ABSTRACT

UNDERWOOD, DANIEL JACOB. Risk-Based Simulation Optimization of PSA-Based Prostate Cancer Screening. (Under the direction of Dr. Brian T. Denton and Dr. James R. Wilson.)

Prostate cancer (PCa) is a serious chronic disease affecting a large number of men and is the second leading cause of men’s cancer deaths in the United States. We present a screening simulation model composed of the following: (i) a PCa natural-history submodel based on a discrete-event stochastic process representing a patient’s progression through underlying health states over his lifetime; and (ii) a statistical change-point submodel representing a patient’s prostate-specific antigen (PSA) level over time. Using specific risk-based parameterizations of this screening simulation model, we search for improved PSA-based PCa screening strategies for certain well-known risk groups based on race (white and African American), family history of PCa, and different levels of comorbid medical conditions.

We first demonstrate how the careful use of common random numbers (synchronized patient histories) allows for more precise estimation of NNS, the expected number of patients needed to be screened in order to prevent 1 death from PCa. We validate the simulation model by comparing model estimates of NNS and other statistics with corresponding estimates from the literature. By comparing 14 strategies from the literature, we found that the strategy of screening annually from age 50 to 75 using the PSA threshold 2.5 ng/mL yielded the smallest estimated NNS.

Next, we present a quality-adjusted life-years (QALYs) parameterization of the simulation model that uses synchronized patient histories to estimate for a given screening strategy the expected QALYs gained (QG) by a patient relative to no screening. This development naturally leads to the statistic NNSQ, the expected number of patients needed to be screened to produce a net gain of 1 QALY in the screened population. We formulate an optimization model to find improved screening strategies with QG as the performance measure. Based on results from this model, we discuss how NNSQ (the reciprocal of QG) can be helpful to policy makers as an alternative to NNS, or as an auxiliary performance measure for evaluation of screening policies.

We develop a three-stage metaheuristic simulation-optimization method based on the combination of a global genetic algorithm (GA), a post-GA clean-up procedure, and an implementation of the COMPASS local search method. The composite global and local metaheuristic is designed to find strategies that yield large expected values of QG while also exhibiting a smoothness in the PSA-threshold time series that render the strategies more suited to actual clinical implementation. We use this method to search for improved strategies in each risk group based on maximizing estimated expected QG. In addition to this metaheuristic, we develop an analytical formulation of a simplified, single-period PCa screening model. Results from both the metaheuristic and the analytical approach suggested that PSA is not beneficial for deciding whether men should have a biopsy. Moreover, our results suggest that for all 5 risk groups a routine biopsy between the ages of 45 and 60, regardless of PSA level, maximizes expected QG.
Finally, we develop a binary encoding scheme using a GA that by construction guarantees perfectly-smooth optimized strategies without requiring the overhead of the COMPASS local search method. Using this alternative approach, we found results generally concurring with the results from using the three-stage approach. In particular, using the binary encoding scheme, we found the near-optimal strategies for all risk groups to consist of biopsying virtually all men, and we found that patients with moderate to severe comorbidities should undergo biopsy earlier in life than patients in other risk groups. We also found the binary encoding approach was just as effective as the three-stage approach in finding strategies that yielded large expected values of QG, but the binary encoding scheme was computationally more efficient.
Risk-Based Simulation Optimization of PSA-Based Prostate Cancer Screening

by
Daniel Jacob Underwood

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Industrial Engineering

Raleigh, North Carolina
2015

APPROVED BY:

__________________________  ____________________________
Dr. Julie S. Ivy          Dr. John W. Baugh

__________________________  ____________________________
Dr. Brian T. Denton          Dr. James R. Wilson
Co-chair of Advisory Committee  Co-chair of Advisory Committee
# TABLE OF CONTENTS

**LIST OF TABLES** ....................................................................................... iv  
**LIST OF FIGURES** .................................................................................... ix  
**LIST OF ALGORITHMS** .......................................................................... xiii  

Chapter 1 Introduction .................................................................................. 1  

Chapter 2 Literature Review ....................................................................... 7  
  2.1 Prostate Cancer Screening ................................................................. 7  
    2.1.1 Guidelines ................................................................................... 7  
    2.1.2 Randomized Control Trials ......................................................... 9  
    2.1.3 Prostate Cancer Screening Models ........................................... 10  
  2.2 Simulation Optimization .................................................................... 15  
  2.3 Contributions of this Dissertation .................................................... 21  

Chapter 3 Simulation of PSA-Based Prostate Cancer Screening for Heterogeneous Patient Risk Groups ................................................. 24  
  3.1 Introduction ....................................................................................... 24  
  3.2 Methodology ..................................................................................... 25  
    3.2.1 Patient Health-State Stochastic Process ................................... 26  
    3.2.2 Model Parameters ..................................................................... 36  
    3.2.3 Screening Strategies ................................................................. 42  
    3.2.4 Estimation Methods .................................................................. 44  
    3.2.5 Procedural Description of Simulation Model ............................ 52  
  3.3 Results .............................................................................................. 54  
    3.3.1 Numerical Estimates for Model Parameters ............................ 54  
    3.3.2 Model Validation ...................................................................... 58  
    3.3.3 Numerical Results Using Base-case Parameter Values ........... 67  
    3.3.4 Sensitivity Analysis ................................................................... 76  
  3.4 Discussion .......................................................................................... 81  

Chapter 4 A Three-Stage Regularization Approach for the Optimization of PSA-Based Prostate Cancer Screening ......................................... 83  
  4.1 Introduction ....................................................................................... 83  
  4.2 Methodology ..................................................................................... 86  
    4.2.1 Single-Period Optimal Screening Analytical Model .................. 86  
    4.2.2 QALYs Parameterized Simulation Model ................................. 107  
    4.2.3 Number Needed to Screen to Gain 1 QALY (NNSQ) ............... 108  
    4.2.4 Prostate Cancer Screening Model ............................................. 110  
    4.2.5 Optimization Model .................................................................. 115  
    4.2.6 Regularizing Objective Function .............................................. 118  
    4.2.7 Three-Stage Simulation-Optimization Method ....................... 125
4.3 Results ................................................................. 147
   4.3.1 Test Problems for the GA ................................. 147
   4.3.2 Optimization Results for Basecase Parameter Values . 154
   4.3.3 Single-Period Screening Model Analysis .................. 162
   4.3.4 Sensitivity Analysis on Maximum Number of Biopsies .... 168
   4.3.5 Sensitivity Analysis on Model Parameters ................. 170
4.4 Discussion ......................................................... 180
   4.4.1 PSA Screening ................................................ 180
   4.4.2 Disease-Screening Modeling and Methodology ........... 183
   4.4.3 Limitations ................................................. 184

Chapter 5  Binary Encoding Scheme to Produce Perfectly-Regular Prostate Cancer Screening Strategies ........................................... 186
5.1 Introduction ..................................................... 186
5.2 Methodology ..................................................... 188
   5.2.1 Encoding Scheme ........................................ 188
   5.2.2 Examples of Encoded Screening Strategies ............... 192
   5.2.3 Optimization Model ....................................... 196
   5.2.4 Simulation Optimization Method ......................... 197
5.3 Results ............................................................ 199
   5.3.1 Optimization Results for Basecase Parameter Values ... 200
   5.3.2 Sensitivity Analysis on Model Parameters ................. 207
5.4 Discussion ......................................................... 216

Chapter 6  Conclusions and Recommendations ............................. 217
6.1 Main Conclusions of the Research .................................. 217
6.2 Limitations and Future Research ................................... 220

REFERENCES .......................................................... 222
| Table 3.1 | Definitions of parameters for transition probability matrices. | 31 |
| Table 3.2 | Parameters involved in the risk-specific parameter derivations for the African American, PCa Family History, (CCI=1), and (CCI≥2) risk groups. | 37 |
| Table 3.3 | Multiplicative adjustment ratios used to derive key parameters of the nonreference risk groups. The term “onset probability” is synonymous with the term “probability of developing PCa.” | 41 |
| Table 3.4 | PSA screening strategies compared in this study. Strategy #1 is no screening. Strategies #1–#8 were taken from a prominent simulation study [86]. Strategy #5 was also used in the PLCO trial. Strategies #9–#14 were used in the ERSPC trial.) | 43 |
| Table 3.5 | Population means and variances of normally-distributed random variables in the log-PSA growth model expressed in Equation (3.10). The values are taken directly from Gulati et al. [40]. | 50 |
| Table 3.6 | Parameters Used in Our PCa Natural-History Simulation, along with Sources for All Parameter Estimates. The value of 1.0 as the probability of other-cause death is a modest assumption made in our model only governing transitions at age 100 years. | 54 |
| Table 3.7 | Lower-bound numerical estimates for parameter $w_t$. The time $t$ (age $40+t$) denotes that the corresponding $w_t$ value applies to transitions occurring during the interval $[t, t+1)$. | 55 |
| Table 3.8 | Basecase numerical estimates for parameter $w_t$. The time $t$ (age $40+t$) denotes that the corresponding $w_t$ value applies to transitions occurring during the interval $[t, t+1)$. | 56 |
| Table 3.9 | Upper-bound numerical estimates for parameter $w_t$. The time $t$ (age $40+t$) denotes that the corresponding $w_t$ value applies to transitions occurring during the interval $[t, t+1)$. | 57 |
| Table 3.10 | Comparing model-estimated PCa mortality and diagnosis rates, lifespan, and metastasized PCa survival time to corresponding estimates from the literature. Whites means white US males with no family history of PCa and CCI=0, AA means African American males with no family history of PCa and CCI=0, and mPCa means metastasized PCa. Model estimates obtained by simulating 50,000 patients screened annually from ages 50 to 75 at PSA threshold 4.0 ng/mL, and 95% confidence-interval half widths are given for model estimates. | 59 |
| Table 3.11 | Description of patient and screening characteristics used to simulate each of the 7 major study centers in the ERSPC trial. These quantities are taken from Schröder et al. [95]. | 60 |
| Table 3.12 | Estimates of the expected NNS for screening in each of the compared risk groups using the screening strategies #1–#14; 95% confidence interval (CI) half-widths are shown in parentheses. Recall that strategy #1 is no screening. | 69 |
| Table 3.13 | Estimates of Clinical Statistics from Screening with Strategy #6 in Each of The Compared Risk Groups (i)–(v). Results are based on samples of 100,000 simulated patients from the corresponding risk groups. For mean months of life saved, the 95% CIs are reported. | 70 |
Table 3.14 Clinical statistics for the whites risk group based on samples of 100,000 simulated patients under the corresponding screening strategies. CI HW means the 95% confidence-interval half width. Recall that strategy #1 is no screening. ............................... 71

Table 3.15 Clinical statistics for the African Americans risk group based on samples of 100,000 simulated patients under the corresponding screening strategies. CI HW means the 95% confidence-interval half width. Recall that strategy #1 is no screening. ............................... 72

Table 3.16 Clinical statistics for the Family History risk group based on samples of 100,000 simulated patients under the corresponding screening strategies. CI HW means the 95% confidence-interval half width. Recall that strategy #1 is no screening. ............................... 73

Table 3.17 Clinical statistics for the CCI=1 risk group based on samples of 100,000 simulated patients under the corresponding screening strategies. CI HW means the 95% confidence-interval half width. Recall that strategy #1 is no screening. ............................... 74

Table 3.18 Clinical statistics for the CCI≥2 risk group based on samples of 100,000 simulated patients under the corresponding screening strategies. CI HW means the 95% confidence-interval half width. Recall that strategy #1 is no screening. ............................... 75

Table 3.19 NNS estimates from sensitivity analysis over the annual probability of other-cause death ($d_t$). For each risk group, the headings “low” and “high” respectively denote the low and high levels of parameter $d_t$. The 95% CI estimates are based on simulated samples of 1 million patients. ................................................................. 77

Table 3.20 NNS estimates from sensitivity analysis over the annual probability of developing PCa ($w_t$). For each risk group, the headings “low” and “high” respectively denote the low and high levels of parameter $w_t$. The 95% CI estimates are based on simulated samples of 1 million patients. ................................................................. 78

Table 3.21 NNS estimates from sensitivity analysis over the annual metastasis probability for patients in state C ($e_t$). For each risk group, the headings “low” and “high” respectively denote the low and high levels of parameter $e_t$. The 95% CI estimates are based on simulated samples of 1 million patients. ................................................................. 79

Table 3.22 NNS estimates from sensitivity analysis over the PCa-screening leadtime ($t_{LT}$). For each risk group, the headings “low” and “high” respectively denote the low and high levels of parameter $t_{LT}$. The 95% CI estimates are based on simulated samples of 1 million patients. ................................................................. 80

Table 4.1 QALY disutilities in the PCa screening simulation model. ................................................................. 108

Table 4.2 Decision-variable ($x_t$) value to PSA-screening threshold rule mappings. ................................................................. 117

Table 4.3 Basecase numerical estimates for QALY disutilities in the natural history model. ............................... 154

Table 4.4 Comparison of the screening strategies from Chapter 3 with the optimized strategy for the whites risk group, $x_{W}^*$, based on estimated expected QG. Estimates are shown with 95% CIs and based on populations (samples) of $10^7$ independently simulated patients. ................................................................. 157

Table 4.5 Comparison of the screening strategies from Chapter 3 with the optimized strategy for the African American risk group, $x_{AA}^*$, based on estimated expected QG. Estimates are shown with 95% CIs and based on populations (samples) of $10^7$ independently simulated patients. ................................................................. 158
Table 4.6 Comparison of the screening strategies from Chapter 3 with the optimized strategy for the family history risk group, $\hat{x}_{FH}$, based on estimated expected QG. Estimates are shown with 95% CIs and based on populations (samples) of $10^7$ independently simulated patients.

Table 4.7 Comparison of the screening strategies from Chapter 3 with the optimized strategy for the CCI=1 risk group, $\hat{x}_{C1}$, based on estimated expected QG. Estimates are shown with 95% CIs and based on populations (samples) of $10^7$ independently simulated patients.

Table 4.8 Comparison of the screening strategies from Chapter 3 with the optimized strategy for the CCI≥2 risk group, $\hat{x}_{C2}$, based on estimated expected QG. Estimates are shown with 95% CIs and based on populations (samples) of $10^7$ independently simulated patients.

Table 4.9 Definitions of parameters for transition probability matrices.

Table 4.10 Numerical estimates used as the low and high levels for QALY disutilities in the sensitivity analysis.

Table 4.11 Expected QG estimates from one-way sensitivity analysis over model parameters for the White risk group. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}, D_{Dia}, D_{Biop}, D_{Tre},$ and $D_{Met}$) were simultaneously varied to their lowest or highest plausible levels.

Table 4.12 Expected QG estimates from one-way sensitivity analysis over model parameters for the African Americans risk group. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}, D_{Dia}, D_{Biop}, D_{Tre},$ and $D_{Met}$) were simultaneously varied to their lowest or highest plausible levels.

Table 4.13 Expected QG estimates from one-way sensitivity analysis over model parameters for the Family History risk group. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}, D_{Dia}, D_{Biop}, D_{Tre},$ and $D_{Met}$) were simultaneously varied to their lowest or highest plausible levels.

Table 4.14 Expected QG estimates from one-way sensitivity analysis over model parameters for the CCI=1 risk group. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}, D_{Dia}, D_{Biop}, D_{Tre},$ and $D_{Met}$) were simultaneously varied to their lowest or highest plausible levels.

Table 4.15 Expected QG estimates from one-way sensitivity analysis over model parameters for the CCI≥2 risk group. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}, D_{Dia}, D_{Biop}, D_{Tre},$ and $D_{Met}$) were simultaneously varied to their lowest or highest plausible levels.

Table 5.1 Samples of optimized strategies produced by the binary-encoding–scheme method for each of the compared risk groups using the basecase parameter estimates. Each QG estimate is based on a sample of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. PSA thresholds have units ng/mL, and ages and screening frequency have units years. The screening frequency means the time between screen events.
**Table 5.2** Sample statistics on the optimized strategies produced by the binary-encoding–scheme method for each of the compared risk groups using the basecase parameter estimates. The CIs are estimated using the Student-t distribution. PSA thresholds have units ng/mL, and ages and screening frequency have units years. The screening frequency means the time between screen events.

**Table 5.3** Means-based strategies constructed on the basis of the sample means from Table 5.2 of the following quantities: the age to start screening; the age to stop screening; the average PSA threshold; and the frequency of screening.

**Table 5.4** Expected QG estimates across risk groups using basecase parameter levels comparing strategies #1–#14 from Chapter 3 to the best strategies found by optimization in Chapters 4 (the strategies in Figures 4.15a through 4.15e), and the representative strategies from Chapter 5 (the strategies in Table 5.3). Each estimate is based on a sample of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. For each risk group, the labels “opt4” and “opt5” denote, respectively, the corresponding optimized strategy from Chapter 4 and representative strategy from Chapter 5.

**Table 5.5** Expected QG estimates from one-way sensitivity analysis over model parameters for the Whites risk group, based on samples of 5 independent runs of the GA for each risk-group–experimental-configuration combination. Each estimate is based on a sample of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. The symbol $D_\star$ denotes that all QALY disutilities ($D_{Scr}$, $D_{Dia}$, $D_{Biop}$, $D_{Tre}$, and $D_{Met}$) were simultaneously varied to their greatest or smallest plausible values. Seed means the random number seed used in the construction of the initial population for the GA and in the crossover and mutation operators of the GA.

**Table 5.6** Expected QG estimates from one-way sensitivity analysis over model parameters for the African Americans risk group, based on samples of 5 independent runs of the GA for each risk-group–experimental-configuration combination. Each estimate is based on a sample of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. The symbol $D_\star$ denotes that all QALY disutilities ($D_{Scr}$, $D_{Dia}$, $D_{Biop}$, $D_{Tre}$, and $D_{Met}$) were simultaneously varied to their greatest or smallest plausible values. Seed means the random number seed used in the construction of the initial population for the GA and in the crossover and mutation operators of the GA.

