

Online Supplement to “Optimization of Active Surveillance Strategies for Heterogeneous Patients with Prostate Cancer”

EC.1. Additional Model Description

EC.1.1. Illustration for the fixed sequence of events leading to treatment

Under Assumption 2, the patient will not be treated before PCa progression. This is true because the false-positive rate is 0. Accordingly, the sequence of events can be illustrated in Figure EC.1 where all patients start from the low-risk PCa (latent state), transition to the high-risk PCa (latent state) at period k ($0 \leq k \leq K$), and then transition to the treated PCa (terminal state) at period t ($k \leq t \leq T$). The transition from high-risk PCa to other states, including metastatic PCa and mortality, is independent of the strategy (Assumptions 1-2), and thus it can be implicitly accounted in the post-treatment parameters r_{pkt} and π_{pkt} .

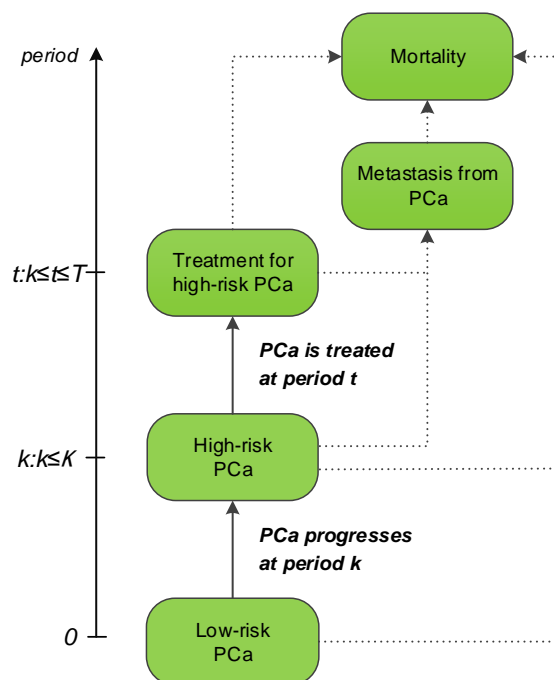


Figure EC.1 The fixed sequence of PCa evolution of each patient type with PCa progression at period k ($k \leq K$) and treatment at period t ($k \leq t \leq T$) where other randomness are denoted by dotted arcs.

EC.1.2. Illustration for predetermined outcome for a fixed combination of p , k , and t

Under Assumptions 1-2, the expected reward is independent of the surveillance strategy if p , k , and t are fixed. This is true because: 1) the sequence of events leading to treatment is fixed as

“low-risk PCa \rightarrow high-risk PCa \rightarrow treated PCa,” and 2) the transition probabilities related to PCa progression, metastasis, and mortality are independent of the biopsy testing (Assumption 1). As a result, the patient’s health outcome is unchanged no matter how many biopsy tests are performed before treatment, and thus the expected reward is independent of the surveillance strategy.

EC.1.3. Modification of \mathbf{P} to account for other criteria

\mathbf{P} can be straightforwardly modified to optimize different criteria, including but not limited to, the following:

- *Delay of detection:* to reduce the delay of detection, we set $r_{pkt} = \sum_{v=k}^T \hat{\pi}_{pkv} - \sum_{v=k}^{t-1} \hat{\pi}_{pkv}$ for all $p \in \mathcal{P}$, $k \in \mathcal{K}$, and $k \leq t \leq T$. Under this setting, earlier detection is always better than later detection. Further, we set $d_{pt} = 1$ for all $p \in \mathcal{P}$, $t \in \mathcal{T}$, and c_p is the unit cost to test patient type p compared with one-year delay of treatment.

- *Survival rate measurement:* to improve the patient’s survival rate, let r_{pkt} denote the rate of survival within a certain period, e.g., 10 years after diagnosis, for patient type p treated at period t (or missed treatment if $t = T + 1$) when PCa progresses at period k . Under this setting, the survival rate will be nondecreasing for earlier treatment, and thus $d_{pt} = 1$ for all $p \in \mathcal{P}$, $t \in \mathcal{T}$. Let c_p denote the unit cost to test patient type p compared with 1% improvement of the survival rate.

- *Cost-effectiveness measurement:* to improve the cost-effectiveness, we set r_{pkt} as a net monetary benefit criterion as $r_{pkt} = \hat{q}_{pkt} - \hat{c}_{pkt}/\kappa_p$ where \hat{q}_{pkt} is the extra QALY gained by treating patient type p at period t compared with “no treatment,” \hat{c}_{pkt} is the cost associated with the treatment, and κ_p is a weight factor meaning the equivalent amount of money the patient is willing to pay for increasing one QALY. We set $d_{pt} = 0$ if treatment is not cost-effective at period t for patient type p and we set $c_p = \tilde{q}_p + \tilde{c}_p/\kappa_p$ where \tilde{q}_p is the instantaneous disutility of QALYs for patient type p , and \tilde{c}_p is the cost of one-time biopsy test for patient type p .

EC.1.4. Proof of NP-completeness of \mathbf{P}

We show that the surveillance strategy optimization problem as defined by \mathbf{P} is NP-complete even if $P = 1$. We prove the NP-completeness using reduction from the 3-CNF satisfiability problem.

Definition: A boolean formula is in 3-conjunctive normal form, or 3-CNF, if it is expressed as conjunctions (by AND) of clauses, where each clause is the disjunction (by OR) of exactly three distinct literals over a set of boolean variables \mathbf{x} . For example, the following is a 3-CNF formula with two clauses,

$$(x_{11} \vee x_{12} \vee x_{13}) \wedge (x_{21} \vee x_{22} \vee x_{23}),$$

and the above formula is satisfiable if $\sum_{j=1}^3 x_{ij} \geq 1, \forall i = 1, 2$. The 3-CNF satisfiability problem is to validate whether a given 3-CNF formula is satisfiable.

We construct the surveillance strategy optimization problem as follows: $\mathcal{P} = \{1\}$, $\mathcal{T} = \{1, \dots, \tau, \dots, 2\tau, \dots, 3\tau, 3\tau + 1\}$, $\mathcal{K} = \mathcal{T} \cup \{0\}$, $c_p = 0$, $\sigma = 100\%$, and

$$\hat{\xi}_{pk} = \begin{cases} 1/\tau, & \text{if } k-1 \text{ is dividable by } 3 \\ 0, & \text{otherwise} \end{cases}, r_{pkt} = \begin{cases} \epsilon, & \text{if } k \leq t \leq k+2, \text{ and } t \leq 3\tau \\ 0, & \text{otherwise} \end{cases},$$

for all $p \in \mathcal{P}$, $k \in \mathcal{K}$, and $k \leq t \leq T$, where τ is a positive integer, and ϵ is a positive constant. The maximum reward of \mathbf{P} is ϵ , which is achieved if and only if all progressed PCa is detected via biopsy and treated within the first three periods after progression, i.e., $t \in \{k, k+1, k+2\}$. When $\sigma = 100\%$, this solution requires that the following boolean formula,

