μ_1	Pre-onset slope mean (preliminary)	0.0443	(0.0431, 0.0455)
μ_0	Pre-onset intercept mean (final)	-1.6094	(1.1421, 2.0767)
μ_1	Pre-onset slope mean (final)	0.0200	(0.0188, 0.0212)
μ_2	Post-onset slope increment mean	0.1094	(0.0919, 0.1269)
σ_0^2	Pre-onset intercept variance	0.0568	(0.0480, 0.0656)
σ_1^2	Pre-onset slope variance	0.0019	(0.0018, 0.0020)
σ_2^2	Post-onset slope increment variance	0.0237	(0.0202, 0.0272)
τ^2	Within-individual variance	0.0829	(0.0817, 0.0841)
Cancer progression parameters			
Parameter	Description	Simplex Mean	Simplex 95% CI
γ_o	Hazard of disease onset	0.0005	(0.0004, 0.0005)
γ_m	Hazard of transition to metastatic disease	0.0004	(0.0004, 0.0005)
γ_c	Hazard of clinical diagnosis	0.0015	(0.0013, 0.0017)
θ_c	Multiplier for hazard of clinical diagnosis	19.1334	(4.2105, 24.0563)
Cause-specific survival parameters			
Parameter	Description	MGSS Mean	MGSS 95% CI
p	Constant cure rate	0.5469	(0.2178, 0.7692)
α	PSA-based cure rate exponent	0.6784	(0.3177, 0.9525)

The calibrated model projections averaged across random seeds reproduces observed incidence trends quite well; observed and projected age-adjusted incidence by year and stage are presented in Figure 1. Confidence interval estimates are shown in each year based on uncertainty due to the random seed. While the model overprojects slightly in the late 1970s for both stages and in the late 1990s for distant stage, the projections generally capture the global shape of observed incidence. The figure also shows projected incidence in the absence of PSA screening, which suggests a relatively constant secular trend continuing at the last level observed before PSA testing was introduced in 1986. This constant secular trend is consistent with that estimated by Etzioni et al. (2008) and Telesca et al. (2008) using different modelling approaches.

Model-projected natural/clinical history measures are broadly consistent with values reported in the literature. For example, we estimate that the lifetime probability of disease onset is 35% while 36% was estimated by Etzioni et al. (1998) based on autopsy studies. We estimate that mean lead time for men aged 50–84 who would have presented clinically in the absence of screening is 7 years, which agrees with other published estimates (Gann et al., 1995; Etzioni et al., 1998; Draisma et al., 2003; Tsodikov et al., 2006; Draisma et al., 2009).

Figure 1: Observed age-adjusted prostate cancer incidence by stage with mean projections and 95% confidence intervals in the presence and absence of PSA screening.



When we project cancer-specific survival based on SEER data among cases diagnosed in 1983–1986 we find that the resulting mortality under screening considerably overestimates that observed in the PLCO trial. We therefore inflate the original SEER survival curves via a range of hazard ratios (ranging from 0.3 to 0.7). The hazard ratios used for model projection are selected as follows. (1) For each hazard ratio, estimate the cure rate associated with screening that produces mortality results that are most consistent with those observed in the screening arm of the PLCO trial (Andriole et al., 2009). (2) For each estimated cure rate, compute the corresponding rate ratio for cancer death that would result from a comparison of a screened versus a control arm in a randomized screening trial with screening patterns similar to the PLCO trial. (3) Select the cure rates (and corresponding hazard ratios) that produce rate ratios consistent with those observed in the ERSPC trial. Thus, in this procedure we are essentially calibrating our survival model to the data observed in the screening arm of the PLCO trial, and our screening efficacy to the results of the European trial. We cannot calibrate screening efficacy to the PLCO trial, even though it was a US study, because of the contamination observed in the control arm (more than half of control group participants were screened).