**Table 5.7** Expected QG estimates from one-way sensitivity analysis over model parameters for the Family History risk group, based on samples of 5 independent runs of the GA for each risk-group–experimental-configuration combination. Each estimate is based on a sample of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. The symbol $D_\star$ denotes that all QALY disutilities ($D_{Scr}$, $D_{Dia}$, $D_{Biop}$, $D_{Tre}$, and $D_{Met}$) were simultaneously varied to their greatest or smallest plausible values. Seed means the random number seed used in the construction of the initial population for the GA and in the crossover and mutation operators of the GA.
Table 5.8 Expected QG estimates from one-way sensitivity analysis over model parameters for the CCI=1 risk group, based on samples of 5 independent runs of the GA for each risk-group–experimental-configuration combination. Each estimate is based on a sample of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}, D_{Dia}, D_{Biop}, D_{Tre},$ and $D_{Met}$) were simultaneously varied to their greatest or smallest plausible values. Seed means the random number seed used in the construction of the initial population for the GA and in the crossover and mutation operators of the GA.

Table 5.9 Expected QG estimates from one-way sensitivity analysis over model parameters for the CCI\(\geq2\) risk group, based on samples of 5 independent runs of the GA for each risk-group–experimental-configuration combination. Each estimate is based on a sample of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}, D_{Dia}, D_{Biop}, D_{Tre},$ and $D_{Met}$) were simultaneously varied to their greatest or smallest plausible values. Seed means the random number seed used in the construction of the initial population for the GA and in the crossover and mutation operators of the GA.

Table 5.10 Samples of optimized strategies produced by the binary-encoding method for the reference risk group at preclinical dwelling time ($t_{PD}$) levels of 0, 4, 8, and 12 years. PSA thresholds have units ng/mL, and ages and screening frequency have units years. The screening frequency means the time between screen events.
LIST OF FIGURES

Figure 2.1 Illustrating how solutions are sampled and the most promising area is defined in our COMPASS implementation. .......................................................... 18

Figure 3.1 Health states and progression paths in the simulation model. Transitions between states are represented by arrows. .................................................. 27

Figure 3.2 Sequence of two decision-making stages that take place during each period in the simulation. ................................................................. 29

Figure 3.3 Illustration of the leadtime clock corresponding to a hypothetical patient with PCa onset at age 42 and with the PSA levels shown from age 40 to 61. .......... 48

Figure 3.4 Illustration depicting the divergence between patient histories in the screening and no-screening simulations. The hypothetical patient depicted in this illustration is shown simulated over 7 periods (years). ........................................ 52

Figure 3.5 Lower-bound, base-case, and upper-bound numerical estimates of parameter \( w \) inferred from Figure 2 of Haas et al. [44]. .................................................. 55

Figure 3.6 Number of patients simulated for each ERSPC study center. These values are equal to the minimums of the sizes of the screen and control groups reported for each center. 60

Figure 3.7 Simulation estimates of NNI in each ERSPC trial study center. .............................. 62

Figure 3.8 Simulation estimates of NNI in the Netherlands study center of the ERSPC trial. . 63

Figure 3.9 Simulation estimates of NNI in the Belgium study center of the ERSPC trial. . 63

Figure 3.10 Simulation estimates of NNI in the Sweden study center of the ERSPC trial. .... 64

Figure 3.11 Simulation estimates of NNI in the Finland study center of the ERSPC trial. .... 64

Figure 3.12 Simulation estimates of NNI in the Italy study center of the ERSPC trial. .... 65

Figure 3.13 Simulation estimates of NNI in the Spain study center of the ERSPC trial. .... 65

Figure 3.14 Simulation estimates of NNI in the Switzerland study center of the ERSPC trial. ... 66

Figure 3.15 Comparison of the aggregated NNI estimates with the NNI values reported by the ERSPC trial. .......................................................... 67

Figure 3.16 The Estimated NNS to Avert 1 PCa Death for Each Risk Group (i)–(v). Also displayed are 95% CIs for expected NNS in each risk group based on a sample of 1 million patients. The NNS estimates and associated 95% CIs for risk groups (i)–(v) were, respectively as follows: 137 [134, 140]; 63 [62, 64]; 111 [109, 113]; 178 [173, 183]; 225 [218, 232]. .................................................. 68

Figure 4.1 Sequence of events in the simulation model in a given time period. PSA testing, prostate biopsy, and definitive PCa treatment take place at the beginning of the period. 110

Figure 4.2 Illustration of a PSA threshold vector \( v \) containing exactly 2 elements, which necessarily exhibits an essentially perfectly smooth PSA threshold pattern. ........... 121

Figure 4.3 Illustration of a PSA threshold vector \( v \) exhibiting an essentially perfectly unsmooth PSA threshold pattern. .................................................. 122
Figure 4.4 Flowchart of three-stage simulation-optimization method. First, the GA broadly explores the feasible region to find a strategy with large expected QG. Then, the clean-up procedure systematically evaluates the strategy from the GA and discards practically insignificant PSA thresholds. Finally, the COMPASS method tries to make regularizing (smoothing) improvements to the cleaned-up strategy without a substantial loss of expected QG.

Figure 4.5 High-level flowchart of the GA. Beginning with a population (set) of initial screening strategies, the GA simulates (evaluates) the population. Then, until some predetermined stopping criterion is met, the GA repeatedly constructs a new population by application of the selection, crossover, and mutation operators, and then simulates the new population.

Figure 4.6 Function call-graph diagram illustrating the organization of algorithms for which pseudocode is provided. The blue boxes beside the abstract functions RECOMBINE, SELECTION, and CROSSOVER list the corresponding concrete functions to which the abstract functions may refer.

Figure 4.7 Illustration of the selection, crossover, and mutation operators in the GA. To construct a single new screening strategy, the GA selects two existing strategies which are then transformed by application of the crossover operator into a single strategy, which is further modified by application of the mutation operator.

Figure 4.8 Screening strategy with practically-insignificant residual DNA from the GA.

Figure 4.9 QALYs-gained observations (i.e., $D_j$ values) for all screen-detected patients who benefited from screening under the strategy depicted in Figure 4.8.

Figure 4.10 Screening strategy from Figure 4.8 after having the practically-insignificant thresholds removed by the clean-up procedure.

Figure 4.11 Comparison of the 3 recombination strategies for the GA on the ACKL test problem.

Figure 4.12 Comparison of the 3 recombination strategies for the GA on the BELA test problem.

Figure 4.13 Comparison of the 3 recombination strategies for the GA on the SALO test problem.

Figure 4.14 Comparison of the 3 recombination strategies for the GA on the GRWK test problem.

Figure 4.15 Optimized strategies $\hat{x}_W^*$, $\hat{x}_{AA}^*$, $\hat{x}_{FH}^*$, $\hat{x}_{C1}^*$, and $\hat{x}_{C2}^*$ produced by the three-stage method for the White, African American, Family History, CCI=1, and CCI$\geq$2 risk groups, respectively, using the basecase parameter estimates.

Figure 4.16 Estimated expected QG for each of the compared risk groups for all the strategies from Chapter 3 as well as the strategies produced by the three-stage method with basecase parameter values. For each risk group, the label “opt” denotes the corresponding optimized strategy from Figures 4.15a through 4.15e.

Figure 4.17 Comparison of the estimated expected NNSQ for the basecase optimized strategies depicted in Figures 4.15a through 4.15e based on samples of $10^7$ simulated patients from the corresponding risk groups. The 95% CIs reflect inter-patient stochastic variability; they do not reflect generalized uncertainty in any of the model parameters.

Figure 4.18 Comparison of the estimated expected NNS for the basecase optimized strategies depicted in Figures 4.15a through 4.15e based on samples of $10^7$ simulated patients from the corresponding risk groups. The 95% CIs reflect inter-patient stochastic variability; they do not reflect generalized uncertainty in any of the model parameters.
Figure 4.19 Plots of the \( \hat{\vartheta} (\phi) \) function showing the effect of variation in the certain model parameters on the single-period expected reward for screening 53-year-old men from the reference risk group (the White risk group) at different PSA thresholds \( (\phi = 0.0, 0.5, \ldots, 10.0 \text{ ng/mL}) \). .......................................................... 164

Figure 4.20 Plots of the \( \hat{\vartheta} (\phi) \) function showing the effect of variation in the certain model parameters on the single-period expected reward for screening 69-year-old men from the reference risk group (the White risk group) at different PSA thresholds \( (\phi = 0.0, 0.5, \ldots, 10.0 \text{ ng/mL}) \). .......................................................... 167

Figure 4.21 Optimized strategies produced by the three-stage method for the White, African American, Family History, CCI\( =1 \), and CCI\( \geq 2 \) risk groups, respectively, where no patient is permitted more than 1 prostate biopsy. .......................................................... 169

Figure 4.22 Optimized strategies produced by the three-stage method for the White risk group at different experiments in the sensitivity analysis. The symbol \( \mathbb{D}_\bullet \) denotes that all QALY disutilities \( (\mathbb{D}_{\text{Scr}}, \mathbb{D}_{\text{Dia}}, \mathbb{D}_{\text{Biop}}, \mathbb{D}_{\text{Tre}}, \text{ and } \mathbb{D}_{\text{Met}}) \) were simultaneously varied to their lowest or highest plausible levels. .......................................................... 174

Figure 4.23 Optimized strategies produced by the three-stage method for the African American risk group at different experiments in the sensitivity analysis. The symbol \( \mathbb{D}_\bullet \) denotes that all QALY disutilities \( (\mathbb{D}_{\text{Scr}}, \mathbb{D}_{\text{Dia}}, \mathbb{D}_{\text{Biop}}, \mathbb{D}_{\text{Tre}}, \text{ and } \mathbb{D}_{\text{Met}}) \) were simultaneously varied to their lowest or highest plausible levels. .......................................................... 175

Figure 4.24 Optimized strategies produced by the three-stage method for the Family History risk group at different experiments in the sensitivity analysis. The symbol \( \mathbb{D}_\bullet \) denotes that all QALY disutilities \( (\mathbb{D}_{\text{Scr}}, \mathbb{D}_{\text{Dia}}, \mathbb{D}_{\text{Biop}}, \mathbb{D}_{\text{Tre}}, \text{ and } \mathbb{D}_{\text{Met}}) \) were simultaneously varied to their lowest or highest plausible levels. .......................................................... 176

Figure 4.25 Optimized strategies produced by the three-stage method for the CCI\( =1 \) risk group at different experiments in the sensitivity analysis. The symbol \( \mathbb{D}_\bullet \) denotes that all QALY disutilities \( (\mathbb{D}_{\text{Scr}}, \mathbb{D}_{\text{Dia}}, \mathbb{D}_{\text{Biop}}, \mathbb{D}_{\text{Tre}}, \text{ and } \mathbb{D}_{\text{Met}}) \) were simultaneously varied to their lowest or highest plausible levels. .......................................................... 177

Figure 4.26 Optimized strategies produced by the three-stage method for the CCI\( \geq 2 \) risk group at different experiments in the sensitivity analysis. The symbol \( \mathbb{D}_\bullet \) denotes that all QALY disutilities \( (\mathbb{D}_{\text{Scr}}, \mathbb{D}_{\text{Dia}}, \mathbb{D}_{\text{Biop}}, \mathbb{D}_{\text{Tre}}, \text{ and } \mathbb{D}_{\text{Met}}) \) were simultaneously varied to their lowest or highest plausible levels. .......................................................... 178

Figure 4.27 Optimized strategies produced by the three-stage method for the reference risk group at different levels of the preclinical dwelling time \( (t_{\text{PD}}) \) parameter. .......................................................... 180

Figure 5.1 PSA-threshold–based screening strategy decoded from \( \tilde{\xi}_1 \). .......................................................... 194

Figure 5.2 PSA-threshold–based screening strategy decoded from \( \tilde{\xi}_2 \). .......................................................... 195

Figure 5.3 Estimated expected QG for each of the compared risk groups for all the strategies from Chapter 3, the basecase optimized strategies from Chapter 4, and the means-based strategies derived from the optimized strategy statistics in Table 5.2. For each risk group, the labels “opt4” and “opt5” denote, respectively, the corresponding optimized strategy from Chapter 4 and representative strategy from Chapter 5. .......................................................... 205
Comparison of the estimated expected NNSQ and NNS for the basecase optimized strategies in Chapters 4 and 5. The strategies from Chapter 4 are those depicted in Figures 4.15a through 4.15e. The strategies from Chapter 5 are the representative strategies described in Table 5.3. All NNSQ and NNS estimates are based on samples of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. The 95% CIs reflect inter-patient stochastic variability; they do not reflect generalized uncertainty in any of the model parameters.
LIST OF ALGORITHMS

Algorithm 1 \texttt{SIMULATION}(m) \hspace{1cm} 53
Algorithm 2 \texttt{SIMULATIONQALYs}(m) \hspace{1cm} 114
Algorithm 3 \texttt{PERIODQALYs} (screen, biopsy, age, ageAtTreatment, state, lastState) \hspace{1cm} 115
Algorithm 4 \texttt{GENETICALGORITHM}(X) \hspace{1cm} 128
Algorithm 5 \texttt{RECOMBINE\_STANDARD}(X, GENERATION) \hspace{1cm} 131
Algorithm 6 \texttt{RECOMBINE\_DETERMINISTICCROWDING}(X, GENERATION) \hspace{1cm} 132
Algorithm 7 \texttt{RECOMBINE\_MERITOCRACY}(X, GENERATION) \hspace{1cm} 133
Algorithm 8 \texttt{SELECT\_ROULETTE\_WHEEL}(X) \hspace{1cm} 135
Algorithm 9 \texttt{SELECT\_TOURNAMENT}(X) \hspace{1cm} 136
Algorithm 10 \texttt{SELECT\_UNIFORM}(X) \hspace{1cm} 136
Algorithm 11 \texttt{CROSSOVER\_ONE\_POINT}(PARENT1, PARENT2) \hspace{1cm} 137
Algorithm 12 \texttt{CROSSOVER\_TWO\_POINT}(PARENT1, PARENT2) \hspace{1cm} 138
Algorithm 13 \texttt{CROSSOVER\_UNIFORM}(PARENT1, PARENT2) \hspace{1cm} 138
Algorithm 14 \texttt{MUTATE}(\text{STRATEGY}) \hspace{1cm} 139
Algorithm 15 \texttt{CLEANUP}(x) \hspace{1cm} 143
Algorithm 16 \texttt{COMPASS}(\text{INITIAL\_STRATEGY}) \hspace{1cm} 146
Algorithm 17 \texttt{DECODE\_SCHEME}(\text{BITS}) \hspace{1cm} 192
Algorithm 18 \texttt{GENETICALGORITHM2}(X) \hspace{1cm} 198
Algorithm 19 \texttt{CLEANUP2}(\text{STRATEGY}) \hspace{1cm} 199
Introduction

In 2015 it is estimated that 220,800 new cases of prostate cancer (PCa) will be diagnosed and 27,540 deaths from PCa will occur in the United States [5]. Over the period from 2007 to 2011, African Americans were estimated to be 1.7 times as likely as white US males to be diagnosed with PCa, and 2.4 times as likely to die from PCa [5]. By these estimates, PCa is one of the most common types of cancer affecting US males, and one in which there are ethnic disparities. But screening for PCa is controversial. Some authorities do not recommend routine screening for PCa [78], while many who do recommend screening disagree on how best to screen [98, 16, 15]. Screening for PCa is controversial for two primary reasons: (i) tests to detect PCa are imperfect, and (ii) many researchers and clinicians are apprehensive of the possible harms incurred from overtreatment.

The standard way to screen for PCa is to administer a digital rectal examination (DRE) and a blood test to measure a patient’s level of prostate-specific antigen (PSA). By performing a DRE, a physician can sometimes detect palpable abnormalities (such as increased size) of the prostate, which can indicate PCa. However, DREs are subjective, and abnormalities in the prostate can be the result of benign prostatic hyperplasia (BPH), a physiologic process that occurs naturally with age. By performing a PSA test, a physician can monitor the level of PSA in a patient. High PSA levels are positively correlated with the presence of PCa, but also with BPH. So, as with abnormal DREs, high PSA levels can result in a falsely positive indication that the patient has PCa. Once a patient is suspected to have PCa by virtue of a positive PSA test or DRE, the next step is a prostate biopsy — a costly and invasive procedure that involves...
inserting large needles into the prostate gland — for more definitive determination of whether the patient has PCa.

Once PCa is identified by a positive prostate biopsy, the patient and physician may elect to treat the cancer. The early detection and treatment of prostate cancer can add decades to a patient’s life. However, because PCa is a very slow-growing disease, patients with prostate cancer often die of other causes before the cancer metastasizes. Thus, not all PCa treatments result in a gain in life expectancy for the patient. All treatments for PCa do, however, have a high probability of significant side effects from treatment, such as urinary and sexual dysfunction, and in rare cases severe infection and possibly death. Consequently, unnecessary treatment of PCa (referred to as overtreatment) can substantially degrade the quality of a patient’s life.

The use of QALYs is a common way of expressing the overall quality of a particular year of life as a number from 0 to 1. The value of 0 QALYs on one extreme represents death (i.e., no benefit of living whatsoever), and the value of 1 QALY on the other extreme represents a year of perfect health. Values between 0 and 1 represent a less-than-perfect year of life with respect to a patient’s subjective appraisal. By assigning decrements in units of QALYs to certain clinical interventions and undesirable health states, one can make model comparisons not solely on the basis of mortality but also on the basis of different non-fatal degradations to quality of life that can result from diagnosis or treatment of PCa.

There is controversy surrounding PCa screening and the potential impact of screening decisions on many people’s lives due to the potential for overtreatment of men with low-risk and likely indolent PCa. This controversy was caused, in part, by conflicting results from two large randomized trials that studied PSA screening [10, 95]. Much research has been devoted to studying how best to use PSA in screening for PCa. Yet, the question remains hotly debated due to problems with randomized trials including contamination of the nonscreened arms by men who were in fact screened [51, 66]. Moreover, these trial results had limited follow-up and do not account for quality-of-life effects for screened patients.

In this research, we examine if and how to utilize PSA in the course of screening for PCa using a simulation model that is carefully validated against randomized trial results. Next, we evaluate previously proposed screening strategies. Finally, we develop multiple optimization approaches for finding improved
screening strategies in terms of QALYs.