$$(x_1 \vee x_2 \vee x_3) \wedge (x_4 \vee x_5 \vee x_6) \wedge \dots \wedge (x_{3\tau-2} \vee x_{3\tau-1} \vee x_{3\tau}),$$

be a feasible 3-CNF, which completes the proof. \square

EC.1.5. Extensive linearized model of \mathbf{P}

The following is a linearized reformulation of \mathbf{P} that considers x_{st} , y_{ps} as the first-stage decision variables, and u_{pkt} and \hat{v}_{pkst} as the second-stage decision variables (note that v_{pkt} is extended to be \hat{v}_{pkst} in this subsection to capture the dependency of the probability of detected PCa on the strategy, and $v_{pkt} = \sum_{s \in \mathcal{S}} \hat{v}_{pkst}$). Since the second-stage program is purely dependent of x_{st} and y_{ps} , we present the model reformulation as a single mixed-integer program as follows:

$$\mathbf{MIP} : \max_{\mathbf{x}, \mathbf{y}, \mathbf{u}, \hat{\mathbf{v}}} \sum_{p \in \mathcal{P}} \sum_{k \in \mathcal{K}} \hat{\xi}_{pk} \left[\sum_{t=k}^{T+1} \sum_{s \in \mathcal{S}} r_{pkt} \hat{v}_{pkst} - \sum_{t \in \mathcal{T}} \hat{\pi}_{pkt} u_{pkt} \right] \quad (\text{EC.1a})$$

s.t. (1b), and

$$\hat{v}_{pkst} + d_{pt} \sigma \sum_{t'=k}^{t-1} \hat{v}_{pkst'} \leq d_{pt} \sigma, \quad \forall p \in \mathcal{P}, k \in \mathcal{K}, s \in \mathcal{S}, k \leq t \leq T, \quad (\text{EC.1b})$$

$$\sum_{k=t}^{K+1} \hat{v}_{pkst} \leq M_{pt} x_{st}, \quad \forall p \in \mathcal{P}, s \in \mathcal{S}, 1 \leq t \leq T, \quad (\text{EC.1c})$$

$$\sum_{k \in \mathcal{K}} \sum_{t=k}^T \hat{v}_{pkst} \leq \hat{M}_p y_{ps}, \quad \forall p \in \mathcal{P}, s \in \mathcal{S}, \quad (\text{EC.1d})$$

$$\sum_{t=k}^{T+1} \hat{v}_{pkst} \leq y_{ps}, \quad \forall p \in \mathcal{P}, k \in \mathcal{K}, s \in \mathcal{S}, \quad (\text{EC.1e})$$

$$-u_{pkt} - \sum_{t'=k}^{t-1} \hat{v}_{pkst'} \leq 2 - d_{pt} - x_{st} - y_{ps}, \quad \forall p \in \mathcal{P}, k \in \mathcal{K}, s \in \mathcal{S}, 1 \leq t \leq T, \quad (\text{EC.1f})$$

$$x_{st}, y_{ps} \in \{0, 1\}, u_{pkt}, \hat{v}_{pkst} \geq 0.$$

where (EC.1a) is the total expected reward that combines (1a) and (2a). (EC.1b)-(EC.1e) reformulate (2b) and (4a)-(4b) which jointly determine the upper bound of \hat{v}_{pkst} , where (EC.1b) enforces

$\hat{v}_{pkst} \leq d_{pt}\sigma(1 - \sum_{t'=k}^{t-1} \hat{v}_{pkst'})$ which is the maximum probability of being detected at period t if the biopsy is performed, and (EC.1c)-(EC.1e) enforce $\hat{v}_{pkst} \leq 0$ if $x_{st} = 0$ or $y_{ps} = 0$. (EC.1f) reformulates (3) that determines the lower bound of u_{pkt} , i.e., $u_{pkt} \geq 1 - \sum_{t'=k}^{t-1} \hat{v}_{pkst'}$ when $d_{pt} = 1$, $x_{st} = 1$, $y_{ps} = 1$, and $u_{pkt} \geq 0$, otherwise. The reformulation can be shown as follows:

When $x_{st} + y_{ps} + d_{pt} \leq 2$, (EC.1f) always holds because the left-hand side is non-positive and the right-hand side is non-negative.

When $x_{st} + y_{ps} + d_{pt} = 3$, i.e., $x_{st} = 1$, $y_{ps} = 1$, and $d_{pt} = 1$, we have $q_{pkt} = d_{pt} \left(1 - \sum_{t'=k}^{t-1} v_{pkt'}\right) = 1 - \sum_{t'=k}^{t-1} \hat{v}_{pkst'}$, and $w_{pt} = x_{st}y_{ps} = 1$, under which (EC.1f) can be rewritten as:

$$-u_{pkt} + q_{pkt} \leq 1 - w_{pt}, \quad \forall p \in \mathcal{P}, k \in \mathcal{K}, 1 \leq t \leq T$$

which is equivalent to (3) when we set the bigM coefficient, \hat{M}_{pt} , as 1.

EC.1.6. Updating \mathbf{b}_{t+1} using Bayesian inference

Let $\bar{q}_t(O | S, A)$ denotes the probability of making observation O at period t when the action chosen is A and the state is S . Let $\bar{p}_t(S' | S, A)$ denote the transition probability from S to S' under action A . Therefore, the probability of transitioning to the absorbing state, $\bar{p}_t(\mathbb{T} | \mathbf{b}_t, A_t) = \sum_{S_t \in \mathcal{S}} \bar{p}_t(\mathbb{T} | S_t, A_t) b_t(S_t)$. The probability of making an observation O_t , $\bar{p}_t(O_t | \mathbf{b}_t, A_t) = \sum_{S_t \in \mathcal{S}} \bar{q}_t(O_t | S_t, A_t) b_t(S_t)$. Then, the belief state in period $t+1$ is updated as follows:

$$b_{t+1}(S_{t+1}) = \sum_{S_t \in \mathcal{S}} \bar{p}_t(S_{t+1} | S_t, A_t) \hat{b}_t(S_t), \quad \forall S_{t+1} \in \mathcal{S}, \quad (\text{EC.2})$$

where $\hat{b}_t(S_t)$ is the updated belief state at the end of period t after observing O_t , which can be determined according to Bayesian inference as follows:

$$\hat{b}_t(S_t) = \frac{\bar{q}_t(O_t | S_t, A_t) b_t(S_t)}{\sum_{S'_t \in \mathcal{S}} \bar{q}_t(O_t | S'_t, A_t) b_t(S'_t)}, \quad \forall S_t \in \mathcal{S}. \quad (\text{EC.3})$$

EC.1.7. Extension of \mathbf{P} to *left to right* Markov models with multiple latent states

We present a simple extension of \mathbf{P} to *left to right* Markov models with more than two states, in which the progression among states is sequential and irreversible (which do arise in cancer surveillance and probably other areas). We let \mathcal{L} denote a set of latent states indexed by $l = 0, 1, \dots, L$, where L is the total number of latent states. To formulate the model extension, we need to make additional assumptions as follows:

ASSUMPTION EC.1. *The progression among latent states in \mathcal{L} is independent of the biopsy testing.*

ASSUMPTION EC.2. All latent states are ordinal, where $l = 1$ is the least severe state and $l = L$ is the most severe state. Moreover, there exists a definitive classifier \tilde{l} for \mathcal{L} such that states $\tilde{l} \leq l \leq L$ require treatment upon detection, and states $0 \leq l \leq \tilde{l} - 1$ are nonmalignant and do not require treatment.