By this selection process, we find that a hazard ratio of 0.5 for improving baseline cancer-specific survival allows for a reasonable approximation to prostate cancer deaths reported in the PLCO trial. Under the constant cure rate assumption, we estimate that just over half of individuals who would have died of prostate cancer in the absence of screening are cured by early detection. Under the PSA-based cure rate, more (fewer) individuals are cured when PSA at screen detection is 40% or less (more) of PSA at clinical diagnosis relative to the change in PSA over the latent disease period; see Figure 2. Calibrated cure rate parameters project numbers of prostate cancer deaths that agree with those reported in the PLCO trial (Andriole et al., 2009); see Figure 3. The implied rate ratio associated with annual screening for six years relative to no screening is 34%.

Figure 2: Calibrated cure rates by ratio of change in PSA over lead time to change in PSA over sojourn time

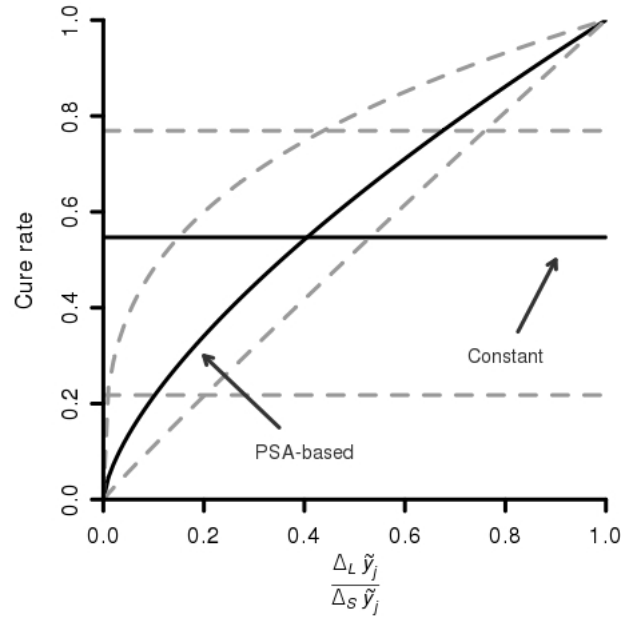
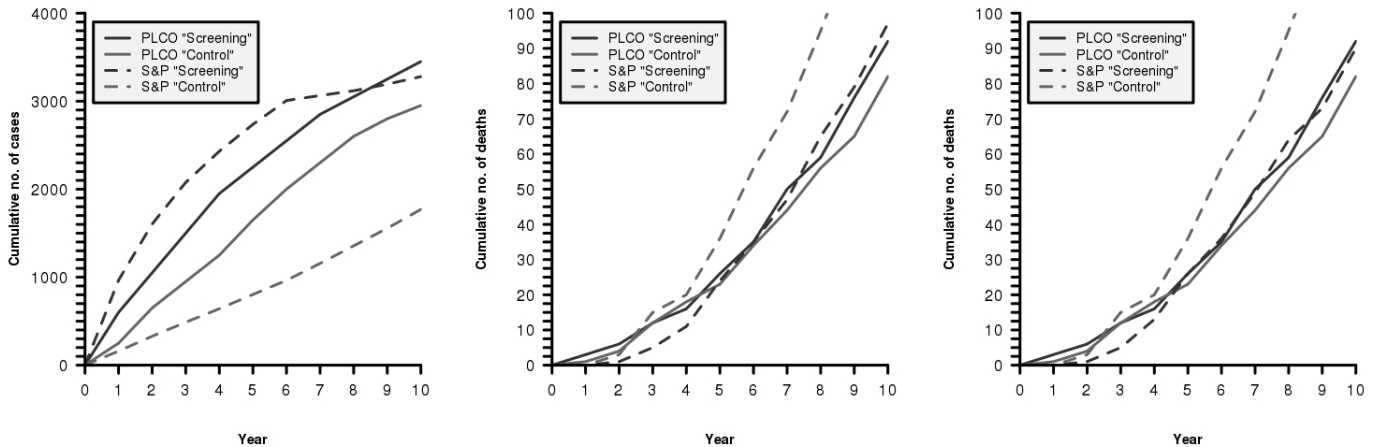