We conduct extensive sensitivity analysis (SA) of the results from optimization to uncertainties in certain key parameter estimates. We vary parameter estimates in two main ways. First, we selectively vary certain combinations of parameters in order to represent specific known risk groups. Second, in each of these risk groups, we systematically conduct one-way SA on a number of influential model parameters. This SA helps establish robustness in the inferences and conclusions we draw from this research.

After examining the results from two substantially different GA-based simulation optimization approaches, we found strong evidence that PSA was influenced by several physiological factors unrelated to PCa and was therefore not a reliable tool either for diagnosing PCa or for indicating the need to perform a biopsy. In fact, rather than using PSA to selectively biopsy a subset of men at some prescribed age, we found evidence that expected QALYs was maximized when all men undergo prostate biopsy (at an age that varies according to risk group). We also found that patients with comorbidities should undergo biopsy earlier in life than patients without comorbidities. These findings were also confirmed using the analytical model based on a formulation of a simplified PCa screening problem in which screening occurs only once for each patient at some defined age. Using simulation to estimate some of the terms in the solution to this analytical model, we found independent support for the conclusions drawn from the simulation optimization methods that virtually all men should undergo prostate biopsy.

The simulation model itself was validated by comparing model estimates with estimates from the literature for key statistics (e.g., patient life expectancy, metastasis survival time, and PCa-specific mortality). The leadtime and clinical detection components of the model were validated by comparing simulation estimates of expected NNS to corresponding estimates reported in the European Randomized Study of Screening for Prostate Cancer (ERSPC), one of the largest randomized control trials (RCTs) ever to study PCa screening [10, 95]. In addition to validating fundamental parts of the simulation model, we also validated the optimization results by comparing two different simulation optimization approaches with one another, and with an analytical formulation of a simplified model of PCa screening. The results from these three methods supported the same basic inference: virtually all men should undergo prostate biopsy without regard to patient PSA level at some point in their lives. Because of the extensive SA we
carried out, we claim that this inference is very unlikely to be an artifact of inaccurate estimation of model parameters.

This dissertation is structured as follows. In Chapter 2 we present a review of the literature on PCa modeling, screening guidelines, and simulation optimization. We discuss recent guidelines on prostate cancer screening issued by authoritative medical organizations and governmental bodies, prominent randomized control trials and empirical studies to evaluate the impact of PSA screening on prostate cancer mortality, and probabilistic models developed to study the merit of various prostate cancer screening strategies on mortality and quality-of-life. Next, we discuss local and global simulation-optimization approaches. Finally, we discuss the ways in which this dissertation contributes to the body of science on prostate cancer screening and simulation optimization.

In Chapter 3 we evaluate previously recommended PSA screening strategies for different risk groups based on NNS, the expected number needed to screen to avert 1 PCa death. A simulation model of the natural history of prostate cancer is used to evaluate PSA screening in the following risk groups: (i) white US males without a family history of prostate cancer and without comorbidities; (ii) African American males without a family history of prostate cancer and without comorbidities; (iii) white US males with a family history of prostate cancer; (iv) white US males with a Charlson Comorbidity Index (CCI) equal to 1; and (v) white US males with a CCI greater than or equal to 2. This modeling study is designed to answer the following questions:

(a) Which risk groups stand to benefit the least/most from screening in terms of PCa-mortality reduction?

(b) Considering a representative sample of different screening strategies from the literature, which strategy is most efficient for each risk group in reducing PCa mortality?

We found that there were significant differences in the potential benefit of screening for the compared risk groups, with the African American risk group (ii) standing to gain the most benefit and the comorbidity risk groups (iv) and (v) standing to gain the least benefit. Despite disparities among risk groups in potential benefit, the same strategy yielded the most PCa-specific-mortality reduction with the fewest patients needed to be screened (i.e., the smallest estimated expected NNS). Of the compared strategies, the best strategy for all risk groups (i)–(v) was to screen annually from age 50 to 75 using a PSA threshold
of 2.5 ng/mL. Therefore, our results suggest that if the sole consideration is reduction of PCa mortality, then a single “one-size-fits-all” strategy may be appropriate for a range of different risk groups.

In Chapter 4, we develop a QALYs-based parameterization of the simulation model from Chapter 3 and use a three-stage simulation-optimization approach to produce clinically realistic improved strategies when PCa mortality is considered along with the quality-of-life degradations stemming from screening and clinical intervention. We also formulate an analytical representation of a simplified single-period PCa screening problem that determines the optimal single-period threshold for screening in terms of maximizing estimated expected QALYs. This model, though much simpler than the real screening problem, is used for supplementary analysis. Following the formulation of this model, we derive point and confidence interval (CI) estimators for the expected reward in QALYs of single-period screening at a given threshold. As an analogue to NNS, we develop and derive a CI estimator for the statistic NNSQ, which is defined as the expected number of patients needed to be screened to produce a net gain in the screened population of 1 QALY. This statistic allows us to utilize variance reduction techniques to improve the efficiency of the simulation optimization method. This three-stage, QALYs-based, simulation-optimization approach is designed to answer the following questions:

(a) When QALYs are considered, which risk groups stand to benefit the least/most from PCa screening?

(b) When QALYs are considered, what is the best way to screen for PCa in each risk group?

Using this approach, we found that there were significant differences in the potential QALYs benefits for different risk groups, with the African American risk group standing to gain the most potential benefit and the comorbidity risk groups standing to gain the least potential benefit. We found the best strategy for each risk group, though not exactly the same across risk groups, was in each case a strategy calling for a biopsy of all men at a particular age (or ages). The most discernible difference between the best strategies for the different risk groups was that biopsy should take place earlier in life for patients with moderate to severe comorbidities (CCI ≥ 2).

In Chapter 5, we develop an alternative simulation-optimization approach that relies upon a binary-encoding scheme for the decision variables and that has a key advantage over the two-stage approach used in Chapter 4. By the binary-encoding scheme, this alternative approach produces clinically realistic
screening strategies without the need for an additional regularizing stage. This alternative approach is designed to answer primarily the same questions we sought to answer in Chapter 4. We use this approach to answer those same questions for the following main reasons: (i) future research on PCa screening may benefit from more than one approach for dealing with the tendency of random-search–based simulation-optimization methods to produce good screening strategies that nevertheless have clinically unrealistic features; and (ii) comparing the results from a second simulation-optimization approach confirms the robustness of our findings with respect to the optimization methodology. This is important due to the likely nonconvex nature of the optimization models. Using this second approach, we obtained results largely concurring with the results from Chapter 4. We also found this approach to be more efficient than the three-stage approach of Chapter 4.

Finally, in Chapter 6 we conclude by recapitulating the main findings and contributions of this dissertation and by discussing opportunities for extending this research in the future.
Literature Review

2.1 Prostate Cancer Screening

In reviewing the literature on PCa screening, we compare and contrast contemporary screening guidelines in Section 2.1.1. We then discuss methods and results from various randomized clinical trials on PCa screening in Section 2.1.2. Finally, in Section 2.1.3, we review PCa screening studies based on probability models.

2.1.1 Guidelines

Some groups, such as the US Preventive Services Task Force (USPSTF) and the American College of Preventive Medicine, recommend against routine PSA-based screening for prostate cancer, basing their recommendation on finding evidence suggesting that screening does not significantly reduce prostate cancer mortality in the general population [70, 78]. Other groups recommend screening in some situations. For example, the American Urological Association (AUA) recommends shared decision-making for men age 55 to 69 years, but does not recommend routine PSA-based screening at other ages [16]. The European Association of Urology (EAU) makes the following statement [47]:

Current evidence is insufficient to warrant widespread population-based screening by prostate-specific antigen (PSA) for PCa.
The National Comprehensive Cancer Network (NCCN) provides an algorithm for recommended PSA screening based on discussing risk factors and the possibility of a baseline PSA test. The recommended screening frequencies and PSA thresholds for biopsy are different for the age groups 45–75, and >75 years. And the recommendations stress that men over age 75 should be PSA-tested with caution, especially if they have comorbidities. The American Cancer Society (ACS) recommends that men with at least a 10-year life expectancy be presented around age 50 years with information to make a shared decision whether to undergo screening. They recommend that this discussion begin as early as age 40 or 45 for men at elevated risk for PCa (such as African Americans or men with a positive family history of PCa).

The ACS recommends that the historical 4.0 ng/mL cutoff point still be used, but also that patients with PSA between 2.5 ng/mL and 4.0 ng/mL consider an individualized risk assessment for PCa, particularly for high-grade cancer. It is especially noteworthy that on p. 95 of the 2013 ACS Guidelines on Cancer Screening, the following statement appears:

Prostate cancer screening should not occur without an informed decision-making process.

Therefore like the USPSTF, the ACS recommends against routine PSA-based screening for PCa without an informed decision-making process.

Finally, the American College of Physicians (ACP) makes the following recommendations about screening for PCa:

Guidance Statement 1: ACP recommends that clinicians inform men between the age of 50 and 69 years about the limited benefits and substantial harms of screening for prostate cancer. ACP recommends that clinicians base the decision to screen for prostate cancer using the prostate-specific antigen test on the risk for prostate cancer, a discussion of the benefits and harms of screening, the patient’s general health and life expectancy, and patient preferences. ACP recommends that clinicians should not screen for prostate cancer using the prostate-specific antigen test in patients who do not express a clear preference for screening. [emphasis added]
Guidance Statement 2: ACP recommends that clinicians should not screen for prostate cancer using the prostate-specific antigen test in average-risk men under the age of 50 years, men over the age of 69 years, or men with a life expectancy of less than 10 to 15 years.

Thus we see that like the USPSTF and the ACS, the ACP recommends against routine PSA-based screening for PCa without informed decision making.

European countries also differ in their recommended PSA screening policies. Based on the ERSPC, most European countries (that participated in the study) use the threshold of 3.0 ng/mL, although, Finland, Italy, Holland, and Belgium use the threshold of 4.0 ng/mL [10, 95]. Most European countries use 4-year screening intervals, although Sweden uses 2-year intervals.

2.1.2 Randomized Control Trials

There have been several RCTs to study prostate cancer screening [67, 64, 88, 10, 95, 59]. The two largest of these are the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial and the ERSPC, the former taking place in the United States. The ERSPC trial found that the rate of death from prostate cancer decreased by 20% as a result of screening, but the PLCO trial found that screening did not achieve a significant mortality reduction [10, 95]. The ERSPC generally used a PSA threshold of 3.0 ng/mL, whereas the PLCO generally used a threshold of 4.0 ng/mL. Thus, the difference in the observed outcomes may be partly due to differences in the screening strategies being evaluated, not to mention differences in prostate cancer incidence and mortality rates between the underlying American and European populations.

There were, however, significant criticisms and sources of bias in the ERSPC and PLCO trials, which are summarized in Schröder and Roobol [94]. For example, because the ERSPC results are based on multiple studies from eight testing centers (with a significantly higher reduction in mortality from screening reported by one of the centers), the results from common analysis across testing centers should not be afforded the weight usually given to clinical trials. Also, treatment was more aggressive in the screen arm of the trial, which may mean that the observed difference in prostate cancer mortality between the two arms is partly due to differences in treatment, not just in screening. For the PLCO trial, one of
the main criticisms was that among participants in the screen arm, those who were prescreened prior to randomization exhibited on average 25% lower prostate cancer mortality than those who were not prescreened. Furthermore, a substantial portion of participants assigned to the control arm actually continued to undergo screening, and subsequent studies have noted that this contamination is likely to weaken the perceived benefit of screening in the PLCO trial [42, 80].

The outcomes of the ERSPC and PLCO trials suggest it may not be possible to accurately evaluate prostate cancer screening strategies using RCTs. This motivates the importance of using quantitative models to investigate the effects of prostate cancer screening on a population.

2.1.3 Prostate Cancer Screening Models

Etzioni et al. [30] formulated an approach for estimating the duration of preclinical prostate cancer, which is prostate cancer that neither has been diagnosed nor has presented symptoms. They derived expressions for estimating the age-specific preclinical incidence rate and prevalence. They then formed the sum weighted by the fraction of a population in each consecutive 1-year age interval from age 30 to 90 of the following quantities: (i) the age-specific preclinical incidence rates; and (ii) the age-specific preclinical prevalences. Finally, they estimated the average preclinical duration as the ratio of the weighted sum of incidence rates to the weighted sum of prevalences. Using data from a Connecticut Surveillance, Epidemiology, and End Results (SEER) registry, US census data, and data from US life tables, they found that the average time spent with preclinical prostate cancer is 11 to 12 years for white males and 10 to 11 years for black males.

Ross et al. [86] used a Monte Carlo simulation model of prostate cancer defined on the states no prostate cancer, organ-confined prostate cancer at two different tumor volumes, and cancer not confined to the prostate. The simulated detection rate of prostate biopsy was made to be positively correlated with the size of an underlying tumor, and no patient was allowed more than 3 biopsies unless the patient had a PSA result greater than 10 ng/mL. PSA results were sampled from a bivariate log-normal distribution with mean and variance depending on the volume of an underlying prostate tumor. Using their model, they compared 7 existing screening strategies with the strategy of no screening, and for each strategy
predicted the number of prostate cancer deaths prevented, the number of person-years of life saved, the number of PSA tests, and the number prostate biopsies. They presented the relative trade-offs between the benefits and resource implications for each strategy, and found that screening at age 40 and 45 and then screening biennially from age 50 reduced prostate cancer mortality, as well as the number of PSA tests and biopsies when compared with the standard US strategy of annual screening from age 50.

A group of researchers from the Cancer Intervention and Surveillance Modeling Network of the National Cancer Institute, the Prostate Working Group, have studied, developed, and compared models of the natural progression of prostate cancer for over 10 years.\(^1\) Using results from the Rotterdam section of the ERSPC, Draisma et al. [25] modeled the development of prostate cancer up to the point of diagnosis as a Markov process with 9 states defined by all combinations of 3 different cancer stages and 3 different cancer grades. Screening for prostate cancer in the model consisted of events with sensitivity depending on the stage of an underlying cancer. Three variants of this basic model were considered. The first variant included the simulation of cancer that would be detectable as localized cancer by biopsy but would never have been otherwise diagnosed, since such cancers were suggested by the high detection rates among screened men in the ERSPC data. The second variant sampled the duration of preclinical cancer stages based on different probability distributions in order to produce higher variance in the length of time a patient spends in a given cancer stage. The third variant prevented the tumor grade from changing once it became detectable by screening. Using these model variants, they predicted lead times and rates of overdiagnosis for screening at different intervals, and found that lead times and overdiagnosis rates were more favorable for screening intervals larger than 1 year. They also produced evidence that screening may advance diagnosis by more than 10 years.

Etzioni et al. [31] used a Markov chain–based natural-history model of prostate cancer to examine the optimal frequency for normal-risk males to undergo PSA testing. Using rates of disease onset and transition rates between stages (defined by the Whitmore-Jewett staging system), the model generates the age of onset of prostate cancer and the duration in each cancer stage for a fixed cohort of US males. For each patient, the model uses reported lifetime probability of clinical presentation to determine randomly whether the patient will ever be clinically diagnosed. An initial PSA value is randomly generated for each

---

\(^1\)See http://cisnet.cancer.gov/prostate/
patient, and each subsequent year the patient’s PSA value increases by a percentage (plus a small random
error) that depends on the patient’s cancer stage and stage duration. Using their model, they consider 5
different screening strategies on the basis of benefits harms, specifically, the number of life-years saved
from screening and the number of false positive PSA tests and rate of overdiagnosis. They found that
screening every 2 years was a cost-effective alternative to the standard US strategy of annual screening.

Etzioni et al. [32] used a simulation model to investigate whether population PSA screening could be
responsible for the decline in prostate cancer mortality observed from 1992 through 1994. The simulation
model identified patients to be PSA-tested and PSA-testing–diagnosed based on a PSA utilization
database, recorded PSA testing rates, and linked SEER–Medicare data. Based on a range of estimates
for how often PSA screening is used for diagnostic purposes and how often it is used for screening,
the model then selected a portion of the PSA-testing–diagnosed patients to have been diagnosed early,
i.e., diagnosed at a time when the cancer is amenable to definitive treatment. Using an estimate of the
decreased risk of death from prostate cancer by screening, and an estimate of the time by which diagnosis
is advanced by screening (lead time), the simulation model generates a date of death for the patient
with screening and a date of death for the patient without screening. Using the decreased risk of dying
from prostate cancer by screening, they computed the rate of death from prostate cancer prevented by
screening. This rate was added to the observed prostate cancer mortality rate to produce a mortality rate
reflecting the absence of screening. Using a rate of screening efficacy matching the rate hypothesized for
the PLCO trial, they computed an adjusted prostate cancer mortality rate for the period from 1988 to
1994, and found that only very short lead times would result in PSA screening being potentially fully
responsible for the observed decline in prostate cancer mortality.

Using the simulation model from Etzioni et al. [32], Etzioni et al. [33] sought to determine the
level of prostate cancer overdiagnosis that would likely result in the increased incidence observed
from 1988 to 1999. They estimated as the overdiagnosis rate the proportion of simulated patients with
prostate cancer detected by screening who died before the cancer became clinically detectable. For
different lead-time values, they compared model-projected incidence trends for blacks and whites against
 corresponding SEER-based incidence trends, finding that the best-fitting lead times for blacks and whites
were, respectively, 7 and 5 years, with corresponding overdiagnosis rates 44% and 29%, respectively. After adjusting for the expected lifetime probability of latent undiagnosed prostate cancer during the era before the advent of PSA screening, the overdiagnosis rates for blacks and whites during that era were estimated to be at most 37% and 15%, respectively.

In order to assess whether PSA screening was responsible for the decline in prostate cancer–specific mortality since the 1990s (a research objective similar to that of Etzioni et al. [32]), Etzioni et al. [35] adapted the model from Etzioni et al. [31] to simulate life histories for the US population corresponding to several different birth cohorts between 1980 and 2000. The model generates a set of natural histories and clinical case histories, and then pairs the clinical histories that include diagnosis with natural histories based on birth cohort year, age of disease onset, and age of diagnosis. For clinical histories where no diagnosis occurs, clinical histories are paired with the remaining unmatched natural histories so that age at other-cause death in the clinical history precedes the age of distant-stage incidence in the natural history. For each birth cohort, PSA screening histories are simulated based on reported data on interscreening intervals. Simulated patients with a PSA value greater than 4 ng/mL underwent prostate biopsy with a probability based on data from the PLCO trial. Results from the model suggested that 80% of the decrease in incidence of advanced-stage prostate cancer since 1990 was attributable to prostate cancer screening.