We expand the notation of $\hat{\xi}_{pk}^l$, $\hat{\pi}_{pkt}$, r_{pkt} , v_{pkt} , u_{pkt} , and q_{pkt} to be l -dependent as $\hat{\xi}_{pk}^l$, $\hat{\pi}_{pkt}^l$, r_{pkt}^l , v_{pkt}^l , u_{pkt}^l , and q_{pkt}^l , respectively. Based on Assumptions 1-2 and 4-5, the model extension (**EP**) accounting for multiple latent states can be formulated as the following two-stage SIP:

$$\begin{aligned} \mathbf{EP}: \max_{\mathbf{x}, \mathbf{y}, \mathbf{w}} Q(\mathbf{x}, \mathbf{y}, \mathbf{w}) &= \sum_{p \in \mathcal{P}} \sum_{k \in \mathcal{K}} \sum_{l \in \mathcal{L}} \hat{\xi}_{pk}^l \tilde{Q}(\mathbf{x}, \mathbf{y}, \mathbf{w}, p, k, l) & (\text{EC.4a}) \\ \text{s.t.} & \text{ (1b), (1c), and } \\ & x_{st}, y_{ps}, w_{pt} \in \{0, 1\}. \end{aligned}$$

where $\tilde{Q}(\mathbf{x}, \mathbf{y}, \mathbf{w}, p, k, l)$ is the maximum reward associated with p , k , and l , which is determined by the following second-stage recourse program:

$$\tilde{Q}(\mathbf{x}, \mathbf{y}, \mathbf{w}, p, k, l) = \max_{\mathbf{u}, \mathbf{v}, \mathbf{q}} \sum_{t=k}^{T+1} r_{pkt}^l v_{pkt}^l - \sum_{t=T} \hat{\pi}_{pkt}^l u_{pkt}^l \quad (\text{EC.5a})$$

$$\text{s.t. } q_{pkt}^l + d_{pt} \sum_{t'=k}^{t-1} v_{pkt'}^l = d_{pt}, \forall 1 \leq t \leq T, 0 \leq l \leq L, \quad (\text{EC.5b})$$

$$-q_{pkt}^l w_{pt} + u_{pkt}^l = 0, \forall 1 \leq t \leq T, 0 \leq l \leq L, \quad (\text{EC.5c})$$

$$v_{pkt}^l = 0, \forall k \leq t \leq T, 0 \leq l \leq \tilde{l} - 1, \quad (\text{EC.5d})$$

$$-q_{pkt}^l w_{pt} \sigma + v_{pkt}^l = 0, \forall k \leq t \leq T, \tilde{l} \leq l \leq L \quad (\text{EC.5e})$$

$$v_{pk, T+1}^l + \sum_{t'=k}^T v_{pkt'}^l = 1, \forall 0 \leq l \leq L, \quad (\text{EC.5f})$$

$$u_{pkt}^l, v_{pkt}^l, q_{pkt}^l \geq 0.$$

EC.1.8. Standard Benders decomposition of $\hat{\mathbf{P}}$ and computational results

Standard Benders decomposition. Following the standard approach for Benders decomposition, we use the dual subproblem (**DSP**) of $\hat{\mathbf{P}}$ to determine the expected reward for a given solution (\mathbf{x}, \mathbf{y}) , and generates the following optimality cut to $\mathcal{F}_v(\eta, \mathbf{x}, \mathbf{y})$:

$$\begin{aligned} \eta \leq & \sum_{p \in \mathcal{P}} \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \sum_{t=k}^{t-1} d_{pt} \sigma \bar{\alpha}_{pkst} + \sum_{p \in \mathcal{P}} \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} M_{pt} \bar{\beta}_{pst} x_{st} + \sum_{p \in \mathcal{P}} \sum_{s \in \mathcal{S}} \hat{M}_p \bar{\gamma}_{ps} y_{ps} \\ & + \sum_{p \in \mathcal{P}} \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \bar{\theta}_{pks} y_{ps} + \sum_{p \in \mathcal{P}} \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} (2 - d_{pt} - x_{st} - y_{ps}) \bar{\chi}_{pkst} \end{aligned} \quad (\text{EC.6})$$

where $\bar{\alpha}$, $\bar{\beta}$, $\bar{\gamma}$, $\bar{\theta}$, and $\bar{\chi}$ are incumbent optimal solutions of the following dual subproblem:

$$\mathbf{DSP}: \min_{\alpha, \beta, \gamma, \theta, \chi} \sum_{p \in \mathcal{P}} \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \sum_{t=k}^{t-1} d_{pt} \sigma \alpha_{pkst} + \sum_{p \in \mathcal{P}} \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} M_{pt} x_{st} \beta_{pst} + \sum_{p \in \mathcal{P}} \sum_{s \in \mathcal{S}} \hat{M}_p y_{ps} \gamma_{ps}$$

$$+ \sum_{p \in \mathcal{P}} \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \theta_{pks} y_{ps} + \sum_{p \in \mathcal{P}} \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} (2 - d_{pt} - x_{st} - y_{ps}) \chi_{pkst} \quad (\text{EC.7a})$$

$$\begin{aligned} \text{s.t. } \alpha_{pkst} + \sigma \sum_{t'=t+1}^T d_{pt'} \alpha_{pkst'} + \beta_{pst} + \gamma_{ps} + \theta_{pks} - \sum_{t'=t+1}^T \chi_{pkst'} &\geq r_{pkt} \hat{\xi}_{pk}, \\ \forall p \in \mathcal{P}, k \in \mathcal{K}, s \in \mathcal{S}, k \leq t \leq T, & \end{aligned} \quad (\text{EC.7b})$$

$$\sum_{s \in \mathcal{S}} \theta_{pks} \geq r_{pk, T+1} \hat{\xi}_{pk}, \quad \forall p \in \mathcal{P}, k \in \mathcal{K}, \quad (\text{EC.7c})$$

$$- \sum_{s \in \mathcal{S}} \chi_{pkst} \geq -\hat{\pi}_{pkt} \hat{\xi}_{pk}, \quad \forall p \in \mathcal{P}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (\text{EC.7d})$$

$$\alpha_{pkst}, \beta_{pst}, \gamma_{ps}, \theta_{pks}, \chi_{pkst} \geq 0, \quad (\text{EC.7e})$$

where α_{pkst} , β_{pst} , γ_{ps} , θ_{pks} , and χ_{pkst} are dual variables of the subproblem of $\hat{\mathbf{P}}$, corresponding to constraints (EC.1b)-(EC.1f), respectively, when the first-stage decisions \mathbf{x}, \mathbf{y} are fixed.