Figure 3: Observed counts of prostate cancers and prostate-cancer deaths from the screening and control arms of the PLCO trial with corresponding model projections under annual PSA screening and under no PSA screening



(a) Prostate cancers

(b) Prostate-cancer deaths under a constant cure rate

(c) Prostate-cancer deaths under a PSA-based cure rate

3.1 Policy outcomes comparison

Table 2 presents outcomes projected by the model for the 18 PSA screening policies averaged across random seeds. Several points are worth noting. First, overdiagnosis and early detections generally increase with more aggressive PSA screening policies, whether this means screening older men or using a lower test-positive threshold for referral to biopsy. The same pattern can be seen for years of overdiagnosis and for years of early detection. Second, while cases with life savings and years of life saved are higher under the PSA-based cure rate assumption than under the constant cure rate assumption, the incremental benefits are comparable. Consequently, rankings

of the 18 policies are highly consistent across the four outcomes-based cost-effectiveness ratios. In light of this high degree of consistency, we present only rankings based on ratios of years of overdiagnosis to years of life saved (YO/YLS) under the constant cure rate assumption (rankings are identical under the PSA-based cure rate assumption). The three most cost-effective policies correspond to quadrennial, biennial, and annual screening of men aged 50–69 with test-positive threshold 4.0 ng/mL. The next three most cost-effective policies involve quadrennial or biennial screening of men aged 50–69 or 50–74 and a test-positive threshold 2.5 or 4.0 ng/mL.

Figure 4 presents selected outcomes-based cost-effectiveness measures coded by policy components. Figure 4(a) illustrates how overdiagnosis and early detections increase with more frequent screening, a lower test-positive threshold, and inclusion of older men. Figure 4(b) similarly illustrates years of overdiagnosis and years of life saved under the constant cure rate assumption. Figure 4(c) presents these outcomes under the PSA-based cure rate assumption.

Implications of the projected outcomes are consistent with recently updated recommendations by the US Preventive Services Task Force (US Preventive Services Task Force, 2008), which advises men to stop screening at age 75 in light of evidence of harms relative to plausible benefits. For example, under annual screening with test-positive threshold 4.0 ng/mL, we find that screening men 50–74 instead of 50–79 results in 471,670 fewer overdiagnoses at a cost of 71,905 fewer early detections over 30 years. Alternatively, this restriction foregoes 2.8 million years of overdiagnosis for 0.15 (constant cure rate) or 0.18 (PSA-based cure rate) million years of life saved, which is justified when each year of life saved is worth less than 15 (PSA-based cure rate) to 19 (constant cure rate) years of overdiagnosis.

The model projections also imply that the net gains associated with lowering the test-positive threshold from 4.0 to 2.5 ng/mL, irrespective of screening frequency or screening ages, greatly increases the harms while increasing the benefits only a small amount. For example, under annual screening of men aged 50–74, we find that lowering the test-positive threshold from 4.0 to 2.5 ng/mL adds 39,950 early detections at a cost of 288,060 additional overdiagnoses. Alternatively, this change adds 0.17 (constant cure rate) or 0.22 (PSA-based cure rate) million years of life saved in exchange for 3.1 million years of overdiagnosis, which corresponds to 14 (PSA-based cure rate) to 19 (constant cure rate) years of overdiagnosis for each year of life saved. Consequently, by roughly the same valuation of a year of life saved relative to a year of overdiagnosis that supports discontinuation of screening after age 75, we conclude that lowering the test-positive threshold from 4.0 to 2.5 ng/mL is not cost-effective in terms of these outcomes.