Etzioni et al. [36] compared projections of the extent to which PSA screening is responsible for the decline in prostate cancer mortality since the 1990s from a model by the Fred Hutchinson Cancer Research Center (FHCRC) with similar projections from a model by the University of Michigan (UMICH). Using autopsy data to derive incidence of latent tumor onset, estimates of transition rates between pathological states, and a database of PSA growth trajectories from men diagnosed with prostate-cancer, the FHCRC model parameterized the models from Etzioni et al. [31] and Etzioni et al. [35] so that age- and stage-specific incidence and mortality rates matched those before and after the advent of PSA screening. The UMICH model estimated population-level age-at-onset and sojourn-time distributions by modeling 3 components [106]. First, they modeled a hazard function of tumor onset as a function of age and birth cohort. Second, they modeled time between tumor onset and clinical diagnosis as a sojourn time hazard.
function incorporating a multiplicative trend function to account for changes in the general practice of prostate cancer screening. Third, they modeled PSA screening using a PSA simulator based on linked SEER–Medicare data and results from the National Health Interview Survey. Using PSA screening trends, the model estimates incidence in the presence of PSA screening as well as the distribution of lead times. By 2000, the FHCRC and UMICH models project that PSA screening may explain, respectively, 45% and 70% of the decline in prostate cancer mortality.

Gulati et al. [41] studied the risk of clinical progression and prostate cancer–specific death after detection of prostate cancer in the absence of definitive treatment using 3 models discussed above: the FHCRC [32], MISCAN [25], and UMICH [106] models. For all 3 models, incidence rates were derived from data in 9 SEER registries covering the period from 1975 to 2000. Disease progression levels were defined as Gleason Scores in the ranges 2–7 and 8–10. Other-cause mortality rates were derived from US life tables. And a common Poisson regression model was used to model prostate cancer survival [2]. The models projected that without early detection or definitive treatment, 12%–25% of US males with preclinical onset would have died due to prostate cancer. If left without definitive treatment, prostate cancers diagnosed by positive biopsy with Gleason Score 2–7 and 8–10 would later cause prostate cancer death with probabilities 0.23–0.34 and 0.63–0.83, respectively.

Building on previous previous models [25, 35], Gulati et al. [43] used a microsimulation model comprising prostate cancer incidence and mortality components calibrated on data from the ERSPC to compare the effectiveness of different screening strategies on the basis of prostate cancer mortality, overdiagnosis, and the number of false-positive PSA tests. The incidence component consists of a PSA growth component based on data from the Prostate Cancer Prevention Trial and a disease progression component that generates, based on simulated PSA level, the ages at disease onset, metastasis, and diagnosis. The mortality component generates other-cause survival and disease-specific survival based on the age of the patient, and the stage and grade of the tumor at diagnosis. They considered 32 alternative prostate cancer screening strategies and projected a range of positive and negative outcomes, finding that PSA thresholds increasing with age, criteria for biopsy referral becoming more conservative with age, and a low frequency of screening for men with low PSA levels can reduce prostate cancer mortality while
also reducing harms associated with overdiagnosis.

2.2 Simulation Optimization

Simulation optimization is any instance of optimization where the objective function is evaluated by simulating a stochastic system; the simulator is effectively a black box that returns a value often with no clear analytic expression of the function. Stochastic problems which lack properties such as convexity and optimal substructure are not amenable to optimization methods such as stochastic programming or stochastic dynamic programming, but are often amenable to simulation optimization. Fu [38] reviews a host of optimization techniques, including tabu search, simulated annealing, and genetic algorithms, which are suited to optimize problems in which the objective function must be approximated by simulation, and discusses the challenges presented by, and the growing interest in, simulation optimization. Other excellent reviews of simulation optimization are [56] and [102]. Andradóttir [8] provides an excellent review of simulation-optimization based on random search methods.

Simulation optimization problems are often solved by implementing a GA, one of the approaches taken in this thesis, such as described in [27] and [24]. GAs are conceptually rooted in the evolutionary laws of natural selection, and are members of the larger class of evolutionary algorithms. Like natural selection, GAs improve a system over time by repeatedly modifying a population of candidate solutions in such a way that better candidate solutions are promoted over time. In 1975, Holland [53] proposed this natural concept as an algorithmic tool for systems modeling and analysis.

GAs have been widely applied across diverse disciplines. Cryptanalysts have used GAs with an initial population comprising guessed keys to attempt to decrypt intercepted cyphertext [75]. Astronomers have used GAs to fit models of the rotation curve of a distant galaxy [18]. Electrical engineers have used GAs to aid in the design of digital signal processing and analysis filters [45]. The common thread that unites these and many other applications, is the use of GAs to overcome the computational intractability of certain problems.

Hong and Nelson [55] introduced a framework for solving discrete simulation-optimization problems called COMPASS, which stands for “convergent optimization via most-promising-area stochastic search.”
COMPASS can be considered a framework because key components of the framework can be significantly modified without violating several mild assumptions that are sufficient conditions for COMPASS to asymptotically converge to a locally optimal solution.

COMPASS is a sampling-based stochastic-search optimization method. The key feature of COMPASS is the “most-promising area” (MPA), which is defined to be the set of feasible solutions that are at least as close to the current sample-best solution as they are to other visited solutions [55]. Thus, the MPA is fully defined given the set of feasible solutions, the set of solutions that have previously been explored (visited solutions), and the current sample-best solution. At any iteration, COMPASS conducts stochastic search confined to the feasible solutions in the MPA.

An implementation of COMPASS begins with an initial solution (or multiple initial solutions). The expectation of the stochastic objective function is estimated for these solutions (usually by simulation), and the current sample-best is identified by comparing the sample-mean objective-function values of all the solutions. Given the current sample-best solution and the other initial solutions (if there are multiple initial solutions), the MPA is updated. If there is only 1 initial solution, which of course must be the current sample-best solution, then the MPA will be the entire feasible region; otherwise, the MPA will be a subset of the feasible region containing solutions that are at least as close to the current sample-best solution as they are to the other initial solutions.

On subsequent iterations of an implementation of COMPASS, solutions are randomly sampled within the MPA, and all visited solutions — the set of visited solutions includes the sample-best solution and the newly sampled solutions — are evaluated based on some defined simulation-allocation rule (SAR). Then, the current sample-best solution is updated by comparing the sample-mean objective-function values of all visited solutions, which may or may not result in the current sample-best solution changing. Once the sample-best solution is updated, then the MPA is updated, and COMPASS proceeds to the next iteration unless some defined stopping criterion is satisfied.

As COMPASS proceeds, the MPA generally becomes smaller and moves toward the direction of a locally optimal solution. Importantly, COMPASS never discards solutions or simulation-observations. Thus, at any iteration of COMPASS, when the entire set of visited solutions is evaluated based on the
SAR, it is necessary only to simulate the number of additional replications (observations) of each solution in order to satisfy the total number of replications required of each solution by the SAR. The accumulation of simulation observations over the iterations of COMPASS reduces the overall computational costs of the method.

Figure 2.1 illustrates four iterations of COMPASS. In Figure 2.1, each of the black dots represents a feasible solution to a hypothetical two-dimensional discrete solution-space. In step I, the MPA (denoted by the dashed border) is the entire set of feasible solutions, and 3 initial solutions (white dots with black borders) are chosen. Because we have sampled or explored these solutions, they are considered visited solutions. In step II, the sample-best (denoted by the star) is selected from the current set of all visited solutions, and the MPA is updated.

The MPA is updated by constructing hyperplanes separating the sample-best solution from the other visited solutions. A hyperplane is constructed between the sample-best solution and each other visited solution. Each hyperplane separates feasible solutions that are at least as close (in Euclidean norm) to the sample-best solution as to the given visited solution from all other feasible solutions that are closer to the given visited solution than to the sample-best solution. In the two-dimensional case, each hyperplane would be the perpendicular bisector of the line segment connecting the sample-best solution to the given non-sample-best visited solution. Note that the MPA can be defined by separating hyperplanes that, together with the implied hyperplanes at the boundaries of the finite feasible region, form a convex polyhedron. In particular, to form the MPA polyhedron, one constructs for each non-sample-best visited solution the hyperplane normal to the line containing that non-sample-best visited solution and the sample-best solution. In the two-dimensional problem in Figure 2.1, the separating hyperplanes are lines perpendicular to lines containing a visited solution and the sample-best solution.

Now, in step III, 3 new solutions are sampled from the updated MPA, and the sample-best solution changes. Finally, in step IV, the MPA is updated based on the entire set of visited solutions and the new sample-best solution discovered in step III.

Hong et al. [57] examined the convergence properties of COMPASS and found that the required running time for COMPASS to locally converge deteriorates rapidly as the number of decision variables
Figure 2.1: Illustrating how solutions are sampled and the most promising area is defined in our COMPASS implementation.
grows larger than 10. They proposed a new scheme for sampling solutions from the MPA called *coordinate sampling*; in coordinate sampling, new solutions are sampled by simply shifting the sample-best solution in one coordinate direction instead of sampling solutions uniformly randomly from the MPA. They found that coordinate sampling reduced the computational overhead of COMPASS by as much as 1000 times.

Recently, Xu et al. [111] developed an local-search alternative to COMPASS called the adaptive hyperbox algorithm (AHA). There are many similarities between COMPASS and AHA; indeed, AHA was inspired by COMPASS. Unlike COMPASS, however, the MPA in AHA is always defined by a \(d\)-dimensional hyperbox, where \(d\) is the number of decision variables. In COMPASS, the MPA is constructed by hyperplanes separating the sample-best solution from each other solution in the visited set. Consequently, when the number of visited solutions grows very large, then construction of the MPA, which occurs at every iteration of COMPASS, can become time-prohibitive. In AHA, the MPA is constructed or modified at each iteration by simple shifting each facet of the hyperbox in one coordinate direction. For problems of many decision variables, AHA is shown to perform as well as COMPASS.

Xu et al. [110] recommend that the optimal computing budget allocation (OCBA) procedure be used as the SAR in COMPASS. Given a set of solutions, OCBA determines the number of simulation replications required for each solution \(x_i\) in order to maximize the probability that the solution with highest sample-mean performance is truly the solution with highest expected performance. Let \(v(x_i)\) denote the expected performance of solution \(x_i\), let \(\overline{v}(x_i)\) denote the sample-mean estimator of \(v(x_i)\) obtained by \(m_i\) independent replications of the simulation at solution \(x_i\), let \(\sigma_i^2\) denote the sample variance of the \(m_i\) simulation-generated responses at the solution \(x_i\), let \(b\) denote the index of the solution with highest sample-mean performance, and let \(k\) denote the number of solutions being compared. Then the probability of correct selection for a maximization problem in which we take solution \(x_b\) to be the best solution is defined as

\[
P\{\text{CS}\} = \Pr\{v(x_b) > v(x_i), i \neq b\}. \tag{2.1}
\]
From Chen [21] we know that a lower bound on $P\{CS\}$ can be computed by

$$P_{\text{CS}}^{\text{LB}} = \prod_{i=1, i \neq b}^{k} \Pr \{ \tilde{v}(x_b) > \tilde{v}(x_i) \}. \quad (2.2)$$

Because Equation (2.2) is a lower bound on $P\{CS\}$ that can be quickly computed from the sample-mean performance estimates for solutions $x_1, \ldots, x_k$, OCBA uses this approximation to estimate the probability of correct selection.

OCBA utilizes the concept of signal-to-noise ratio. The “signal” in OCBA for a given solution $x_i$ is the difference $\delta_{b,i} = \tilde{v}(x_i) - \tilde{v}(x_b)$ between the sample mean $\tilde{v}(x_i)$ of that solution and the sample mean $\tilde{v}(x_b)$ of the current sample-best solution. The “noise” in OCBA for a given solution is that solution’s sample variance $\sigma_i^2$. The fundamental principle of OCBA is to allocate simulation replications to the various solutions in proportion to their signal-to-noise ratios. If a solution has a relatively high signal-to-noise ratio, we know with a high probability that this solution is not actually the best of the solutions under consideration. On the other hand, if a solution has a relatively low signal-to-noise ratio, then we know with a lower probability that this solution is not actually better than the current sample-best solution.

From Chen and Lee [22], we know that given a computing budget of $B$ simulation replications, we can maximize the approximate probability of correct selection asymptotically as $B \rightarrow \infty$ by solving the equations

$$m_i = \left( \frac{\sigma_i}{\delta_{b,i}} \right)^2, \quad i, j \in \{1, 2, \ldots, k\}, \text{ and } i \neq j \neq b, \quad (2.3)$$

$$m_b = \sigma_b \times \sqrt{\sum_{i=1, i \neq b}^{k} \frac{m_i^2}{\sigma_i^2}}, \quad (2.4)$$

$$\sum_{i=1}^{k} m_i = B. \quad (2.5)$$

In OCBA, the simulation budget, $B$, is initially small and is gradually increased over a series of iterations. The specific values for $B$ initially and after each iteration are problem specific and can be best
chosen by experimentation for a given problem. Every time $B$ is increased, Equations (2.3), (2.4), and (2.5) are solved and each solution $x_i$ is evaluated by executing $m_i - m'_i$ additional simulation replications. This iterative process allows OCBA to gradually learn more about the true mean and variance of each solution, and, at each iteration after the first, correct poorer mean and variance estimates from the preceding iteration.

2.3 Contributions of this Dissertation

This dissertation contributes to the existing literature in the following ways. First, our research constitutes to the best of our knowledge the first simulation-optimization study of PCa screening across ethnic, hereditary, and comorbidity risk groups. Although randomized clinical trials have studied the effect of prostate cancer screening in different populations, they are costly, time-consuming, and lack the freedom to explore more than a handful of alternative screening strategies. Randomized clinical trials also suffer from contamination and other sources of bias, as described previously in this chapter. Our research, which uses simulation rather than real trial patients, is unencumbered by the financial and ethical constraints that preclude consideration of a vast array of untried and unexplored possible screening strategies. This research can be extended by modeling additional risk groups, such as subgroups within the groups we considered. Additionally, this model could be modified to consider additional non-PSA-based PCa biomarkers.

Second, our research stands favorably among many other PCa-screening studies in the way that we have rigorously, precisely, and accurately computed the effect of PCa screening on expected quality-adjusted survival. Many other similar studies have relied on Markovian problem formulations—such as Partially Observable Markov Decision Processes (POMDPs)—or simulations of patients at higher (less granular) levels. And many studies do not expressly estimate quality-adjusted survival, but rather work to minimize, maximize, or trade off other PCa-outcome–related measures. Our research makes a number of modeling and methodological contributions. Our model optimization model abandons the “memoryless” Markovian property, which is a restriction on previous approaches such as POMDPs. While this makes for a challenging optimization problem (i.e., requiring the use of stochastic simulation
optimization), it also allows us to model a range of important aspects of the problem that would otherwise be extremely challenging, if not practically infeasible. For example, we model clinical detection of PCa on the individual patient level by monitoring the time that a patient spends with PCa, and we use this time in conjunction with empirical PCa-leadtime estimates to determine when (if ever) a given patient will present symptoms of PCa that lead to clinical detection. To model the same monitoring process in a Markov model would require a large state-space expansion, and the transition probabilities between states in the expanded space might be very difficult to estimate accurately. Another example of the advantage of relaxing the Markov assumption is our use of a widely accepted method for sampling patient PSA levels based on a linear changepoint model of PSA growth [40]. In this method, each patient’s PSA trajectory over time is based on the age at PCa onset (if onset occurs for the given patient) and several randomly-sampled, continuous, patient-specific, PSA-model parameters. Such a PSA model does not have a tractable Markov representation, because each patient’s future PSA trajectory depends not merely on the patient’s present PSA level but also on previous values of the PSA level and on the time of PCa onset (if applicable). These two examples demonstrate the value in using a simulation-optimization approach as opposed to a dynamic-programming approach because both clinical detection and PSA growth over time are important components of the PCa screening problem.

Third, we have carefully validated important aspects of our simulation model such as life expectancy, PCa diagnosis and mortality, and metastasis survival. We validated the novel PCa leadtime–based implementation of clinical detection together with other components of the simulation model by using simulation to reconstruct the ERSPC trial and compare the corresponding simulation NNS estimates with the published reports on the outcomes of the ERSPC.

Fourth, we demonstrate how careful synchronization of simulated patient histories can be used as a variance-reduction technique to evaluate and compare screening strategies on the basis of QALYs. The NNSQ statistic is easily interpreted by a clinical audience because of its conceptual similarity to the widely-used NNS statistic. Therefore, we contend that NNSQ is an excellent performance measure for a comparative simulation study of disease screening or treatment strategies.

Fifth, we develop two independent novel methods for producing screening strategies with regularity
in the time series of PSA test thresholds. Because of the tendency of simulation-optimization methods
to produce strategies with erratic patterns of PSA test thresholds, and because such irregular strategies
would be difficult to implement in practice and easy to misinterpret, these methods may prove valuable
to future simulation-optimization research in disease screening and treatment. This research could be
extended by investigating additional approaches for inducing regularity in the decision-variable time
series. For example, the binary-encoding scheme could be modified for the PCa screening problem by
assuming a nonlinear rate of change in the PSA threshold over time.