Computational settings and results. Our main purpose is to investigate the computational performance of different approaches, including the extensive formulation (B&C), the standard Benders decomposition (SBD), and the logic-based Benders decomposition (LBD). We tested the approaches on 80 instances that were classified into eight categories depending on the number of patient types, $P \in \{40, 50\}$, and the number of strategies, $S \in \{2, 5\}$, and the distribution type of the reward parameters corresponding to deterministically sampled r_{pkt} and randomly sampled r_{pkt} , respectively. Moreover, we set $T = K = 10$, and ε be a random variable following a standard uniform distribution, $\text{UNIF}(0, 1)$. r_{pkt} was generated as follows:

- Deterministic \mathbf{r} : set $r_{pkt} = T + 1 + t - k$ for all $p \in \mathcal{P}$, $k \in \mathcal{K}$, $k \leq t \leq T + 1$,
- Random \mathbf{r} : initialize $r_{P,1,T+1} = 0$, and then iteratively sample other r_{pkt} s as follows: $r_{pkt} = r_{p,k,t+1} + \varepsilon$ and $r_{pk,T+1} = r_{p,k-1,T+1} + \varepsilon$ for all $p \in \mathcal{P}$, $2 \leq k \leq K + 1$, and $k \leq t \leq T$.

Similarly, we set $\pi_{pk1} = 1$, and iteratively sample π_{pkt} as follows:

$$\pi_{pkt} = \begin{cases} \pi_{pk,t-1} + \frac{0.2\varepsilon}{T} & \text{if } t < k, \\ \pi_{pk,t-1} + \frac{0.5\varepsilon}{T} & \text{if } t \geq k, \end{cases}$$

for all $p \in \mathcal{P}$, $k \in \mathcal{K}$, and $t = 2, \dots, T$. We set $\hat{\xi}_{p0} = \varepsilon$ and iteratively sample $\hat{\xi}_{pk} = \varepsilon \left(1 - \sum_{k' \in \mathcal{K} | k' < k} \hat{\xi}_{pk'}\right)$ for all $p \in \mathcal{P}$ and $k = 2, \dots, K$.

All algorithms were implemented in Microsoft Visual Studio.NET 2017 linking with the CPLEX 12.8 callable library. Experiments were conducted on an Intel Core i7-9700 PC with processors running at 3.00 GHz and 16 GB memory under Windows 10. All algorithms were stopped when either they obtained the optimal solution (concerning a relative gap $\leq 0.01\%$) or the computation time reached a maximum of 3,600 seconds.

We report the average and maximum solution times and optimality gaps across ten instances in Table EC.1. The main results are summarized as follows:

Table EC.1 Computational performance of the extensive formulation (B&C) method, the standard Benders decomposition (SBD) method, and the logic-based Benders decomposition (LBD) method (CPU time limit is 3,600 seconds)

P	S	Method	Deterministically sampled \mathbf{r}				Randomly sampled \mathbf{r}			
			Solution time (CPU seconds)		Optimality gap, %		Solution time (CPU seconds)		Optimality gap, %	
			Avg	Max	Avg	Max	Avg	Max	Avg	Max
40	2	B&C	818	2424	0.00	0.00	1429	2178	0.00	0.00
		SBD	4	5	0.00	0.00	3600	3600	0.12	0.21
		LBD	4	4	0.00	0.00	37	51	0.00	0.00
	5	B&C	3601	3602	16.81	22.57	3601	3602	5.58	6.80
		SBD	6	9	0.00	0.00	3600	3601	0.10	0.19
		LBD	5	9	0.00	0.00	3216	3601	0.01	0.03
50	2	B&C	1634	3601	0.18	1.77	2238	3601	0.10	0.42
		SBD	6	7	0.00	0.00	3399	3600	0.07	0.16
		LBD	6	7	0.00	0.00	80	99	0.00	0.00
	5	B&C	3601	3601	17.12	21.13	3601	3601	5.92	7.15
		SBD	7	8	0.00	0.00	3390	3601	0.10	0.16
		LBD	6	8	0.00	0.00	3311	3603	0.02	0.04

- The LBD method was 208% and 110% faster than the SBD and B&C methods, respectively, which demonstrates that the logic-based Benders decomposition formulation helped significantly in reducing the computational complexity compared with the original formulation.

- For instances unsolved to optimality within the time limit, the maximum optimality gap was 22.57%, 0.21%, and 0.04% for the B&C, SBD, and LBD methods, respectively. The value of the logic-based Benders decomposition is particularly significant for active surveillance practice because it has a large fixed reward associated with no surveillance due to the low progression rate to high-risk PCa, which is the fundamental motivation for surveillance instead of immediate treatment. Therefore, the optimality gap is often minimal, but the relative term corresponds to a large absolute difference by public health standards. In such a situation, the LBD method is beneficial to analyze the structure of the optimal solution and improve the overall performance in a large population.

EC.1.9. Results of parameter estimation in the case study

The total expected reward is a weighted sum of rewards across all patient types, where the number of patients for each type (n_p) is estimated based on enrolled patients in the timeframe 1995–2014 as published in Inoue et al. (2018). See Table EC.2 for the number of patients for each patient type. The other parameters including π_{pkt} and r_{pkt} are estimated based on a Markov process model as shown in Figure EC.2. For the ease of readability, we use figures to highlight the representative π_{pkt} and r_{pkt} under different k and t . The results of parameter estimation are given in order:

- Estimation of the probability of PCa progression at period k for patient type p , ξ_{pk} , is illustrated in Figure EC.3.

- Estimation of the probability of patient type p staying in active surveillance at period t when PCa progresses at period k and the patient is not treated, π_{pkt} , is illustrated in Figure EC.4.
- Estimation of the expected QALYs if patient type p is treated at period t , r_{pkt} , for $1 \leq t \leq T$, or not treated, $r_{pk,T+1}$, when PCa progresses at period k , which is illustrated in Figure EC.5.

Table EC.2 Number of enrolled patients in each type among 1,000 patients (Inoue et al. 2018, Table 2)

	$50 \leq A_p \leq 59$	$60 \leq A_p \leq 69$	$70 \leq A_p \leq 75$
JH	7.1	27.2	12.2
SF	20.5	30.7	12

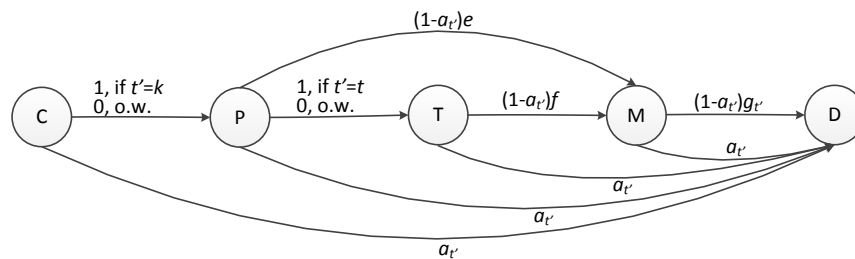


Figure EC.2 Simplified Markov process model of PCa progression based on Figure 2 for fixed p , k , and t where nodes C, P, T, M, and D denote the states of “Low-risk PCa”, “High-risk PCa”, “Treatment for High-risk PCa”, “Metastasis from PCa”, and “Mortality” which combines “Other-cause mortality” and “PCa mortality”, respectively, arcs denote the transition probabilities.