4 Discussion

This paper describes an extension of a microsimulation model to project outcomes-based cost-effectiveness measures to evaluate selected PSA screening policies. In the basic model, individual-specific PSA growth trajectories determine natural/clinical history events, and parameters are estimated using a simulated likelihood based on observed incidence trends and PSA testing patterns. In the extended model, Gleason grade at onset, treatment at diagnosis, and baseline cancer-specific survival following clinical diagnosis are randomly assigned based on observed distributions in the population. To reflect improvements in prostate cancer survival over the PSA era, we calibrate a hazard ratio to improve baseline cancer-specific survival in the pre-PSA era. To incorporate cancer-survival following screen detection, we consider two cure rate assumptions—a constant cure rate for all

Table 2: Projected outcomes associated with selected PSA screening policies (million) and example ranking

Policy components			Incidence outcomes				Survival outcomes				Example ranking
							Constant ^a		PSA-based ^b		
Frequency	Ages	Threshold	O ^c	YO ^d	ED ^e	YED ^f	LS ^g	YLS ^h	LS	YLS	YO/YLS
Annual	50–69	4.0	0.36	2.71	0.21	1.87	0.19	1.57	0.24	1.91	3
Annual	50–69	2.5	0.56	4.97	0.25	2.50	0.22	1.75	0.27	2.13	9
Annual	50–74	4.0	0.61	4.41	0.28	2.32	0.22	1.75	0.27	2.13	8
Annual	50–74	2.5	0.90	7.55	0.32	3.02	0.25	1.92	0.31	2.35	15
Annual	50–79	4.0	1.08	7.20	0.35	2.79	0.25	1.89	0.31	2.31	14
Annual	50–79	2.5	1.49	11.29	0.39	3.50	0.28	2.06	0.34	2.52	18
Biennial	50–69	4.0	0.26	1.87	0.17	1.37	0.15	1.24	0.18	1.51	2
Biennial	50–69	2.5	0.41	3.53	0.20	1.88	0.17	1.41	0.21	1.72	7
Biennial	50–74	4.0	0.44	3.07	0.21	1.70	0.18	1.39	0.21	1.69	5
Biennial	50–74	2.5	0.66	5.38	0.25	2.26	0.20	1.56	0.24	1.90	11
Biennial	50–79	4.0	0.88	5.61	0.29	2.16	0.21	1.54	0.25	1.87	12
Biennial	50–79	2.5	1.22	8.86	0.32	2.75	0.23	1.71	0.28	2.09	17
Quadrennial	50–69	4.0	0.19	1.34	0.13	1.01	0.12	0.96	0.14	1.16	1
Quadrennial	50–69	2.5	0.30	2.56	0.15	1.40	0.14	1.10	0.17	1.34	6
Quadrennial	50–74	4.0	0.32	2.22	0.16	1.26	0.14	1.08	0.17	1.30	4
Quadrennial	50–74	2.5	0.49	3.92	0.19	1.70	0.16	1.23	0.19	1.49	10
Quadrennial	50–79	4.0	0.73	4.48	0.24	1.68	0.17	1.22	0.20	1.48	13
Quadrennial	50–79	2.5	1.01	7.11	0.27	2.16	0.19	1.38	0.23	1.66	16

^a Constant: constant cure rate assumption

^b PSA-based: PSA-based cure rate assumption

^c Overdiagnosis: number of PSA detected cases who would not have presented clinically

^d Years of overdiagnosis: years from PSA detection to other-cause death among overdiagnosed cases

^e Early detections: number of PSA detected local-regional cases who would have presented clinically with advanced disease