Finally, we developed an analytical formulation of a simplified single-period screening problem that
can be helpful in trying to understand PCa screening. We derived a closed-form solution to this analytical
optimization problem as well as point and CI estimators for the objective function. Results from this
analytical model were used to support results from the simulation-optimization methods, implying that
optimal screening may be characterized by biopsying virtually all men. And with careful sensitivity
analysis, we established that this conclusion was robust to most plausible variation in estimates of key
model parameters.
Chapter 3

Simulation of PSA-Based Prostate Cancer Screening for Heterogeneous Patient Risk Groups

3.1 Introduction

PCa is the most common noncutaneous malignancy and the second leading cause of cancer mortality in US men [91]; and screening for PCa is controversial [79]. Several factors have been associated with the risks of PCa incidence and mortality. For example, African American males have higher PCa incidence and mortality than white US males [7, 103]; and men with a family history of PCa have a higher risk of diagnosis than those without a family history [100]. On the other hand, men with competing risks are less likely to experience morbidity and mortality from PCa [3]; and thus the burden of comorbidities may contextually define a lower PCa risk group. Screening for PCa with prostate specific antigen (PSA) has been broadly adopted in the United States over the past two decades [28]; however, this practice has been a subject of great controversy, as discussed in Chapter 2. In 2012, the US Preventive Services Task Force (USPSTF) recommended against PSA screening in average-risk men [78]. However, groups such as the American Urological Association (AUA) and the American Cancer Society (ACS) continue to
recommend screening in the context of a shared decision-making process for some patients that considers individual patient preferences, life expectancy, and risk factors such as race and family history of PCa [16, 98].

Differences in incidence and mortality rates, as well as variation in the predictive value of PSA, can affect the balance of harms and benefits of screening within demographic subpopulations of US males [85, 105]. Estimating the public-health implications of screening in these various subgroups is further complicated by both the large number of proposed screening strategies and the difficulty of estimating common performance measures for screening such as the expected number needed to screen (NNS) to avert 1 PCa death [10, 95].

We developed and used a computer simulation of PCa screening, derived from a stochastic-process model of the natural history of the disease, to compare screening strategies for five demographic groups based on race (white and African American), family history of prostate cancer, and different levels of comorbid medical conditions. Using estimates of the incidence, other-cause mortality, and PCa mortality rates associated with each risk group, we compared published PSA screening strategies to find the strategy that minimized the expected NNS for each subgroup.

3.2 Methodology

A stochastic-process model of the natural history of PCa from Underwood et al. [107] was adapted and combined with a method for sampling patient PSA histories and a method for incorporating clinical incidence of PCa to predict the effects of published screening strategies on different risk groups. We developed a method of synchronizing patient histories between screening and no-screening simulations in order to estimate expected NNS. Different risk groups were modeled by varying relevant model parameters based on ratios reported in the literature.

There are a number of important differences between the model in Underwood et al. [107] and the model in this research. The annual PCa onset probability is here based on estimates from much more recent autopsy data. The annual other-cause mortality rate is here calculated from more recent population statistics. The annual probability that PCa metastasizes after being definitely treated is here estimated
from a multi-center Scandinavian study instead of a largely regionally-local set of patients under care at Mayo Clinic. The annual probability that untreated PCa metastasizes is now estimated according to length of time since PCa onset (with a metastasis probability of 0 during a specific “preclinical dwelling interval”), whereas in Underwood et al. [107] a single fixed probability was used independent of how long a given patient had PCa. The annual probability of PCa-specific death is based on SEER data both in Underwood et al. [107] and in this research, but the estimates in this research are based on a more-recently published SEER Survival Monograph.

In addition to improved estimates of various model transition probabilities, the model in this research surpasses the model in Underwood et al. [107] by the use of PCa leadtime in the simulation of clinical (nonscreen) detection of PCa. Accounting for clinical incidence of PCa is important because it influences the overall absolute benefit that can be realized by screening.

In the model in this research, we also have a more detailed set of QALY disutilities, and the disutility estimates come from a prominent recent study of quality-of-life and PCa screening. Previously, in Underwood et al. [107], the estimates for the various QALY disutilities were pooled from more than a single source. Furthermore, previously the disutility estimate for biopsy was based on a disutility estimate from another similar biopsy procedure; in this research, we have a much better estimate for the disutility of biopsy, and it is significantly smaller than the previous estimate inferred from a nonprostate biopsy procedure. Finally, in much of this research we permit men to undergo more than 1 prostate biopsy during their lives. Previously, we restricted all patients to at most 1 biopsy.

3.2.1 Patient Health-State Stochastic Process

The health states and progression paths are depicted in Figure 3.1. In our model, all patients diagnosed with PCa are treated by radical prostatectomy. All simulated patients are initially disease free but may later develop PCa according to the “PCa onset” transition depicted in Figure 3.1.
Figure 3.1: Health states and progression paths in the simulation model. Transitions between states are represented by arrows.
Underlying the simulation model are one-step transition probability matrices defined on a finite set of states. The states in this finite set are the following: no cancer (NC), undiagnosed cancer (C), posttreatment (T), metastasis (M), and death (D). State D is the only absorbing state. These states correspond to the states in Figure 3.1 in the following way: NC corresponds to “No PCa”; C corresponds to “PCa (undetected)”; T corresponds to “Posttreatment for PCa”; M corresponds to “Metastasis from PCa”; and D is an absorbing all-cause mortality state that encompasses both “All-other-cause mortality” and “PCa specific mortality.” Statistics are maintained separately for PCa-specific and non-PCa-specific mortality. For more details on these health states, see Underwood et al. [107].

Not all PCa tumors are considered to be equally threatening. There are low and high risk tumors represented by different grades, as based on the Gleason scoring system [76]. The grade of prostate tumor is an important predictor of survival. Although our model does not explicitly consider tumor grade by differentiating PCa with different states having different corresponding transition probabilities, our estimates of quality-adjusted survival are based on an average over different possible grades at diagnosis by biopsy. This aggregation is motivated by a lack of adequate data for different tumor grades.

We formulate a finite-horizon problem consisting of a set of $N + 1$ time periods,

$$\mathcal{T} \equiv \{0, 1, \ldots, N\},$$

each of which correspond to an age in a patient’s life. Time period $t$ corresponds to a patient’s life over the age interval $[t, t+1)$. The state set,

$$\mathcal{S} \equiv \{\text{NC, C, T, M, D}\},$$

comprises the patient health states in the model. In every time period in this process, there are two stages of decision making. In the first stage, it is determined whether the patient should undergo a PSA test, and, if so, what should be the corresponding PSA threshold that will trigger a biopsy. In the second stage, it is determined whether the patient should undergo a prostate biopsy based on the result of the PSA test and the selected PSA threshold. The decision-making sequence is illustrated in Figure 3.2.
Figure 3.2: Sequence of two decision-making stages that take place during each period in the simulation.

The decision (action) set,

\[ \mathcal{A}_1 \equiv \{\text{No PSA test, PSA test}\}, \]

comprises the possible decisions in the first decision-making stage; and the decision set,

\[ \mathcal{A}_2 \equiv \{\text{No biopsy, Biopsy}\}, \]

comprises the possible decisions in the second decision-making stage. We denote by the double

\[ a_t \equiv (a_{1,t}, a_{2,t}) \]

the decisions taken in the first and second decision-making stages in period \( t \) for \( 1 \leq t \leq N - 1 \). We denote by

\[ \mathcal{A} \equiv \{(a_{1,t}, a_{2,t}) : a_{1,t} \in \mathcal{A}_1, a_{2,t} \in \mathcal{A}_2\} \]

the set of possible decision doubles. For \( t \in T \) and \( 1 \leq t \leq N - 1 \), the random variable \( s_t \in \mathcal{S} \) denotes the patient’s health state that prevailed over the time period \( (t - 1, t] \) so that the patient makes a transition to state \( s_{t+1} \) just after time \( t \), and state \( s_{t+1} \) prevails over the time period \( (t, t + 1] \). The state \( s_0 \) represents
the initial state of a patient upon beginning simulation, and we assume \( s_0 = \text{NC} \) for all patients.

The random variable \( \rho_t \) has outcomes that are the results of a PSA test of a patient at epoch \( t \) for \( 1 \leq t \leq N - 1 \), where the outcomes are denoted by the set

\[
P \equiv \{\text{No PSA result} \} \cup \mathbb{R}^+.\]

The random variable \( \eta_t \) has outcomes that are the results of a prostate biopsy at epoch \( t \) for \( 1 \leq t \leq N - 1 \), where the outcomes are denoted by the set

\[
E \equiv \{\text{No biopsy result, Negative, Positive}\}.
\]

Clearly, then, the random variables \( \rho_t \) and \( \eta_t \) have the following respective conditional probabilities:

\[
\Pr\{\rho_t = \text{No PSA result} \mid a_{1,t} = \text{No PSA test}\} = 1,
\]

and

\[
\Pr\{\eta_t = \text{No biopsy result} \mid a_{2,t} = \text{No biopsy} \text{ or } a_{1,t} = \text{No PSA test}\} = 1.
\]

Because decisions at an epoch \( t \) may depend on the history of decisions \( a_k \in \mathcal{A} \) at epochs \( k = 1, \ldots, t-1 \), we let

\[
h_t = (a_1, a_2, \ldots, a_{t-1}) \quad \text{for} \quad t = 1, \ldots, N - 1
\]

denote the history of decisions taken for a given patient prior to epoch \( t \).

The transitions between states occur according to a one-step transition probability matrix that depends
on the current time $t$ and has the following form:

$$P(t) = \begin{pmatrix}
NC & C & T & M & D \\
NC & P_{1,1}(t) & P_{1,2}(t) & 0 & 0 & P_{1,5}(t) \\
C & 0 & P_{2,2}(t) & P_{2,3}(t) & P_{2,4}(t) & P_{2,5}(t) \\
T & 0 & 0 & P_{3,3}(t) & P_{3,4}(t) & P_{3,5}(t) \\
M & 0 & 0 & 0 & P_{4,4}(t) & P_{4,5}(t) \\
D & 0 & 0 & 0 & 0 & 1
\end{pmatrix}.$$

The various parameters used to define the individual transition probabilities are defined in Table 3.1.

<table>
<thead>
<tr>
<th>Param.</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_t$</td>
<td>Annual probability of developing biopsy-detectable PCa</td>
</tr>
<tr>
<td>$d_t$</td>
<td>Annual probability of non–PCa-specific (other cause) death</td>
</tr>
<tr>
<td>$b_t$</td>
<td>Annual probability of PCa metastasizing for patients in state $T$</td>
</tr>
<tr>
<td>$e_t$</td>
<td>Annual probability of PCa metastasizing for patients in state $C$</td>
</tr>
<tr>
<td>$z_t$</td>
<td>Annual probability of PCa-specific death for patients in state $M$</td>
</tr>
<tr>
<td>$f$</td>
<td>Sensitivity of the prostate biopsy procedure</td>
</tr>
</tbody>
</table>

The transition probability matrix is a function of whether a prostate biopsy is performed. That is, at the beginning of the current time period $(t, t + 1]$ in the model (that is, just after time $t$), a state transition from state $s_t$ to state $s_{t+1}$ occurs according to matrix $P^B(t)$ if a prostate biopsy was performed at the beginning of the current period; otherwise, the state transition occurs according to matrix $P^{NB}(t)$.

It is important to note that the one-step transition matrices are patient-history dependent, and thus the stochastic process governed by the transition matrices is not Markovian in the usual sense. For example, as explained below, the value of transition-probability parameters $b_t$ and $e_t$ for a given patient at a given time $t$ are dependent upon the dwelling time of the patient in state $C$. And this is an example of the benefit obtained by not requiring the Markovian property: it would be difficult to obtain accurate estimates of the transition probability parameters for a Markov version of this stochastic process with states defined by
Given the decision to biopsy (B), the one-step transition probability matrix has the following probabilities:

\[ P_{1,1}^B(t) = \Pr\{s_{t+1} = \text{NC} \mid s_t = \text{NC}, a_{2,t} = B\} = (1 - d_t)(1 - w_t). \]

\[ P_{1,2}^B(t) = \Pr\{s_{t+1} = C \mid s_t = \text{NC}, a_{2,t} = B\} = (1 - d_t)w_t. \]

\[ P_{1,5}^B(t) = \Pr\{s_{t+1} = D \mid s_t = \text{NC}, a_{2,t} = B\} = d_t. \]

\[ P_{2,2}^B(t) = \Pr\{s_{t+1} = C \mid s_t = C, a_{2,t} = B\} = (1 - f)(1 - d_t)(1 - e_t). \]

\[ P_{2,3}^B(t) = \Pr\{s_{t+1} = T \mid s_t = C, a_{2,t} = B\} = f(1 - b_t)(1 - d_t). \]

\[ P_{2,4}^B(t) = \Pr\{s_{t+1} = M \mid s_t = C, a_{2,t} = B\} = f b_t(1 - d_t) + e_t(1 - f)(1 - d_t). \]

\[ P_{2,5}^B(t) = \Pr\{s_{t+1} = D \mid s_t = C, a_{2,t} = B\} = d_t. \]

\[ P_{3,3}^B(t) = \Pr\{s_{t+1} = T \mid s_t = T, a_{2,t} = B\} = (1 - d_t)(1 - b_t). \]

\[ P_{3,4}^B(t) = \Pr\{s_{t+1} = M \mid s_t = T, a_{2,t} = B\} = b_t(1 - d_t). \]

\[ P_{3,5}^B(t) = \Pr\{s_{t+1} = D \mid s_t = T, a_{2,t} = B\} = d_t. \]

\[ P_{4,4}^B(t) = \Pr\{s_{t+1} = M \mid s_t = M, a_{2,t} = B\} = (1 - d_t)(1 - z_t). \]

\[ P_{4,5}^B(t) = \Pr\{s_{t+1} = D \mid s_t = M, a_{2,t} = B\} = d_t + z_t(1 - d_t). \]

If no biopsy is performed (NB), the one-step transition probability matrix has the following probabilities:
For each of the compared risk groups, the $P^B(t)$ and $P^{NB}(t)$ matrices are computed as functions of the parameters listed in Table 3.1. The estimates of parameters $d_t$, $w_t$, and $z_t$ vary by risk group. We describe the risk groups in more detail in Section 3.2.2. In the remainder of this subsection, we describe important model assumptions related to the transition probabilities.

**Transition probabilities $P^B_{1,1}(t)$ and $P^{NB}_{1,1}(t)$**. A patient remains in state NC assuming that the independent events of developing PCa and dying of other (non-PCa) causes do not occur. Biopsying has no effect since, in state NC, there is no PCa to be detected.

**Transition probabilities $P^B_{1,2}(t)$ and $P^{NB}_{1,2}(t)$**. A patient transitions from state NC to C assuming that the independent events of developing PCa and not dying of other causes occur. Biopsying has no effect.

\[
P_{1,1}^{NB}(t) = \Pr\{s_{t+1} = \text{NC} \mid s_t = \text{NC}, \quad a_{2,t} = \text{NB}\} = (1 - d_t)(1 - w_t).
\]

\[
P_{1,2}^{NB}(t) = \Pr\{s_{t+1} = C \mid s_t = \text{NC}, \quad a_{2,t} = \text{NB}\} = (1 - d_t)w_t.
\]

\[
P_{1,5}^{NB}(t) = \Pr\{s_{t+1} = D \mid s_t = \text{NC}, \quad a_{2,t} = \text{NB}\} = d_t.
\]

\[
P_{2,2}^{NB}(t) = \Pr\{s_{t+1} = C \mid s_t = C, \quad a_{2,t} = \text{NB}\} = (1 - d_t)(1 - e_t).
\]

\[
P_{2,3}^{NB}(t) = \Pr\{s_{t+1} = T \mid s_t = C, \quad a_{2,t} = \text{NB}\} = 0.
\]

\[
P_{2,4}^{NB}(t) = \Pr\{s_{t+1} = M \mid s_t = C, \quad a_{2,t} = \text{NB}\} = e_t(1 - d_t).
\]

\[
P_{2,5}^{NB}(t) = \Pr\{s_{t+1} = D \mid s_t = C, \quad a_{2,t} = \text{NB}\} = d_t.
\]

\[
P_{3,3}^{NB}(t) = \Pr\{s_{t+1} = T \mid s_t = T, \quad a_{2,t} = \text{NB}\} = (1 - d_t)(1 - b_t).
\]

\[
P_{3,4}^{NB}(t) = \Pr\{s_{t+1} = M \mid s_t = T, \quad a_{2,t} = \text{NB}\} = b_t(1 - d_t).
\]

\[
P_{3,5}^{NB}(t) = \Pr\{s_{t+1} = D \mid s_t = T, \quad a_{2,t} = \text{NB}\} = d_t.
\]

\[
P_{4,4}^{NB}(t) = \Pr\{s_{t+1} = M \mid s_t = M, \quad a_{2,t} = \text{NB}\} = (1 - d_t)(1 - z_t).
\]

\[
P_{4,5}^{NB}(t) = \Pr\{s_{t+1} = D \mid s_t = M, \quad a_{2,t} = \text{NB}\} = d_t + z_t(1 - d_t).
\]
since, in state NC, there is no PCa to be detected. We assume that screening, biopsying, and treating occur at the first instant of a time period, immediately prior to making the state transition.

**Transition probabilities** $P^B_{i,5}(t)$ and $P^\text{NB}_{i,5}(t)$ for $i = 1, 2, 3$. Because there is no metastatic PCa in state NC, C, or T, death can only be other-cause mortality. In our model, we assume that only metastasized PCa is capable of immediately causing PCa-specific mortality. Biopsying has no effect since biopsy-triggered treatment is only capable of averting PCa-specific mortality, not other-cause death.

**Transition probabilities** $P^\text{NB}_{2,2}(t)$ and $P^B_{2,2}(t)$. We assume that to remain in the undetected PCa state, the patient simply needs both to remain alive and to avoid metastasis, if no biopsy is performed. On the other hand, if there is a biopsy, then in order to remain in the undetected PCa state, the patient’s biopsy must fail to find the PCa, because we assume that every detected non-metastasized PCa is immediately treated with radical prostatectomy. Because we assume that the biopsy procedure has a fixed sensitivity equal to $f$ and has perfect specificity, the NB probability is simply multiplied by $(1 - f)$ to obtain the corresponding B probability.

**Transition probabilities** $P^\text{NB}_{2,3}(t)$. We assume that treatment of undetected PCa can only be triggered by a positive biopsy, unless clinical detection occurs. In our model, clinical detection is triggered at the onset time plus the leadtime. When this occurs, the patient transitions deterministically to the treatment state. That is, when the leadtime clock runs down, the transition probability matrices are superseded by a one-time deterministic transition to state T in that period. Thus the use of the transition probability matrices at time $t$ is based on the background assumption that clinical detection does not occur at time $t$. For these reasons, $P^\text{NB}_{2,3}(t)$ is assumed to be zero.