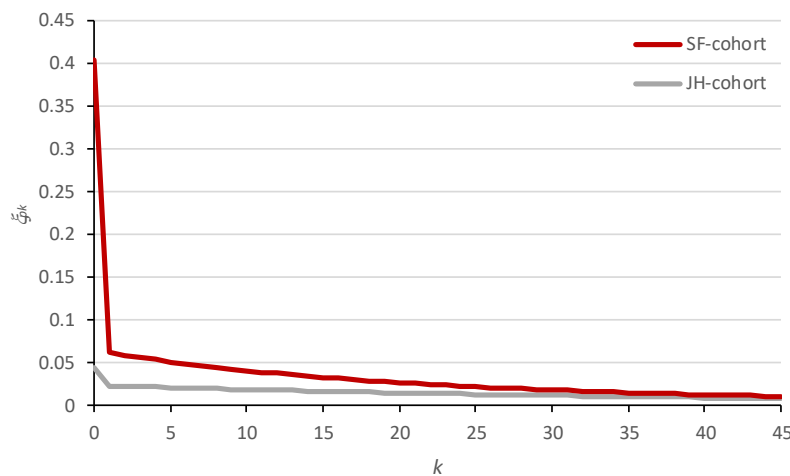


Figure EC.3 Estimation of ξ_{pk} for patients from the JH and SF cohorts, respectively, where $\xi_{p,K+1} = 1 - \sum_{k=1}^{45} \xi_{pk}$ is the probability of unprogressed PCa at the end of K periods.

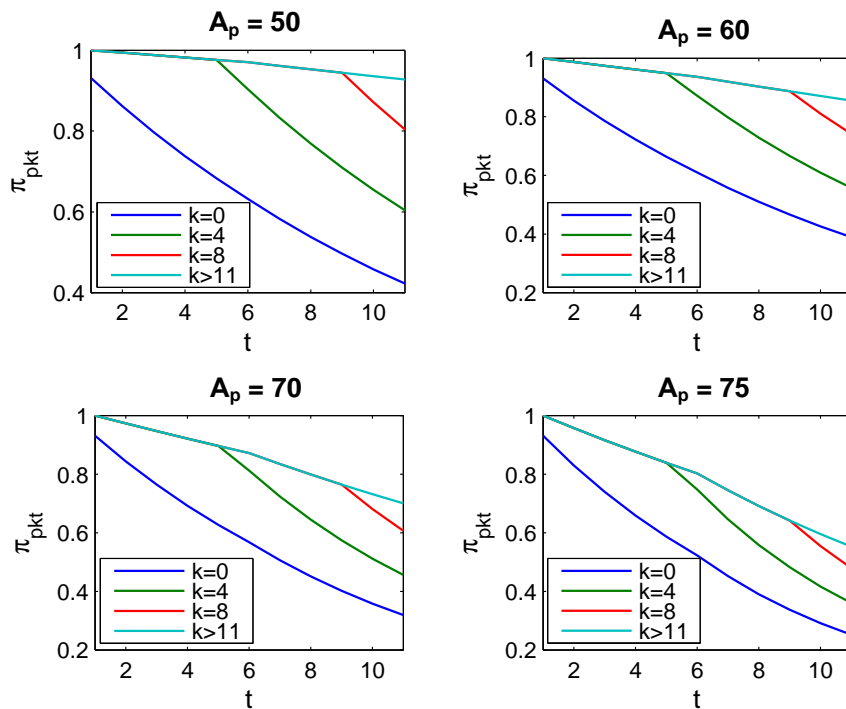


Figure EC.4 Probability of patients staying in surveillance at period t , π_{pkt}

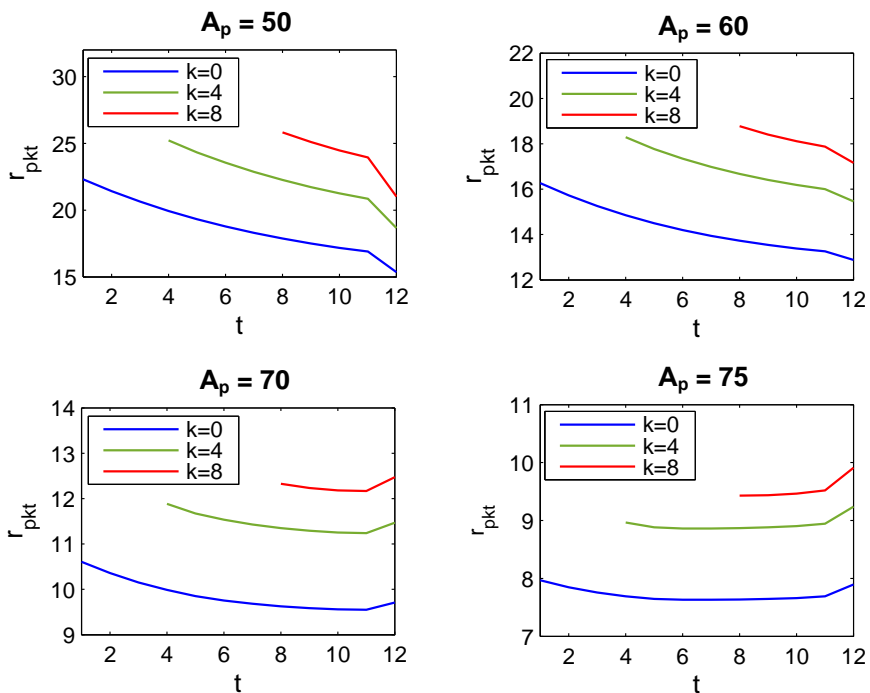


Figure EC.5 The expected reward associated with being treated at period t , r_{pkt} , where $t = 12$ ($T + 1$) denotes being missed of treatment

EC.1.10. Results of model validation

Our main purpose is to validate the disease model by comparing the model outputs to published sources on PCa outcomes in this subsection. Specifically, we compared the following indicators of PCa outcomes:

- Expected time between beginning surveillance and being detected via biopsy for progressed PCa, which we refer to as the *expected time to detection*,
- PCa specific mortality (PCSM) rate.

In terms of the expected time to detection, Inoue et al. (2018) reported their estimated outcomes (with the title “time to upgrading”) based on men diagnosed at 60 years old in the JH and the SF cohorts. We compare the model outputs with their outcomes. The results are summarized in Table EC.3, which show that the two outcomes are closely matched with a small difference ($\leq 6.0\%$ in the worst case across all comparisons) in both JH and SF cohorts and under different strategies. This result means that the disease model accurately reflects PCa disease progression. We also reported more model-based estimates of the expected time to detection in Table EC.4 corresponding to men diagnosed at different ages and under different surveillance strategies.

In terms of the PCSM rate, Tosoian et al. (2016) estimated that the PCSM rate was between 0% and 1.5% in the intermediate-term (5–10 years) while our estimate was between 0.2% and 1.0% in 5-10 years under different strategies on men with various diagnosis ages. However, comprehensive assessments of the post surveillance outcomes across multiple cohorts are currently lacking. We also report more model-based estimates of post surveillance outcomes such as the PCSM rate and the life expectancy in Table EC.4 for future comparison.