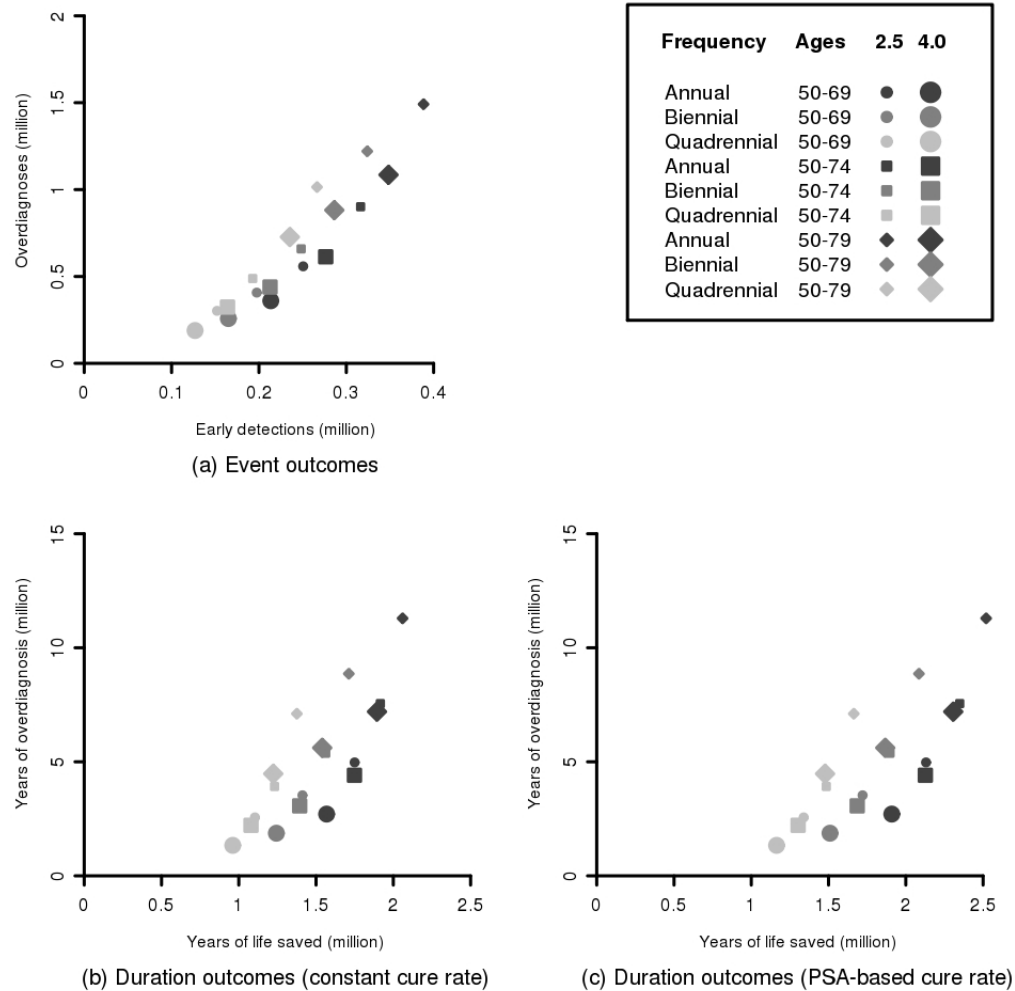
^f Years of early detections: years from PSA detection to clinical diagnosis among early detections

^g Cases with life savings: number of cases who live longer as a result of PSA detection

^h Years of life saved: years from death in the absence of PSA testing to death in the presence of PSA testing

individuals and a PSA-based cure rate with higher chances of cure for earlier detections or slower-growing cancers. Cure rate parameters are calibrated based on prostate cancer deaths from the screening arm of the PLCO trial. The model reproduces incidence and mortality patterns observed in both the population and trial settings and validates well in terms of key natural/clinical history measures.

Figure 4: Projected outcomes associated with selected PSA screening policies



The model presented here has the potential to be useful for policymakers and guidelines panels for three primary reasons: (1) it explicitly links PSA growth with cancer progression, (2) it is calibrated to the population and therefore likely to project representative population outcomes under different screening policies, and (3) it is designed to quantify clinically relevant benefits and harms essential to informing cancer control policy decisions. However, like all models, there are important limitations. First, estimates of the progression rates from latent to clinical and from localized to metastatic states are conditional on inputs that are themselves estimated from a variety of sources. For example, the PSA growth model and its variance components are derived from studies of the PSA distribution in younger men and the PCPT. Each of these data sources is fairly representative of the population, so that the statistical errors in the parameter estimates appear to be small. The inputs relating to the practice of PSA screening in the population have all been estimated or retrospectively inferred based on observational data or published studies. The cancer progression estimates in Table 1 are all conditional on the frequency of PSA use, the probability of a biopsy, and the biopsy sensitivity, which is tied to the number of biopsy cores. Since these quantities are almost certainly estimated with some error, we are understating the noise in the model results. While we do not believe that this noise will substantially affect the rankings of candidate policies, it may affect our uncertainty in the sizes of the differences between policies, and this must be kept in