**Transition probabilities** $P^B_{2,3}(t)$. As already stated, the use of the transition probability matrices at time $t$ assumes that clinical detection does not occur at time $t$, and aside from clinical detection, the only way for undetected PCa to be detected and treated is via a positive biopsy. Given that the patient has undetected PCa and a biopsy is performed at time $t$, the conditional probability that the patient makes an immediate transition to the treatment state is the product of the probabilities of three independent events:
(i) the biopsy detects PCa with probability $f$ so that the patient makes the transition $C \rightarrow T$ immediately; (ii) during the current treatment period, the PCa does not metastasize (i.e., the transition $T \rightarrow M$ does not occur) with probability $1 - b_t$; and (iii) during the current treatment period, the patient does not die with probability $1 - d_t$.

**Transition probabilities $P_{2,4}^{NB}(t)$ and $P_{2,4}^{B}(t)$**. In order to transition from state $C$ to $M$ when no biopsy occurs, the patient must survive (not die of other causes) and must develop metastatic PCa (governed by probability $e_t$). If a biopsy occurs and the patient is in state $C$, there are two mutually exclusive ways to metastasize in the current time period. The first way is for PCa to be detected and treated; the second is for PCa to remain undetected (no treatment performed). If treatment is performed, then as explained in the case of $P_{2,3}^{B}(t)$ the governing metastasis probability is $b_t$. In the absence of treatment, the governing metastasis probability is $e_t$, since no curative benefit is afforded. Thus the terms $fb_t$ and $e_t(1 - f)$ respectively express the treatment and no-treatment ways that undetected PCa can metastasize when there is a biopsy at the commencement of the time period. And, of course, those terms are multiplied by the probability that the independent event other-cause-death does not occur in the period.

Regardless of whether a biopsy occurs, our model assumes that there is an interval of time immediately following PCa onset during which it is impossible for PCa to metastasize [25]. We refer to this time as the preclinical dwelling time ($t_{PD}$). During the $t_{PD}$ years immediately following a transition from state NC to state C, the value of parameters $b_t$ and $e_t$ are set to 0, disallowing transitions into state M during the preclinical-dwelling-time interval. In particular, we assume that if the transition $NC \rightarrow C$ occurred at time $t_c$, then

$$b_t = e_t = 0 \text{ for } t \in \{t_c + 1, \ldots, t_c + [t_{PD}]\}. \quad (3.1)$$

An immediate implication of Equation (3.1) is that at each time $t \in T$, the one-step transition probability $P_{i,j}^{B}(t)$ and $P_{i,j}^{NB}(t)$ depend in general not only on the current health state $s_t$ but also on the previous health history $\{s_{t-\ell} : \ell = 1, \ldots, [t_{PD}]\}$ so that the stochastic process $\{s_t : t \in T\}$ is not a Markov chain in the usual sense. The $t_{PD}$ of localized PCa in Draisma et al. [25] is approximately 11.86 years. Therefore we use 12 years as the length of the preclinical dwelling time ($t_{PD} = 12$).
**Transition probabilities** \( P_{3,3}^{NB}(t) \) and \( P_{3,3}^{B}(t) \). After a patient has undergone treatment, in order to remain in state T in the current time period, the patient simply must in the same period avoid metastasis and death from other causes. Because our model assumes that no PSA screens or biopsies take place after a patient has undergone treatment, there is no reason to differentiate \( P_{3,3}^{B}(t) \) from \( P_{3,3}^{NB}(t) \). In fact, in light of that model assumption, the following parameters are never actually needed: \( P_{3,j}^{B}(t) \) for any \( j, t \). Thus, we assume that \( s_t = T \) implies \( a_{2,t} = NB \).

**Transition probabilities** \( P_{3,4}^{NB}(t) \). Given that treatment has been performed in a prior period, a patient transitions from state T to M if the patient remains alive and the treated PCa metastasizes. Such a case is representative of a real-world situation in which treatment fails to remove all of the cancerous tissue, and the remaining portion metastasizes. The governing metastasis probability is \( b_t \).

**Transition probabilities** \( P_{4,4}^{NB}(t) \). We assume that treatment in a previous time period does not affect the survival outcome for a patient with metastatic PCa. This means for a patient in state \( s_t = M \), the likelihood of avoiding PCa-specific mortality is independent of whether there exists some \( t' < t \) such that \( s_{t'} = T \). We also assume that there is no remission of PCa (metastatic or non-metastatic). Therefore, to remain in state M, a patient has to avoid both dying due to PCa and dying by some other cause. Because we assume that no PSA screens, biopsies, or treatments take place for patients with metastatic PCa, there is no reason to differentiate \( P_{4,4}^{B}(t) \) from \( P_{4,4}^{NB}(t) \). In fact, as with transitions from state T, transitions from state M never involve biopsy. That is, the transition probabilities \( P_{4,j}^{B}(t) \) are never actually used for any \( j, t \), since we assume that \( s_t = M \) implies \( a_{2,t} = NB \).

**Transition probabilities** \( P_{4,5}^{NB}(t) \). Given metastatic PCa, there are two mutually exclusive ways of dying. The patient may die of other causes, or the patient may die due to PCa (and not to other causes).

### 3.2.2 Model Parameters

The risk group consisting of white US males without a family history of PCa and without comorbidities is the reference subpopulation from which we derived parameters for the other risk groups. The derived
parameters of the various risk groups, including the parameters of the reference risk group that the derived parameters are based upon, are listed in Table 3.2. In the following paragraphs, we describe how the parameters in Table 3.2 are estimated.

Table 3.2: Parameters involved in the risk-specific parameter derivations for the African American, PCa Family History, (CCI=1), and (CCI≥2) risk groups.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_t$</td>
<td>Annual probability of other-cause death for whites of age $t + 40$</td>
</tr>
<tr>
<td>$d_{t}^{AA}$</td>
<td>Annual probability of other-cause death for African Americans of age $t + 40$</td>
</tr>
<tr>
<td>$d_{t}^{C1}$</td>
<td>Annual probability of other-cause death for men with CCI=1 of age $t + 40$</td>
</tr>
<tr>
<td>$d_{t}^{C2}$</td>
<td>Annual probability of other-cause death for men with CCI≥2 of age $t + 40$</td>
</tr>
<tr>
<td>$w_t$</td>
<td>Annual probability of developing PCa for whites of age $t + 40$</td>
</tr>
<tr>
<td>$w_{t}^{AA}$</td>
<td>Annual probability of developing PCa for African Americans of age $t + 40$</td>
</tr>
<tr>
<td>$w_{t}^{FH}(\ell)$</td>
<td>Annual probability of developing PCa for men with PCa Family History of age $t + 40$</td>
</tr>
<tr>
<td>$z_t$</td>
<td>Annual probability of PCa-specific death given mPCa for whites of age $t + 40$</td>
</tr>
<tr>
<td>$z_{t}^{AA}$</td>
<td>Annual probability of PCa-specific death given mPCa for African Americans of age $t + 40$</td>
</tr>
<tr>
<td>$q_t$</td>
<td>Annual probability of PCa-specific death for whites of age $t + 40$</td>
</tr>
<tr>
<td>$q_{t}^{AA}$</td>
<td>Annual probability of PCa-specific death for African Americans of age $t + 40$</td>
</tr>
<tr>
<td>$D_{t}^{AA}$</td>
<td>Annual all-cause probability of death for African Americans of age $t + 40$</td>
</tr>
<tr>
<td>$R_{z}^{AA}$</td>
<td>Ratio for converting $z_t$ to $z_{t}^{AA}$</td>
</tr>
<tr>
<td>$R_{w,t}^{AA}$</td>
<td>Ratio for converting $w_t$ to $w_{t}^{AA}$</td>
</tr>
<tr>
<td>$R_{q,t}^{AA}$</td>
<td>Ratio for converting $q_t$ to $q_{t}^{AA}$</td>
</tr>
<tr>
<td>$R_{w}^{FH}(\ell)$</td>
<td>Ratio for converting $w_t$ to $w_{t}^{FH}$</td>
</tr>
<tr>
<td>$R_{d,t}^{C1}$</td>
<td>Ratio for converting $d_t$ to $d_{t}^{C1}$</td>
</tr>
<tr>
<td>$R_{d,t}^{C2}$</td>
<td>Ratio for converting $d_t$ to $d_{t}^{C2}$</td>
</tr>
<tr>
<td>$\ell$</td>
<td>Index to PSA interval $[0, 1), [1, 2), [2, 3), [3, 4), [4, 7), [7, 10), or [10, \infty)$</td>
</tr>
</tbody>
</table>
**Incidence Rates**

To estimate the age-specific annual PCa incidence rate for African American men ($w_{t}^{AA}$), we multiplied the following quantities: (a) the ratio $R_{w,t}^{AA}$ of the relevant 5-year average annual incidence rate for African American men to that for the reference subpopulation; and (b) the age-specific annual incidence rate for the reference subpopulation ($w_{t}$).

To account for family history of white males, we used the logistic regression model for PCa risk based on biopsy data from the PCPT [105]. Specifically, using freely-accessible R code provided by the PCPT, we computed the following quantities for PSA levels 0.5, 1.5, 2.5, 3.5, 5.5, 8.5, and 10 ng/mL: (i) the risk of biopsy-detectable PCa for Caucasian males with a family history of PCa and without a DRE or prostate biopsy; and (ii) the risk of biopsy-detectable PCa for Caucasian males without a family history of PCa and without a DRE or prostate biopsy. At each PSA level, by dividing the risk given a family history of PCa by the risk given no family history of PCa, we obtained ratios for converting the annual probability of developing PCa for the reference risk group to an estimate for white US men with a family history of PCa. Let $w_{t}^{FH}(\ell)$ denote the annual probability of developing PCa for the men of age $t + 40$ in the Family History risk group, where $w_{t}^{FH}(\ell)$ is a function of the index $\ell \in \{1, 2, \ldots, 7\}$ that corresponds to a patient’s PSA level as follows:

<table>
<thead>
<tr>
<th>$\ell$</th>
<th>PSA (ng/mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[0, 1)</td>
</tr>
<tr>
<td>2</td>
<td>[1, 2)</td>
</tr>
<tr>
<td>3</td>
<td>[2, 3)</td>
</tr>
<tr>
<td>4</td>
<td>[3, 4)</td>
</tr>
<tr>
<td>5</td>
<td>[4, 7)</td>
</tr>
<tr>
<td>6</td>
<td>[7, 10)</td>
</tr>
<tr>
<td>7</td>
<td>[10, \infty)</td>
</tr>
</tbody>
</table>

Each PSA-specific annual probability of developing PCa for the men in the Family History risk group is computed as follows:

$$w_{t}^{FH}(\ell) = R_{w}^{FH}(\ell) \times w_{t} \quad \text{for} \quad \ell = 1, 2, \ldots, 7.$$  \hspace{1cm} (3.2)
**Mortality Rates**

The annual mortality rate from causes other than PCa, \( d_t \), was estimated by subtracting the annual PCa-specific mortality rate, \( q_t \), estimated from the 5-year rate given in SEER data [58], from the annual all-cause mortality rate, estimated from the 5-year rate reported by the CDC [50]. For each cancer grade in Ghani et al. [39], we multiplied the probability that PCa upon detection is found to be of that grade [39] by the probability that a prostate tumor of that grade metastasizes [17]. We then estimated \( e_t \), the probability of metastasis from undetected PCa, as the sum of these products. This summing was done to derive an overall metastasis probability over all grades, since our model state-space does not differentiate PCa by tumor grade.

The relationship between survival and the Charlson Comorbidity Index (CCI) [19] was established by Albertsen et al. [3], who conducted a 10-year survival and cause-of-death analysis of PCa patients from SEER data linked to Medicare insurance program files. Let \( d_{C1}^{t} \) and \( d_{C2}^{t} \) denote the annual probability of other-cause death for the men of age \( t + 40 \) in the (CCI=1) and (CCI≥2) risk groups, respectively. The percentage increase over time in the other-cause mortality rate for patients based on age and CCI was estimated by comparing the reported PCa-specific and all-cause mortality rates for males having CCI=0 with the corresponding rates for males with CCI>0. Using these comparisons, we computed adjusted annual other-cause mortality rates for the selected subpopulations with different comorbid medical conditions and different age ranges. For white men with CCI=1 in the age groups 66–74 and 75+, the ratios \( R_{C1}^{d,t} \) of their annual other-cause mortality rate to the corresponding rate for the reference subpopulation were, respectively, 2.449 and 1.343. The corresponding ratios \( R_{C2}^{d,t} \) for white men with CCI≥2 were, respectively, 3.648 and 1.821. The other-cause mortality rates for the age group <66 were the same as for the reference subpopulation under age 66.

To estimate each age-specific annual PCa mortality rate for African American males \( q_{t}^{AA} \), we multiplied the following quantities: (a) the ratio \( R_{q,t}^{AA} \) of the relevant 5-year average annual PCa-specific mortality rate for African American men to the corresponding 5-year average annual PCa specific mortality rate for the reference population; and (b) the age-specific annual PCa-specific mortality rate for the reference population \( q_{t} \). To estimate each age-specific annual other-cause mortality rate for African
American males ($d_{t}^{AA}$), we then subtracted our estimate $q_{t}^{AA}$ of the corresponding age-specific annual PCa mortality rate from the annual all-cause mortality rate for African American males $D_{t}^{AA}$ as given in US life-tables [26].

The annual probability of PCa-specific death for African American men with metastatic PCa (mPCa), $z_{t}^{AA}$, was approximated by the product of the following quantities: (a) the annual probability $z_{t}$ of PCa-specific death for men from the reference subpopulation with mPCa; and (b) the African American-vs-white hazard ratio $R_{z}^{AA}$ of all-cause mortality reported in a study on the outcomes of African American and white men with mPCa [103].

For each of the nonreference risk groups, we multiplied relevant parameter estimates from the base-case reference risk group by adjustment ratios derived from the literature. In Table 3.3, we present our estimates of the multiplicative adjustment ratios along with their sources.

The numerical estimates and sources for the multiplicative adjustment ratios used to derive the risk-specific parameter estimates for each of the nonreference risk groups from the corresponding estimates for the reference risk group are shown in Table 3.3.
Table 3.3: Multiplicative adjustment ratios used to derive key parameters of the nonreference risk groups. The term “onset probability” is synonymous with the term “probability of developing PCa.”

<table>
<thead>
<tr>
<th>Description of Ratio*</th>
<th>Age (yr.)</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{AA}^{\frac{z}{2}}$</td>
<td>Ratio of African American–to–white annual probability of PCa-specific death given mPCa [103]</td>
<td>40+</td>
</tr>
<tr>
<td>$R_{AA}^{\frac{q,t}{t}}$</td>
<td>Ratio of African American–to–white annual probability of PCa-specific death [83]</td>
<td>40–44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45–49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50–54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55–59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60–64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>65–69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70–74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75–79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80–84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>85+</td>
</tr>
<tr>
<td>$R_{w,t}^{\frac{AA}{w}}$</td>
<td>Ratio of African American–to–white PCa onset probability [83]</td>
<td>40–44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45–49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50–54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55–59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60–64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>65–69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70–74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75–79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80–84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>85+</td>
</tr>
<tr>
<td>$w_{FH}^{1}(1)$</td>
<td>Ratio of Family History–to–white PCa onset probability for PSA ∈ [0, 1) ng/mL [105]</td>
<td>40+</td>
</tr>
<tr>
<td>$w_{FH}^{1}(2)$</td>
<td>Ratio of Family History–to–white PCa onset probability for PSA ∈ [1, 2) ng/mL [105]</td>
<td>40+</td>
</tr>
<tr>
<td>$w_{FH}^{1}(3)$</td>
<td>Ratio of Family History–to–white PCa onset probability for PSA ∈ [2, 3) ng/mL [105]</td>
<td>40+</td>
</tr>
<tr>
<td>$w_{FH}^{1}(4)$</td>
<td>Ratio of Family History–to–white PCa onset probability for PSA ∈ [3, 4) ng/mL [105]</td>
<td>40+</td>
</tr>
<tr>
<td>$w_{FH}^{1}(5)$</td>
<td>Ratio of Family History–to–white PCa onset probability for PSA ∈ [4, 7) ng/mL [105]</td>
<td>40+</td>
</tr>
<tr>
<td>$w_{FH}^{1}(6)$</td>
<td>Ratio of Family History–to–white PCa onset probability for PSA ∈ [7, 10] ng/mL [105]</td>
<td>40+</td>
</tr>
<tr>
<td>$w_{FH}^{1}(7)$</td>
<td>Ratio of Family History–to–white PCa onset probability for PSA ≥ 10 ng/mL [105]</td>
<td>40+</td>
</tr>
<tr>
<td>$R_{d,d}^{C1}$</td>
<td>Ratio of (CCI=1)–to–white annual probability of other-cause death [3]</td>
<td>40–65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>66–74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75+</td>
</tr>
<tr>
<td>$R_{d,d}^{C2}$</td>
<td>Ratio of (CCI≥2)–to–white annual probability of other-cause death [3]</td>
<td>40–65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>66–74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75+</td>
</tr>
</tbody>
</table>

* For each ratio with the time subscript $t$, the corresponding age is $t + 40$. 
**PSA Sampling**

The method for sampling PSA values was based on a linear random effects model [40]. The model uses a linear changepoint formulation for log(PSA) such that the growth rate of log(PSA) accelerates after the onset of PCa. Parameters of the model were based on data from the control arm of the Prostrate Cancer Prevention Trial (PCPT) [104, 34]. This model was previously validated against data from the Prostate, Lung, Colorectal, and Ovarian Cancer Screening (PLCO) Trial [9]. To estimate the PSA samples for African American patients without PCa, we used the age-dependent ratio of the mean PSA level for African American men to the mean PSA level for white US men [49]. To estimate the PSA samples for African American patients with nonmetastatic PCa, we used the ratio of the mean PSA level for African American men to the mean PSA level for white US men from a Radiation Therapy Oncology Group prospective registration study [108], which found no correlation between PSA level and age.