Table EC.3 Comparing our model outputs to Inoue et al. (2018) in terms of the expected time to detection under surveillance strategies that biopsy every 1 or 2 years up to 11 years for men diagnosed at 60 years old

Cohort	Number of years between biopsies	Expected time to detection (years) by Inoue et al. (2018)		Expected time to detection (years) by our model	
		Mean	SD	Mean	SD
JH	1	5.18	3.22	5.16	3.23
	2	5.58	3.41	5.66	3.36
SF	1	3.64	2.92	3.42	2.94
	2	3.92	3.14	3.85	3.18

Table EC.4 Model-based estimates of the expected time to detection, the PCa specific mortality rate, and the life expectancy under surveillance strategies that biopsy every 1 or 2 or 3 years up to 11 years for men diagnosed at different ages

Cohort	Number of years between biopsies	Expected time to detection (years)		PCa specific mortality, %			Life expectancy (years)
		Mean	SD	5 years	10 years	15 years	
Men diagnosed at 50 years old							
JH	1	5.25	3.25	0.2	0.5	1.0	28.2
	2	5.76	3.37	0.3	0.8	1.5	28.0
	3	5.70	3.24	0.3	1.0	1.9	27.8
SF	1	3.49	2.99	1.5	2.9	4.4	27.1
	2	3.94	3.22	1.8	3.6	5.6	26.8
	3	3.96	3.19	1.9	4.2	6.6	26.4
Men diagnosed at 60 years old							
JH	1	5.16	3.23	0.2	0.5	0.9	20.7
	2	5.66	3.36	0.3	0.7	1.3	20.6
	3	5.61	3.24	0.3	0.9	1.7	20.5
SF	1	3.42	2.94	1.5	2.7	4.0	20.1
	2	3.85	3.18	1.7	3.4	5.0	19.9
	3	3.88	3.16	1.8	3.9	5.9	19.7
Men diagnosed at 70 years old							
JH	1	4.94	3.18	0.2	0.4	0.7	13.7
	2	5.41	3.32	0.2	0.6	1.0	13.7
	3	5.39	3.22	0.3	0.8	1.3	13.6
SF	1	3.25	2.83	1.3	2.4	3.3	13.4
	2	3.65	3.07	1.5	3.0	4.1	13.3
	3	3.69	3.08	1.7	3.4	4.8	13.3

EC.1.11. Results of sensitivity analysis

We conducted a series of one-way sensitivity analyses based on varying selected data (Table EC.5).

The results of the sensitivity analysis are presented in Figure EC.6 (page 47).

Table EC.5 Summary of data for one-way sensitivity analysis

Data	Range	Reference
Sensitivity of biopsy, σ	0.582 - 0.688	Barnett et al. (2018)
Disutility in the year of biopsy, c_B	$\pm 20\%$	
Misclassification rate for the JH cohort, $\hat{w}(\text{JH})$	$\pm 20\%$	
Misclassification rate for the SF cohort, $\hat{w}(\text{SF})$	$\pm 20\%$	
Annual progression rate for the JH cohort, $w(\text{JH})$	$\pm 20\%$	
Annual progression rate for the SF cohort, $w(\text{SF})$	$\pm 20\%$	
Annual metastasis rate from treated high-risk PCa, e	$\pm 20\%$	
Annual metastasis rate from untreated high-risk PCa, f	$\pm 20\%$	
Annual death rate from all-other causes, a_t	$\pm 95\%$ CI	US Life Tables (2012)

EC.2. Technical Proofs

EC.2.1. Proof of Lemma 1

Proof: let $\delta(x)$ denote the change in x . Suppose $t \geq k$ and the patient is undetected, then he will be detected among periods $\{t, \dots, T\}$ or remain undetected at period $T + 1$. According to (2e), as v_{pkt} increases (i.e., $\delta(v_{pkt}) > 0$), the following equation always holds for all $p \in \mathcal{P}$ and $k \in \mathcal{K}$:

$$\delta(v_{pkt}) + \sum_{t'=t}^{T+1} \delta(v_{pkt'}) = 0.$$

According to Assumption 3, $r_{pkt} \geq d_{pt} r_{pkt'}$, $\forall t + 1 \leq t' \leq T + 1$, and thus

$$\begin{aligned} & r_{pkt} \delta(v_{pkt}) + \sum_{t'=t}^{T+1} r_{pkt'} \delta(v_{pkt'}) \\ & \geq r_{pkt} \left(\delta(v_{pkt}) + \sum_{t'=t}^{T+1} r_{pkt'} \delta(v_{pkt'}) \right) \\ & \geq 0 \end{aligned} \tag{EC.8}$$

for all $p \in \mathcal{P}$, $k \in \mathcal{K}$, and $k \leq t \leq T$ when $d_{pt} = 1$. Moreover, as v_{pkt} increases, $q_{pkt'}$ is nonincreasing for all $t + 1 \leq t' \leq T$ when $d_{pt'} = 1$ because of the following transformation:

$$\begin{aligned} q_{pk,t+1} &= 1 - \sum_{t'=k}^{t-1} v_{pkt'} - v_{pkt} \\ &= q_{pkt} - (q_{pkt} w_{pt} \sigma) \\ &= q_{pkt} (1 - w_{pt} \sigma) \end{aligned}$$

where the first equation is from (2b) and the second equation is from (2d) when $d_{pt} = d_{p,t+1} = 1$. We further have $u_{pk,t'}$ is nonincreasing for all $t + 1 \leq t' \leq T$ according to (2c). Therefore, as v_{pkt} increases, $\tilde{Q}(\mathbf{x}, \mathbf{y}, \mathbf{w}, p, k)$ is nondecreasing for all $p \in \mathcal{P}$, $k \in \mathcal{K}$, and $k \leq t \leq T$ when $d_{pt} = 1$, which completes the proof. \square

EC.2.2. Proof of Theorem 1

Proof: the proof is straightforward because:

- When $d_{pt} = 0$, v_{pkt} is enforced to be zero according to (2b) and (2d),
- When $d_{pt} = 1$, v_{pkt} will achieve its upper bound value to maximize $\tilde{Q}(\mathbf{x}, \mathbf{y}, \mathbf{w}, p, k)$ (Lemma 1),

which completes the proof. \square

EC.2.3. Proof of Theorem 2

Proof. We consider $\xi_{pk} = \hat{\xi}_{pk}$ in this proof because there is only one patient type. Let ΔV_t^Π denote the expected reward received at period t in the POMDP-surveillance model under policy Π , and thus it can be determined according to (6) as follows:

$$\Delta V_t^\Pi = \begin{cases} (R_t(\mathbf{b}_t, A_t^\Pi) + \bar{p}_t(\mathbb{T} | \mathbf{b}_t, A_t^\Pi) \bar{R}_t(\mathbb{T})) \prod_{t'=1}^{t-1} (1 - \bar{p}_{t'}(\mathbb{T} | \mathbf{b}_{t'}, A_{t'}^\Pi)), & \text{if } 1 \leq t \leq \bar{T} \\ \bar{R}_{\bar{T}+1}(\mathbb{T}) \prod_{t'=1}^{t-1} (1 - \bar{p}_{t'}(\mathbb{T} | \mathbf{b}_{t'}, A_{t'}^\Pi)), & \text{if } t = \bar{T} + 1. \end{cases} \tag{EC.9}$$