mind when choosing between competing policies that yield similar outcomes. Finally, functional relationships specified in the model are also subject to uncertainty. For example, we have specified that the risks of metastasis and clinical diagnosis depend on PSA (and not some function of PSA), that subject-specific PSA slopes have truncated normal distributions (and not some other positive distributions), and that random draws determining compliance at consecutive PSA screens are independent (and not correlated with some correlation structure). While we considered many more model variants than we have reported here, there is no guarantee that our final model structure is the best of a class.

Anticipated future extensions of the model framework will include allowing for grade-dependent PSA growth trajectories, introducing a finer stage breakdown, and incorporating economic costs. This model is one of few emerging model frameworks for projecting the benefits and harms associated with competing PSA screening policies, and we are confident that it will become an important tool for evidence-based policymaking.

Acknowledgements

We thank Eric Feuer, Angela Mariotto, and members of the CISNET prostate group for support and feedback. We thank Georg Luebeck for providing source code used for likelihood maximization. We also thank Paul Pinsky, Catherine Tangen, and Phyllis Goodman for providing extended PCPT and PLCO datasets.

This research was supported by Award Numbers U01CA88160 and R01CA131874 of the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

References

- Andriole, G. L., Grubb, R. L., Buys, S. S., Chia, D., Church, T. R., Fouad, M. N., Gelmann, E. P., Kvale, P. A., Reding, D. J., Weissfeld, J. L., Yokochi, L. A., Crawford, E. D., O'Brien, B., Clapp, J. D., Rathmell, J. M., Riley, T. L., Hayes, R. B., Kramer, B. S., Izmirlian, G., Miller, A. B., Pinsky, P. F., Prorok, P. C., Gohagan, J. K., and Berg, C. D. (2009). Mortality results from a randomized prostate-cancer screening trial. *New England Journal of Medicine*, 360(13):1310–1319.
- Andriole, G. L., Levin, D. L., Crawford, E. D., Gelmann, E. P., Pinsky, P. F., Chia, D., Kramer, B. S., Reding, D., Church, T. R., Grubb, R. L., Izmirlian, G., Ragard, L. R., Clapp, J. D., Prorok, P. C., and Gohagan, J. K. (2005). Prostate cancer screening in the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial: Findings from the initial screening round of a randomized trial. *Journal of the National Cancer Institute*, 97(6):433–438.
- Bill-Axelsson, A., Holmberg, L., Ruutu, M., Haggman, M., Andersson, S.-O., Bratell, S., Spangberg, A., Busch, C., Nordling, S., Garmo, H., Palmgren, J., Adami, H.-O., Norlen, B. J., Johansson, J.-E., and the Scandinavian Prostate Cancer Group Study No. 4 (2005). Radical prostatectomy versus watchful waiting in early prostate cancer. *New England Journal of Medicine*, 352(19):1977–1984.
- Draisma, G., Boer, R., Otto, S. J., van der Crujisen, I. W., Damhuis, R. A., Schröder, F. H., and de Koning, H. J. (2003). Lead times and overdiagnosis due to prostate-specific antigen screening: Estimates from the