**3.2.3 Screening Strategies**

For our comparative analysis, we selected the PSA-threshold-based PCa screening strategies that were used in two of the largest randomized control trials on PCa screening — namely, the PLCO [10] and the ERSPC [95] trials — as well as strategies considered in other prominent studies [86, 48]. The selected strategies are listed in Table 3.4.
Table 3.4: PSA screening strategies compared in this study. Strategy #1 is no screening. Strategies #1–#8 were taken from a prominent simulation study [86]. Strategy #5 was also used in the PLCO trial. Strategies #9–#14 were used in the ERSPC trial.

<table>
<thead>
<tr>
<th>Strategy Label</th>
<th>Range of Ages (yr)</th>
<th>Screen Interval (yr)</th>
<th>PSA Threshold (ng/mL)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>#2</td>
<td>40–75</td>
<td>5</td>
<td>4.0</td>
<td>[86]</td>
</tr>
<tr>
<td>#3</td>
<td>50–75</td>
<td>2</td>
<td>4.0</td>
<td>[86]</td>
</tr>
<tr>
<td>#4</td>
<td>50–59</td>
<td>1</td>
<td>3.5</td>
<td>[86]</td>
</tr>
<tr>
<td></td>
<td>60–69</td>
<td>1</td>
<td>4.5</td>
<td>[86]</td>
</tr>
<tr>
<td></td>
<td>70–75</td>
<td>1</td>
<td>6.5</td>
<td></td>
</tr>
<tr>
<td>#5</td>
<td>50–75</td>
<td>1</td>
<td>4.0</td>
<td>[86, 10]</td>
</tr>
<tr>
<td>#6</td>
<td>50–75</td>
<td>1</td>
<td>2.5</td>
<td>[86]</td>
</tr>
<tr>
<td>#7</td>
<td>40, 45</td>
<td>–</td>
<td>4.0</td>
<td>[86]</td>
</tr>
<tr>
<td></td>
<td>50–75</td>
<td>1</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>#8</td>
<td>40, 45</td>
<td>–</td>
<td>4.0</td>
<td>[86]</td>
</tr>
<tr>
<td></td>
<td>50–75</td>
<td>2</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>#9</td>
<td>55–69</td>
<td>4</td>
<td>3.0</td>
<td>[48]</td>
</tr>
<tr>
<td>#10</td>
<td>55–74</td>
<td>1</td>
<td>3.0</td>
<td>[48]</td>
</tr>
<tr>
<td>#11</td>
<td>55–69</td>
<td>4</td>
<td>3.0</td>
<td>[48]</td>
</tr>
<tr>
<td>#12</td>
<td>55</td>
<td>–</td>
<td>3.0</td>
<td>[48]</td>
</tr>
<tr>
<td>#13</td>
<td>60</td>
<td>–</td>
<td>3.0</td>
<td>[48]</td>
</tr>
<tr>
<td>#14</td>
<td>65</td>
<td>–</td>
<td>3.0</td>
<td>[48]</td>
</tr>
</tbody>
</table>
3.2.4 Estimation Methods

To compare screening strategies, we compute an estimate of the expected NNS for a given strategy, which we have defined as the expected number needed to screen under that strategy to avert 1 PCa death compared with no screening. Estimating this quantity requires the synchronization of patient histories, which is achieved by careful synchronization of the random number streams underlying the simulations.

Estimating Expected NNS

To estimate the expected value of the NNS, we simulated a population of patients both under screening and under no-screening, and computed our NNS estimate based on a comparison of the observed PCa-specific mortality rates in the screening and no-screening simulations. In particular, to estimate NNS, we simulated a population of \( m \) patients under the given screening strategy; and then we simulated the same population of patients under the no-screening strategy, allowing the patient histories to diverge only upon screen detection of PCa. In this way, for the \( j \)th patient in the population without screening (where \( j = 1, \ldots, m \)), we obtain an estimate of what the \( j \)th patient’s health history would have been had he not been screen detected. We synchronized patient histories up to the point of screen detection by using the same sequence of random numbers to simulate the PSA and health trajectories of the \( j \)th simulated patient in both the screening and no-screening simulations, where different (independent) sequences of random numbers were used for different patients. By synchronizing patient histories in this way, we could compute the total number \( D \) of PCa deaths averted by the given screening strategy. We estimated NNS for the given screening strategy by the quantity \( m/D \). We derived a 95% confidence interval (CI) estimator of the expected NNS by the univariate delta method [52].

Formally, from a target population of men following a given screening strategy, we simulated a random sample of \( m \) patients who adhered to that screening strategy in the screening simulations and who did not undergo any screening in the no-screening simulations; and for the \( j \)th such patient (where
\( j = 1, \ldots, m \), let

\[
J_{S,j} \equiv \begin{cases} 
1 & \text{if patient } j \text{ did not die due to PCa in the screening (S) simulation during the follow-up period,} \\
0 & \text{otherwise}; 
\end{cases}
\tag{3.3}
\]

and

\[
J_{NS,j} \equiv \begin{cases} 
1 & \text{if patient } j \text{ did not die due to PCa in the no-screening (NS) simulation during the follow-up period,} \\
0 & \text{otherwise.} 
\end{cases}
\tag{3.4}
\]

The follow-up period is simply a length of time after the commencement of simulation of a single patient at time \( t = 0 \) during which the patient’s history is monitored. If a patient does not die from PCa within this follow-up period, then the patient is deemed to have \textit{not} died due to PCa. In this way, we are able to emulate RCTs that censor patient data in order to analyze trial-participant mortality results at different points in time. Except for the experiment described in Section 3.3.2, all experiments in this thesis use an infinite follow-up period. Using an infinite follow-up period, every patient is followed until he dies, and there is no censoring.

With the setup in Equations (3.3) and (3.4), we say that screening with the given screening strategy averted PCa death for patient \( j \) if and only if \( J_{S,j} - J_{NS,j} = 1 \). Recognizing that \( J_{S,j} - J_{NS,j} \) can take the values 1, 0, or \(-1\) for each patient in the simulated population, we compute the total number of PCa deaths averted,

\[
D = \sum_{j=1}^{m} (J_{S,j} - J_{NS,j}) , \tag{3.5}
\]

which has the corresponding sample mean

\[
\bar{D} = \frac{1}{m} \sum_{j=1}^{m} (J_{S,j} - J_{NS,j}) .
\]
Let 
\[ D_j = J_{S,j} - J_{NS,j} \quad \text{for} \quad j = 1, \ldots, m. \]

Our point estimate of NNS is
\[
\hat{NNS} = \frac{m}{D} = \frac{1}{\bar{D}}; \tag{3.6}
\]
and an asymptotically valid 100(1 − α)% CI for NNS is
\[
\frac{1}{\bar{D}} \pm (\hat{z}_{1-\alpha/2}) \frac{S_D}{(\bar{D})^2 \sqrt{m}}. \tag{3.7}
\]
where \( \hat{z}_{1-\alpha/2} \) is the \( 1 - \alpha/2 \) quantile of the standard normal distribution and
\[
S_D^2 = \frac{1}{m-1} \sum_{j=1}^{m} (D_j - \bar{D})^2. \tag{3.8}
\]

The CI in Equation (3.7) is derived from the univariate delta method. Hogg et al. [52] show that if a sequence of random variables \( \{X_m\} \) is such that \( \sqrt{m} (X_m - \theta) \) converges in distribution to \( N(0, \sigma^2) \) as \( m \to \infty \), and also the function \( h(x) \) is differentiable at \( \theta \) with \( h'(\theta) \neq 0 \), then
\[
\sqrt{m} (h(X_m) - h(\theta))
\]
converges in distribution to
\[
N \left( 0, \sigma^2[h'(\theta)]^2 \right)
\]
as \( m \to \infty \). In the case of our NNS estimator, we have \( X_m = \bar{D} \) and \( h(x) = 1/x \), which is differentiable for \( x \neq 0 \). Also, \( \theta = E[\bar{D}] = \mu_D \) and \( h'(\mu_D) = 1/\mu_D^2 \neq 0 \) if \( \mu_D > 0 \). From Hogg et al. [52], the
asymptotic variance of the distribution of \( h(\bar{D}) = 1/\bar{D} = \hat{\text{NNS}} \) is equal to

\[
\text{Var}[h(\bar{D})] \approx \text{Var}[\bar{D}] \left( h'(E[\bar{D}]) \right)^2 = \frac{\sigma_D^2}{m} \left( \frac{1}{\mu_D^2} \right)^2 = \frac{\sigma_D^2}{m \cdot \mu_D^4}.
\]

for which we use the sample-variance estimator

\[
\frac{S_D^2}{m (\bar{D})^4}.
\]

From Equation (3.9), we have the standard error term

\[
\frac{1}{\sqrt{m}} \sqrt{\frac{S_D^2}{(\bar{D})^4}} = \frac{S_D}{(\bar{D})^2 \sqrt{m}}.
\]

which when used to form a CI on \( \hat{\text{NNS}} \) produces the result shown in Equation (3.7).

Because some PCa cases are diagnosed clinically, apart from PSA screening, an unbiased estimation of NNS must account for clinical detection. Our model accounts for clinical detection by incorporating PSA-screening leadtime. We assume that once a patient is in state C and has a PSA level \( \geq 3 \) ng/mL, the “leadtime clock” starts. If a patient with PCa first achieves a PSA level \( \geq 3 \) ng/mL at time \( t^\dagger \in \mathcal{T} \), then the time of clinical diagnosis is

\[
t^\ddagger = t^\dagger + [t_{LT}] \in \mathcal{T}.
\]

If at time \( t^\ddagger \) the patient is still alive and has been neither diagnosed nor treated for PCa, then the patient transitions deterministically from state C to T.

Figure 3.3 illustrates the leadtime clock on a particular patient with PCa onset at age 42 and with the shown PSA sample path from ages 40 to 61. From Figure 3.3, we see that the earliest age at which the patient has PCa and has a PSA level \( \geq 3 \) ng/mL is 48 years so that \( t^\dagger = 48 - 40 = 8 \). Therefore, the leadtime clock starts at age 48 (time \( t^\dagger = 8 \)) for this patient. With a leadtime estimate of \( t_{LT} = 12.8 \) years, the leadtime clock will run down at time \( t^\ddagger = t^\dagger + [t_{LT}] = 8 + 13 = 21 \) (age \( t^\ddagger + 40 = 61 \).
years). Therefore, clinical detection for this patient will occur at age 61 if the patient is still alive and in state C at that time. If, on the other hand, the patient is screen detected or develops metastatic PCa or dies before age 61, then clinical detection will not occur.

In our model, we assume that no patient ≥95 years of age is treated upon clinical detection. We also assume that $t_{LT}$ is fixed for all patients. The PSA level of 3 ng/mL used to trigger the leadtime clock was chosen because 3 ng/mL is the baseline PSA level used by Savage et al. [90].

![Diagram of the leadtime clock](image)

**Figure 3.3:** Illustration of the leadtime clock corresponding to a hypothetical patient with PCa onset at age 42 and with the PSA levels shown from age 40 to 61.
Synchronized Patient Histories

To most accurately gauge the effect of a given screening strategy, we use carefully synchronized streams of random numbers, which the simulation literature refers to as the method of common random numbers (CRN) [69, 68], to synchronize patient histories. Each patient in the simulated population has his own dedicated sequences (streams) of random numbers that are used in the simulation to determine his health history—that is, the sequence of health states and PSA levels for that patient over the course of his lifetime. When the patient is simulated without screening, then we obtain a “natural” health history in the absence of any screening-based medical intervention to avoid death caused by PCa. When the same patient is simulated under a given screening strategy, then the same (i.e., synchronized) sequences of random numbers are used to re-create his health history up to the time at which a screening exam detects PCa; and beyond this point, the rest of his dedicated sequences of random numbers are used to generate his modified health history, which may, for example, involve transitions between different health states following treatment when compared with the unmodified health history in the no-screening simulation. And the divergence of the modified health history from the unmodified health history is precisely what we are interested in for the purpose of gauging the impact of a given screening strategy.

Such a comparative analysis of each patient would be impossible in an RCT since it is impossible to observe the counterfactual case. But by careful use of CRN, we are able to simulate the same patient life under two different experimental conditions (screening and no screening). This induces dependence that allows for paired comparisons across patients and significantly reduces the variance in the outcomes of the model that compare screening versus no screening.

The synchronization of patient histories is done by separately synchronizing two components of the simulation model: the simulated PSA levels and the state transitions. The PSA sampling model from Gulati et al. [40] mentioned above is based upon the following stochastic formula for computing the $j$th patient’s PSA level at time $t$ (age $40 + t$):

$$\log(\text{PSA}_j(t)) = \beta_{0,j} + \beta_{1,j} \cdot t + \beta_{2,j}(t - t_{0,j}) \cdot I(t > t_{0,j}) + \epsilon_{t,j},$$

(3.10)

where $\beta_{0,j}, \beta_{1,j}, \beta_{2,j}$, and $\epsilon_{t,j}$ are normal random variables from distributions specified in Gulati et al.
that vary from patient to patient, \( t_{0,j} \) is the time at which patient \( j \) developed PCa (equal to \(+\infty\) if the patient has not yet developed PCa), and where \( \mathbb{I}(\cdot) \) is the indicator function. The means and variances of the normal random variables \( \beta_{k,j} \) for \( k = 0, 1, 2 \) and \( \varepsilon_{t,j} \) are listed in Table 3.5.

Table 3.5: Population means and variances of normally-distributed random variables in the log-PSA growth model expressed in Equation (3.10). The values are taken directly from Gulati et al. [40].

<table>
<thead>
<tr>
<th>Parameter description</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_{0,j} )  Preonset intercept</td>
<td>-1.6094</td>
<td>0.0568</td>
</tr>
<tr>
<td>( \beta_{1,j} )  Preonset slope</td>
<td>0.0200</td>
<td>0.0019</td>
</tr>
<tr>
<td>( \beta_{2,j} )  Postonset slope increment</td>
<td>0.1094</td>
<td>0.0237</td>
</tr>
<tr>
<td>( \varepsilon_{t,j} )  Within-patient noise</td>
<td>0.0000</td>
<td>0.0829</td>
</tr>
</tbody>
</table>

Whenever commencing to simulate patient \( j \), either in the screening or the no-screening simulation, a single stream of random numbers is initialized with seed value \( S_j \), and then this stream is used to sample the following quantities in order: \( \beta_{0,j}, \beta_{1,j}, \beta_{2,j}, \varepsilon_{1,j}, \varepsilon_{2,j}, \ldots, \varepsilon_{60,j} \). Having computed these 63 quantities using the first 63 uniform random variates from the PSA random-number stream devoted to PSA sampling, we use Equation (3.10) to compute patient \( j \)'s PSA level at time \( t \), \( \text{PSA}_j(t) \), depending on the time at which patient \( j \) develops PCa, \( t_{0,j} \). In this way, by precomputing these 63 random quantities, we can deterministically compute a sampled PSA value for patient \( j \) at time \( t \) that will always be exactly the same under either the screening or no-screening simulation, as long as the time at which PCa develops is also synchronized, which is guaranteed by virtue of the synchronization of the underlying model that generates patient health-state transitions.

To synchronize the health-state trajectory over time for patient \( j \), we utilize another stream of uniform random variates, which is devoted to the state transitions. In particular, whenever commencing to simulate patient \( j \), either in the screening or the no-screening simulation, this random number stream is initialized with seed value \( H_j \). Each state transition of the underlying model, utilizes one uniform random variate from the random number stream. The state to which the patient transitions depends only upon the following: (i) whether a prostate biopsy was performed in the current period; and (ii) the health
state occupied by the patient during the current period (i.e., the state from which the patient transitions). Consequently, as long as patient $j$ is always simulated with the exact same starting time $t = 0$ (age 40) and starting state $s_0$, and the exact same sequence of biopsy decisions, the health state trajectory for patient $j$ will always be exactly identical. The biopsy decisions are based on the screening strategy and the sampled PSA values. As described above, the sampled PSA values are already synchronized. Therefore, the only way in which the patient histories for two different simulations of patient $j$ could be different is due to the screening strategy. And the screening strategy is by design different between the screening and no-screening simulations, and therefore the sequences of biopsy decisions in the screening and no-screening simulations will not always be identical. This divergence of patient histories is precisely what permits the evaluation of screening strategies.

The synchronization of patient histories by the separate synchronization of PSA levels and patient health states is depicted in Figure 3.4. In Figure 3.4, it is shown how a hypothetical patient’s health-state histories diverge immediately following screen detection (and by our model assumption treatment) of PCa. The PSA trajectory does not diverge following diagnosis of PCa, because PSA information is not used at all in the no-screening simulations. Realistically, a patient’s PSA level is affected by treatment for PCa, but this physiological aspect has absolutely no effect in the model, and so is not modeled.
Figure 3.4: Illustration depicting the divergence between patient histories in the screening and no-screening simulations. The hypothetical patient depicted in this illustration is shown simulated over 7 periods (years).

### 3.2.5 Procedural Description of Simulation Model

In view of the foregoing description of the simulation model’s state transition, PSA sampling, synchronized patient-histories, and NNS estimation components, we present in Algorithm 1 a concise but detailed procedural description of the simulation model. In the algorithm, we use mathematical notation to refer to already defined variables, but for the sake of brevity and clarity, we use self-descriptive nonsymbolic language to describe variables that have not been defined already. Also, certain parts of the algorithm are abstracted away and replaced with simple verbal descriptions. We do this because the primary purpose of the algorithm is to describe the overall process of the simulation, and we judge that certain parts of the algorithm need not be technically expounded.
Algorithm 1: SIMULATION(m)

Input: Number m of patients to simulate.

1 begin
2   foreach patient j ∈ {1, . . . , m} do
3     foreach subsim in {Screening, No Screening} do
4       Set age ← 40;
5       Set state ← NC;
6       Set ageAtTreatment ← -1;
7       Set screen ← false;
8       Set biopsy ← false;
9       while state ≠ D do
10          Compute patient PSA level;
11          Check whether to start leadtime clock;
12          Set lastState ← state;
13          if Clinical detection conditions met then
14            Set state ← T;
15          else
16            if Screen this period then
17              Set screen ← true;
18              if PSA level ≥ threshold then
19                Set biopsy ← true;
20                Transition state using P^B;
21              else
22                Transition state using P^NB;
23            end
24          else
25            Transition state using P^NB;
26          end
27          if state ≠ D then
28            if lastState ≠ T and state = T then
29              Set ageAtTreatment ← age;
30            end
31            Set age ← age + 1;
32          end
33        end
34      end
35    end
36 end
37
38 Compute \( D_j \);
3.3 Results

3.3.1 Numerical Estimates for Model Parameters

Table 3.6 lists the parameters of the simulation model along with the base-case numerical estimates and their sources. The estimates in Table 3.6 are for the reference risk group. Although the specificity of the prostate biopsy procedure is not a parameter of our model, our model does operate on the assumption of perfect specificity. That is, our model does not account for false-positive biopsy results.