where $R_t(\mathbf{b}_t, \mathbf{B}) = -d_{pt}\tilde{\pi}_{pt}$, $R_t(\mathbf{b}_t, \mathbf{W}) = 0$, $\bar{p}_t(\mathbb{T} | \mathbf{b}_t, \mathbf{W}) = 0$ for all \mathbf{b}_t according to (8). The belief state \mathbf{b}_t is independent of Π and it is specified as (7). Let ΔQ_t^Π denote the expected reward received by one patient at period t in \mathbf{P} under policy Π , and it can be determined according to (1a), and (2a)-(2e) as follows:

$$\Delta Q_t^\Pi = \begin{cases} \sum_{k \leq t} \xi_{pk} r_{pkt} v_{pkt}^\Pi - \sum_{k \in \mathcal{K}} \xi_{pk} \hat{\pi}_{pkt} u_{pkt}^\Pi, & \text{if } 1 \leq t \leq \bar{T}, \\ 0, & \text{if } \bar{T} + 1 \leq t \leq T, \\ \sum_{k \in \mathcal{K}} \xi_{pk} \left(1 - \sum_{t'=k}^T v_{pkt'}^\Pi\right) r_{pk, T+1}, & \text{if } t = T + 1, \end{cases} \quad (\text{EC.10})$$

where v_{pkt}^Π and u_{pkt}^Π are the second stage decision of \mathbf{P} under policy Π . In the following proof, we show that $\Delta Q_t^\Pi = \Delta V_t^\Pi$ for all $1 \leq t \leq \bar{T}$ in **Step 1**, and $\Delta Q_{T+1}^\Pi = \Delta V_{T+1}^\Pi$ in **Step 2**.

Step 1: we prove that $\Delta Q_t^\Pi = \Delta V_t^\Pi$ for all $1 \leq t \leq \bar{T}$. According to (EC.10), we reformulate ΔQ_t^Π as follows:

$$\Delta Q_t^\Pi = \sum_{k \leq t} \xi_{pk} r_{pkt} q_{pkt}^\Pi w_{pt}^\Pi \sigma - \sum_{k \in \mathcal{K}} \xi_{pk} \hat{\pi}_{pkt} q_{pkt}^\Pi w_{pt}^\Pi \quad (\text{EC.11a})$$

$$= \sum_{k \in \mathcal{K}} \xi_{pk} q_{pkt}^\Pi \left[-w_{pt}^\Pi \tilde{\pi}_{pt} + 1_{\{k \leq t\}} w_{pt}^\Pi \sigma \bar{R}_t(\mathbb{T}) \right] \quad (\text{EC.11b})$$

$$= \sum_{k \in \mathcal{K}} \xi_{pk} q_{pkt}^\Pi \left[R_t(\mathbf{b}_t, A_t^\Pi) + \bar{p}_t(\mathbb{T} | \mathbf{b}_t, A_t^\Pi) \bar{R}_t(\mathbb{T}) \right] \quad (\text{EC.11c})$$

$$= \left[R_t(\mathbf{b}_t, A_t^\Pi) + \bar{p}_t(\mathbb{T} | \mathbf{b}_t, A_t^\Pi) \bar{R}_t(\mathbb{T}) \right] \prod_{t'=1}^{t-1} \left(1 - \bar{p}_{t'}(\mathbb{T} | \mathbf{b}_{t'}, A_{t'}^\Pi) \right) \quad (\text{EC.11d})$$

where $1_{\{x\}}$ returns 1 if x is true and 0 otherwise. (EC.11a) is according to $u_{pkt} = q_{pkt} w_{pt}$ (2c) and $v_{pkt} = q_{pkt} w_{pt} \sigma$ (2d). (EC.11b) is according to the conditions that $\hat{\pi}_{pkt} = \tilde{\pi}_{pt}$ and $r_{pkt} = R_t(\mathbb{T})$ for all k . (EC.11c) is according to the definitions of $R_t(\mathbf{b}_t, A_t^\Pi)$ and $\bar{p}_t(\mathbb{T} | \mathbf{b}_t, A_t^\Pi) \bar{R}_t$. The last equation, (EC.11d), is according to Lemma EC.1, which is given following this proof.

Step 2: we prove that $\Delta Q_{T+1}^\Pi = \Delta V_{T+1}^\Pi$. According to (2b), (2d), and $d_{pt} = 0$, we have $v_{pkt}^\Pi = 0$ for all $\bar{T} + 1 \leq t \leq T$, and thus, $\sum_{t'=k}^T v_{pkt'}^\Pi = \sum_{t'=k}^{\bar{T}} v_{pkt'}^\Pi$. Further, $\sum_{k \in \mathcal{K}} \xi_{pk} \left(1 - \sum_{t'=k}^T v_{pkt'}^\Pi\right) r_{pk, T+1} = \sum_{k \in \mathcal{K}} \xi_{pk} \left(1 - \sum_{t'=k}^{\bar{T}} v_{pkt'}^\Pi\right) R_{T+1}(\mathbb{T}) = \prod_{t'=1}^{T-1} \left(1 - \bar{p}_{t'}(\mathbb{T} | \mathbf{b}_{t'}, A_{t'}^\Pi)\right) R_{T+1}(\mathbb{T})$.

Combining the steps 1-2, the POMDP-surveillance model and \mathbf{P} achieve the same expected reward in every period of the surveillance, and thus they achieve the same total expected reward under policy Π , which completes the proof. \square

LEMMA EC.1. *The total probability of undetected PCa in surveillance at period t is equivalent to the following probability in the POMDP-surveillance model:*

$$\sum_{k \in \mathcal{K}} q_{pkt}^\Pi \xi_{pk} = \prod_{t'=1}^{t-1} \left(1 - \bar{p}_{t'}(\mathbb{T} | \mathbf{b}_{t'}, A_{t'}^\Pi) \right), \quad \forall 1 \leq t \leq \bar{T}.$$

Proof. Recall that $\bar{p}_t(\mathbb{T} \mid \mathbf{b}_t, \mathbf{W}) = 0$, and $d_{pt} = 1$ for all $1 \leq t \leq \bar{T}$, we then prove this lemma using induction by the following two steps:

Step 1: According to (2b), $\sum_{k \in \mathcal{K}} q_{pk1} \xi_{pk} = \sum_{k \in \mathcal{K}} d_{p1} \xi_{pk} = 1$, and thus, Lemma EC.1 is true.

Step 2: Given Lemma EC.1 is true for t , we prove that it holds for $t + 1$.