- European Randomized Study of Screening for Prostate Cancer. *Journal of the National Cancer Institute*, 95(12):868–878.
- Draisma, G., Etzioni, R., Tsodikov, A., Mariotto, A., Wever, E., Gulati, R., Feuer, E., and de Koning, H. (2009). Lead times and overdiagnosis in prostate-specific antigen screening: Importance of methods and context. *Journal of the National Cancer Institute*, 101(6):374–383.
- Ellis, W. J., Etzioni, R., Vessella, R. L., Hu, C., and Goodman, G. E. (2001). Serial prostate specific antigen, free-to-total prostate specific antigen ratio and complexed prostate specific antigen for the diagnosis of prostate cancer. *Journal of Urology*, 166(1):93–98; discussion 98–99.
- Etzioni, R., Cha, R., Feuer, E. J., and Davidov, O. (1998). Asymptomatic incidence and duration of prostate cancer. *American Journal of Epidemiology*, 148(8):775–785.
- Etzioni, R., Gulati, R., Falcon, S., and Penson, D. (2008). Impact of psa screening on the incidence of advanced stage prostate cancer in the US: A surveillance modeling approach. *Medical Decision Making*, 28:323–331.
- Gann, P. H., Hennekens, C. H., and Stampfer, M. J. (1995). A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. *Journal of the American Medical Association*, 273(4):289–294.
- Gulati, R., Inoue, L. Y. T., Katcher, J., Hazelton, W., and Etzioni, R. (2009). Calibrating disease progression models using population data: A critical precursor to policy development in cancer control. *The Milbank Quarterly*. Submitted.
- Horner, M. J., Ries, L. A. G., Krapcho, M., Neyman, N., Aminou, R., Howlader, N., Altekruse, S. F., Feuer, E. J., Huang, L., Mariotto, A., Miller, B. A., Lewis, D. R., Eisner, M. P., Stinchcomb, D. G., and Edwards, B. K. (2009). *SEER Cancer Statistics Review, 1975-2006*. National Cancer Institute, Bethesda, MD.
- Inoue, L. Y., Etzioni, R., Slate, E. H., Morrell, C., and Penson, D. F. (2004). Combining longitudinal studies of PSA. *Biostatistics*, 5(3):483–500. 1465-4644.
- Inoue, L. Y. T., Etzioni, R., C., M., and P., M. (2008). Modeling disease progression with longitudinal markers. *Journal of the American Statistical Association*, 103(481):259–270.
- Mariotto, A. B., Etzioni, R., Krapcho, M., and Feuer, E. J. (2007). Reconstructing PSA testing patterns between black and white men in the US from Medicare claims and the National Health Interview Survey. *Cancer*, 109:1877–1886.
- Nelder, J. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, pages 308–313.
- Oesterling, J. E., Jacobsen, S. J., Chute, C. G., Guess, H. A., Girman, C. J., Panser, L. A., and Lieber, M. M. (1993). Serum prostate-specific antigen in a community-based population of healthy men. establishment of age-specific reference ranges. *Journal of the American Medical Association*, 270(7):860–864.
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

- Schröder, F. H., Hugosson, J., Roobol, M. J., Tammela, T. L., Ciatto, S., Nelen, V., Kwiatkowski, M., Lujan, M., Lilja, H., Zappa, M., Denis, L. J., Recker, F., Berenguer, A., Määttänen, L., Bangma, C. H., Aus, G., Villers, A., Rebillard, X., van der Kwast, T., Blijenberg, B. G., Moss, S. M., de Koning, H. J., and Auvinen, A. (2009). Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine*, 360(13):1320–1328.
- Spall, J. (2003). *Introduction to stochastic search and optimization: Estimation, simulation, and control*. Wiley, Hoboken, NJ.
- Telesca, D., Etzioni, R., and Gulati, R. (2008). Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. *Biometrics*, 64(1):10–19.
- Tsodikov, A., Szabo, A., and Wegelin, J. (2006). A population model of prostate cancer incidence. *Statistics in Medicine*, 25(16):2846–2866.
- US Preventive Services Task Force (2008). Screening for prostate cancer: US Preventive Services Task Force recommendation statement. *Annals of Internal Medicine*, 149(3):185–191.
- Whittemore, A. S., Lele, C., Friedman, G. D., Stamey, T., Vogelman, J. H., and Orentreich, N. (1995). Prostate-specific antigen as predictor of prostate cancer in black men and white men. *Journal of the National Cancer Institute*, 87(5):354–360.