According to (2b), $\sum_{k \in \mathcal{K}} q_{pkt}^{\Pi} \xi_{pk}$ can be rewritten as follows:

$$\begin{aligned} \sum_{k \in \mathcal{K}} q_{pkt}^{\Pi} \xi_{pk} &= \sum_{k \in \mathcal{K}} \left(1 - \sum_{t'=k}^t v_{pkt'}^{\Pi} \right) \xi_{pk} \\ &= \sum_{k \in \mathcal{K}} \left(1 - \sum_{t'=k}^{t-1} v_{pkt'}^{\Pi} - v_{pkt}^{\Pi} \right) \xi_{pk} \\ &= \sum_{k \in \mathcal{K}} \left(q_{pkt}^{\Pi} - 1_{\{k \leq t\}} q_{pkt}^{\Pi} w_{pt}^{\Pi} \sigma \right) \xi_{pk} \\ &= \sum_{k \in \mathcal{K}} q_{pkt}^{\Pi} \xi_{pk} \left(1 - 1_{\{k \leq t\}} w_{pt}^{\Pi} \sigma \right) \end{aligned}$$

where $1_{\{x\}}$ returns 1 if x is true and 0 otherwise. Observe that $1_{\{k \leq t\}} w_{pt}^{\Pi} \sigma$ is probability of newly detected PCa at period t , and thus, $1_{\{k \leq t\}} w_{pt}^{\Pi} \sigma = \bar{p}_t(\mathbb{T} \mid \mathbf{b}_t, A_t^{\Pi})$ for all k . Moreover, $\sum_{k \in \mathcal{K}} q_{pkt}^{\Pi} \xi_{pk}$ is equal to $\prod_{t'=1}^{t-1} \left(1 - \bar{p}_{t'}(\mathbb{T} \mid \mathbf{b}_{t'}, A_{t'}^{\Pi}) \right)$ by induction at period t . We can further rewrite $\sum_{k \in \mathcal{K}} q_{pkt}^{\Pi} \xi_{pk}$ as follows:

$$\begin{aligned} \sum_{k \in \mathcal{K}} q_{pkt}^{\Pi} \xi_{pk} &= \left(1 - \bar{p}_t(\mathbb{T} \mid \mathbf{b}_t, A_t^{\Pi}) \right) \prod_{t'=1}^{t-1} \left(1 - \bar{p}_{t'}(\mathbb{T} \mid \mathbf{b}_{t'}, A_{t'}^{\Pi}) \right) \\ &= \prod_{t'=1}^t \left(1 - \bar{p}_{t'}(\mathbb{T} \mid \mathbf{b}_{t'}, A_{t'}^{\Pi}) \right) \end{aligned}$$

Therefore, Lemma EC.1 is true for $t + 1$, which completes the proof. \square

EC.2.4. Proof of Proposition 1

Proof: let $x_{\hat{\mathcal{P}}}^*$ denote the optimal solution that maximizes $\hat{Q}(\hat{\mathcal{P}})$. Then, $x_p = x_{\hat{\mathcal{P}}}^*$ is a feasible solution to $\hat{Q}(\{p\})$ for all $p \in \hat{\mathcal{P}}$, and they achieve the same reward with $\hat{Q}(\hat{\mathcal{P}})$. Therefore, $\sum_{p \in \hat{\mathcal{P}}} \hat{Q}(\{p\}) \geq \hat{Q}(\hat{\mathcal{P}})$. \square

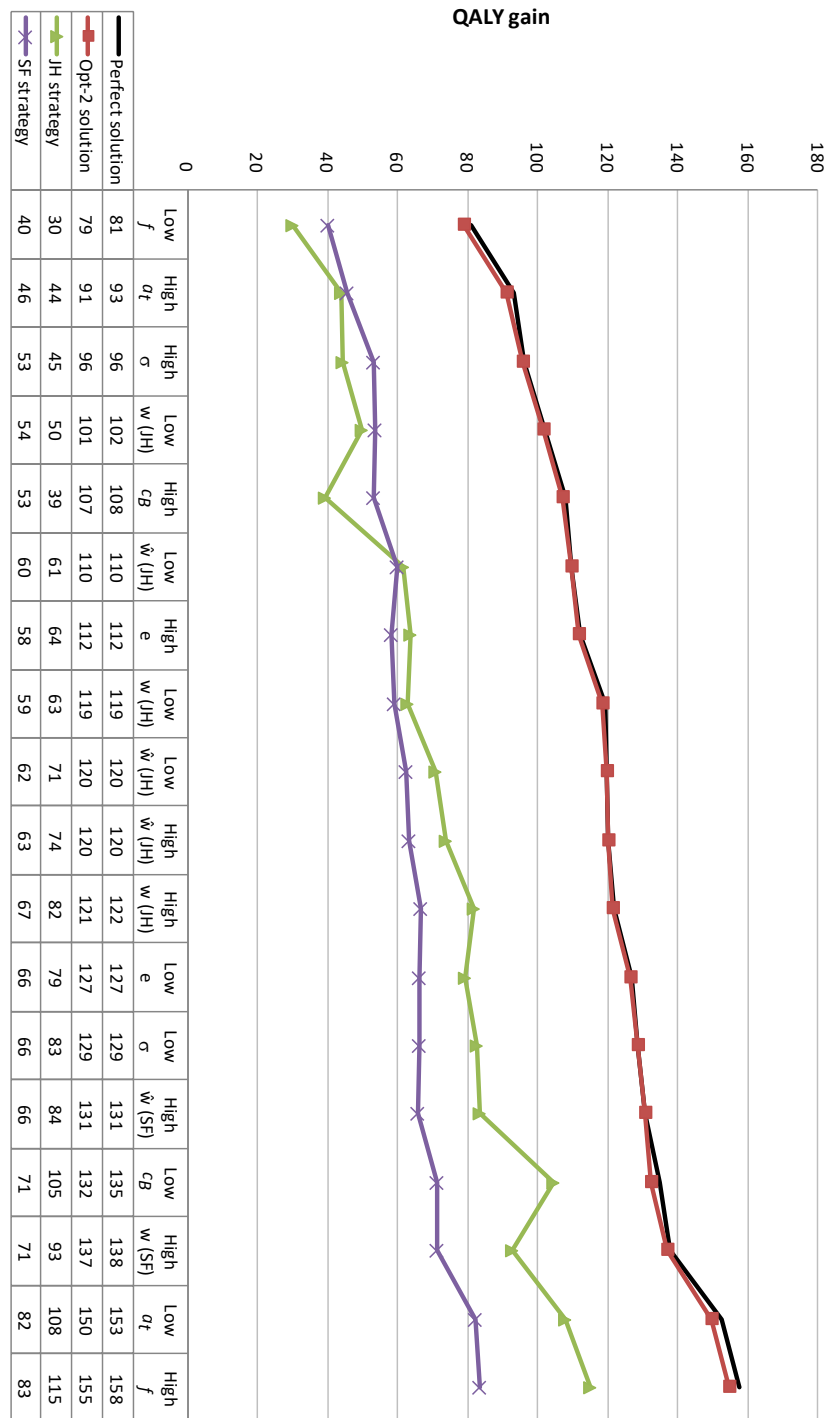


Figure EC.6 One-way sensitivity analyses of gain in expected QALYs compared with the UT strategy, where the baseline gain is 120 QALYs based on parameters in Table 3, and the x-axis is ordered by increasing gains of the *Opt-2* solution