Table 3.6: Parameters Used in Our PCa Natural-History Simulation, along with Sources for All Parameter Estimates. The value of 1.0 as the probability of other-cause death is a modest assumption made in our model only governing transitions at age 100 years.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Numerical Estimates</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_t$ annual probability of developing PCa</td>
<td>0.000–0.069</td>
<td>[44]</td>
</tr>
<tr>
<td>$d_t$ annual probability of other-cause death</td>
<td>0.000–1.000</td>
<td>[58, 50]</td>
</tr>
<tr>
<td>$b_t$ annual metastasis probability for patients in state T</td>
<td>0.000–0.007</td>
<td>[13]</td>
</tr>
<tr>
<td>$e_t$ annual metastasis probability for patients in state C</td>
<td>0.000–0.041</td>
<td>[62, 25]</td>
</tr>
<tr>
<td>$t_{PD}$ Preclinical dwelling time (yr)</td>
<td>11.86</td>
<td>[25]</td>
</tr>
<tr>
<td>$z_t$ annual probability of PCa-specific death given mPCa</td>
<td>0.181–0.204</td>
<td>[84]</td>
</tr>
<tr>
<td>$f$ Sensitivity of the prostate biopsy procedure</td>
<td>0.8</td>
<td>[44]</td>
</tr>
<tr>
<td>$t_{LT}$ PCa-screening leadtime (yr)</td>
<td>12.8</td>
<td>[90]</td>
</tr>
</tbody>
</table>

The estimates of parameter $w_t$, the annual probability of developing PCa, were inferred from Figure 2 of Haas et al. [44]. (We were unable to obtain the actual data points underlying the figure.) Since the figure only depicts probabilities extending to approximately 90 years, we used SAS JMP software to construct a spline fit from which we obtained probabilities extending to age 100. The inferred lower-bound, base-case, and upper-bound numerical estimates of $w_t$ are depicted in Figure 3.5 and listed in Tables 3.7, 3.8, and 3.9.
Figure 3.5: Lower-bound, base-case, and upper-bound numerical estimates of parameter $w_t$ inferred from Figure 2 of Haas et al. [44].

Table 3.7: Lower-bound numerical estimates for parameter $w_t$. The time $t$ (age $40 + t$) denotes that the corresponding $w_t$ value applies to transitions occurring during the interval $[t, t + 1)$.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$w_t$</th>
<th>$t$</th>
<th>$w_t$</th>
<th>$t$</th>
<th>$w_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0052</td>
<td>20</td>
<td>0.0190</td>
<td>40</td>
<td>0.0064</td>
</tr>
<tr>
<td>1</td>
<td>0.0010</td>
<td>21</td>
<td>0.0097</td>
<td>41</td>
<td>0.0226</td>
</tr>
<tr>
<td>2</td>
<td>0.0021</td>
<td>22</td>
<td>0.0172</td>
<td>42</td>
<td>0.0099</td>
</tr>
<tr>
<td>3</td>
<td>0.0062</td>
<td>23</td>
<td>0.0150</td>
<td>43</td>
<td>0.0167</td>
</tr>
<tr>
<td>4</td>
<td>0.0042</td>
<td>24</td>
<td>0.0127</td>
<td>44</td>
<td>0.0441</td>
</tr>
<tr>
<td>5</td>
<td>0.0021</td>
<td>25</td>
<td>0.0128</td>
<td>45</td>
<td>0.0142</td>
</tr>
<tr>
<td>6</td>
<td>0.0063</td>
<td>26</td>
<td>0.0130</td>
<td>46</td>
<td>0.0288</td>
</tr>
<tr>
<td>7</td>
<td>0.0042</td>
<td>27</td>
<td>0.0211</td>
<td>47</td>
<td>0.0093</td>
</tr>
<tr>
<td>8</td>
<td>0.0053</td>
<td>28</td>
<td>0.0108</td>
<td>48</td>
<td>0.0114</td>
</tr>
<tr>
<td>9</td>
<td>0.0053</td>
<td>29</td>
<td>0.0190</td>
<td>49</td>
<td>0.0196</td>
</tr>
<tr>
<td>10</td>
<td>0.0108</td>
<td>30</td>
<td>0.0097</td>
<td>50</td>
<td>0.0200</td>
</tr>
<tr>
<td>11</td>
<td>0.0109</td>
<td>31</td>
<td>0.0182</td>
<td>51</td>
<td>0.0204</td>
</tr>
<tr>
<td>12</td>
<td>0.0066</td>
<td>32</td>
<td>0.0085</td>
<td>52</td>
<td>0.0209</td>
</tr>
<tr>
<td>13</td>
<td>0.0088</td>
<td>33</td>
<td>0.0201</td>
<td>53</td>
<td>0.0213</td>
</tr>
<tr>
<td>14</td>
<td>0.0067</td>
<td>34</td>
<td>0.0059</td>
<td>54</td>
<td>0.0218</td>
</tr>
<tr>
<td>15</td>
<td>0.0112</td>
<td>35</td>
<td>0.0236</td>
<td>55</td>
<td>0.0223</td>
</tr>
<tr>
<td>16</td>
<td>0.0057</td>
<td>36</td>
<td>0.0151</td>
<td>56</td>
<td>0.0228</td>
</tr>
<tr>
<td>17</td>
<td>0.0171</td>
<td>37</td>
<td>0.0153</td>
<td>57</td>
<td>0.0233</td>
</tr>
<tr>
<td>18</td>
<td>0.0116</td>
<td>38</td>
<td>0.0093</td>
<td>58</td>
<td>0.0238</td>
</tr>
<tr>
<td>19</td>
<td>0.0118</td>
<td>39</td>
<td>0.0189</td>
<td>59</td>
<td>0.0238</td>
</tr>
</tbody>
</table>
Table 3.8: Basecase numerical estimates for parameter $w_t$. The time $t$ (age $40 + t$) denotes that the corresponding $w_t$ value applies to transitions occurring during the interval $[t, t+1)$.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$w_t$</th>
<th>$t$</th>
<th>$w_t$</th>
<th>$t$</th>
<th>$w_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0054</td>
<td>20</td>
<td>0.0168</td>
<td>40</td>
<td>0.0327</td>
</tr>
<tr>
<td>1</td>
<td>0.0043</td>
<td>21</td>
<td>0.0131</td>
<td>41</td>
<td>0.0295</td>
</tr>
<tr>
<td>2</td>
<td>0.0087</td>
<td>22</td>
<td>0.0186</td>
<td>42</td>
<td>0.0435</td>
</tr>
<tr>
<td>3</td>
<td>0.0044</td>
<td>23</td>
<td>0.0190</td>
<td>43</td>
<td>0.0227</td>
</tr>
<tr>
<td>4</td>
<td>0.0044</td>
<td>24</td>
<td>0.0110</td>
<td>44</td>
<td>0.0326</td>
</tr>
<tr>
<td>5</td>
<td>0.0056</td>
<td>25</td>
<td>0.0223</td>
<td>45</td>
<td>0.0625</td>
</tr>
<tr>
<td>6</td>
<td>0.0112</td>
<td>26</td>
<td>0.0143</td>
<td>46</td>
<td>0.0256</td>
</tr>
<tr>
<td>7</td>
<td>0.0056</td>
<td>27</td>
<td>0.0261</td>
<td>47</td>
<td>0.0316</td>
</tr>
<tr>
<td>8</td>
<td>0.0045</td>
<td>28</td>
<td>0.0179</td>
<td>48</td>
<td>0.0061</td>
</tr>
<tr>
<td>9</td>
<td>0.0091</td>
<td>29</td>
<td>0.0303</td>
<td>49</td>
<td>0.0425</td>
</tr>
<tr>
<td>10</td>
<td>0.0092</td>
<td>30</td>
<td>0.0156</td>
<td>50</td>
<td>0.0443</td>
</tr>
<tr>
<td>11</td>
<td>0.0116</td>
<td>31</td>
<td>0.0238</td>
<td>51</td>
<td>0.0464</td>
</tr>
<tr>
<td>12</td>
<td>0.0059</td>
<td>32</td>
<td>0.0244</td>
<td>52</td>
<td>0.0487</td>
</tr>
<tr>
<td>13</td>
<td>0.0130</td>
<td>33</td>
<td>0.0250</td>
<td>53</td>
<td>0.0512</td>
</tr>
<tr>
<td>14</td>
<td>0.0168</td>
<td>34</td>
<td>0.0256</td>
<td>54</td>
<td>0.0539</td>
</tr>
<tr>
<td>15</td>
<td>0.0061</td>
<td>35</td>
<td>0.0351</td>
<td>55</td>
<td>0.0570</td>
</tr>
<tr>
<td>16</td>
<td>0.0061</td>
<td>36</td>
<td>0.0182</td>
<td>56</td>
<td>0.0604</td>
</tr>
<tr>
<td>17</td>
<td>0.0173</td>
<td>37</td>
<td>0.0370</td>
<td>57</td>
<td>0.0643</td>
</tr>
<tr>
<td>18</td>
<td>0.0201</td>
<td>38</td>
<td>0.0192</td>
<td>58</td>
<td>0.0687</td>
</tr>
<tr>
<td>19</td>
<td>0.0064</td>
<td>39</td>
<td>0.0392</td>
<td>59</td>
<td>0.0687</td>
</tr>
</tbody>
</table>

56
Table 3.9: Upper-bound numerical estimates for parameter $w_t$. The time $t$ (age $40 + t$) denotes that the corresponding $w_t$ value applies to transitions occurring during the interval $[t, t+1)$.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$w_t$</th>
<th>$t$</th>
<th>$w_t$</th>
<th>$t$</th>
<th>$w_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0024</td>
<td>20</td>
<td>0.0145</td>
<td>40</td>
<td>0.0286</td>
</tr>
<tr>
<td>1</td>
<td>0.0049</td>
<td>21</td>
<td>0.0088</td>
<td>41</td>
<td>0.0588</td>
</tr>
<tr>
<td>2</td>
<td>0.0074</td>
<td>22</td>
<td>0.0208</td>
<td>42</td>
<td>0.0750</td>
</tr>
<tr>
<td>3</td>
<td>0.0062</td>
<td>23</td>
<td>0.0227</td>
<td>43</td>
<td>0.0541</td>
</tr>
<tr>
<td>4</td>
<td>0.0062</td>
<td>24</td>
<td>0.0171</td>
<td>44</td>
<td>0.0714</td>
</tr>
<tr>
<td>5</td>
<td>0.0025</td>
<td>25</td>
<td>0.0221</td>
<td>45</td>
<td>0.0692</td>
</tr>
<tr>
<td>6</td>
<td>0.0100</td>
<td>26</td>
<td>0.0290</td>
<td>46</td>
<td>0.0496</td>
</tr>
<tr>
<td>7</td>
<td>0.0101</td>
<td>27</td>
<td>0.0199</td>
<td>47</td>
<td>0.0870</td>
</tr>
<tr>
<td>8</td>
<td>0.0077</td>
<td>28</td>
<td>0.0237</td>
<td>48</td>
<td>0.1136</td>
</tr>
<tr>
<td>9</td>
<td>0.0013</td>
<td>29</td>
<td>0.0382</td>
<td>49</td>
<td>0.1041</td>
</tr>
<tr>
<td>10</td>
<td>0.0142</td>
<td>30</td>
<td>0.0325</td>
<td>50</td>
<td>0.1162</td>
</tr>
<tr>
<td>11</td>
<td>0.0052</td>
<td>31</td>
<td>0.0336</td>
<td>51</td>
<td>0.1314</td>
</tr>
<tr>
<td>12</td>
<td>0.0053</td>
<td>32</td>
<td>0.0347</td>
<td>52</td>
<td>0.1513</td>
</tr>
<tr>
<td>13</td>
<td>0.0185</td>
<td>33</td>
<td>0.0400</td>
<td>53</td>
<td>0.1783</td>
</tr>
<tr>
<td>14</td>
<td>0.0081</td>
<td>34</td>
<td>0.0417</td>
<td>54</td>
<td>0.2170</td>
</tr>
<tr>
<td>15</td>
<td>0.0149</td>
<td>35</td>
<td>0.0435</td>
<td>55</td>
<td>0.2772</td>
</tr>
<tr>
<td>16</td>
<td>0.0069</td>
<td>36</td>
<td>0.0455</td>
<td>56</td>
<td>0.3835</td>
</tr>
<tr>
<td>17</td>
<td>0.0139</td>
<td>37</td>
<td>0.0476</td>
<td>57</td>
<td>0.6220</td>
</tr>
<tr>
<td>18</td>
<td>0.0141</td>
<td>38</td>
<td>0.0700</td>
<td>58</td>
<td>1.0000</td>
</tr>
<tr>
<td>19</td>
<td>0.0143</td>
<td>39</td>
<td>0.0591</td>
<td>59</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
3.3.2 Model Validation

Our validation has two prongs: (i) validation of the fundamental PCa natural-history and PSA sampling simulation; and (ii) validation of the paired-comparisons–based estimation of NNS.

Validation of Trends in Diagnosis, Mortality, and Survival

For prong (i), we validated our model by examining PCa diagnosis and mortality rates, expected lifespan, and survival time for patients with metastatic PCa. Model estimates for whites and African Americans were compared with corresponding estimates in the literature. All model estimates were generated from a sample of 50,000 simulated patients who adhered to annual screening from age 50 to 75 with a PSA threshold of 4.0 ng/mL. For lifespan, we compared model-predicted life expectancy for 40-year-old white US males and African American males with the corresponding estimates from US life tables [26]. For metastatic PCa survival time, we compared the mean survival time for men having metastatic PCa from our model with the corresponding estimates based on a weighted average of survival times across patients with different PSA values. The latter survival times are based on a report concerning data from a phase III trial to evaluate the effect of androgen-deprivation therapy on survival of men with metastatic PCa [60]. For PCa diagnosis and mortality rates, we compared model estimates to observed rates reported in SEER statistics tables [58]. The model-estimated statistics used for validation and their corresponding literature estimates are shown in Table 3.10.

There are plausible reasons to expect some deviation of model estimates from literature estimates. For example, the model-estimated life expectancy may be elevated because our model assumes all simulated patients are disease free at age 40. Also, the model-estimated PCa mortality rate may be elevated if the literature estimates underrepresent the full PCa mortality in a given population; in the model, there are no unreported PCa deaths. Also, the model-estimated PCa diagnosis rate may deviate from the literature because the model assumes that all patients are screened under the prevailing screening strategy, and thus the model would not reflect any existing disparities in screening adherence or participation in a given demographic group.
Table 3.10: Comparing model-estimated PCa mortality and diagnosis rates, lifespan, and metastasized PCa survival time to corresponding estimates from the literature. Whites means white US males with no family history of PCa and CCI=0, AA means African American males with no family history of PCa and CCI=0, and mPCa means metastasized PCa. Model estimates obtained by simulating 50,000 patients screened annually from ages 50 to 75 at PSA threshold 4.0 ng/mL, and 95% confidence-interval half widths are given for model estimates.

<table>
<thead>
<tr>
<th></th>
<th>Model Estimate</th>
<th>Literature Estimate</th>
<th>Absolute Deviation Model vs Lit. Est.</th>
<th>% Relative Deviation Model vs Lit. Est.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PCa mortality rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whites</td>
<td>0.027 ± 0.001</td>
<td>0.027</td>
<td>0.000</td>
<td>0.0</td>
</tr>
<tr>
<td>AA</td>
<td>0.043 ± 0.002</td>
<td>0.047</td>
<td>−0.004</td>
<td>8.5</td>
</tr>
<tr>
<td><strong>PCa diagnosis rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whites</td>
<td>0.185 ± 0.003</td>
<td>0.159</td>
<td>+0.026</td>
<td>16.4</td>
</tr>
<tr>
<td>AA</td>
<td>0.292 ± 0.004</td>
<td>0.191</td>
<td>+0.101</td>
<td>52.9</td>
</tr>
<tr>
<td><strong>Average life expectancy (y)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whites</td>
<td>37.10 ± 0.11</td>
<td>37.7</td>
<td>−0.60</td>
<td>1.6</td>
</tr>
<tr>
<td>AA</td>
<td>32.84 ± 0.12</td>
<td>33.2</td>
<td>−0.01</td>
<td>1.1</td>
</tr>
<tr>
<td><strong>Average mPCa survival time (mo)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whites</td>
<td>4.0 ± 0.1</td>
<td>3.5</td>
<td>+0.5</td>
<td>14.3</td>
</tr>
<tr>
<td>AA</td>
<td>3.5 ± 0.1</td>
<td>3.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Validation of NNS Estimation

For prong (ii), validation is made especially challenging due to the paucity of reliable RCT-based NNS estimates and to the variability in the available RCT-based estimates. As previously mentioned, one of the largest RCTs to study PCa screening was the ERSPC. This trial comprised seven major study centers, referred to in Schröder et al. [95] by corresponding region as the following: Netherlands, Belgium, Sweden, Finland, Italy, Spain, and Switzerland. To validate our NNS estimation, we simulated the trial at each center using statistical data on each center’s participant population. Table 3.11 lists the data for each center used in our simulation of the trial.
Table 3.11: Description of patient and screening characteristics used to simulate each of the 7 major study centers in the ERSPC trial. These quantities are taken from Schröder et al. [95].

<table>
<thead>
<tr>
<th></th>
<th>Netherlands</th>
<th>Belgium</th>
<th>Sweden</th>
<th>Finland</th>
<th>Italy</th>
<th>Spain</th>
<th>Switzerland</th>
</tr>
</thead>
<tbody>
<tr>
<td># of patients</td>
<td>34,833</td>
<td>8,562</td>
<td>11,852</td>
<td>80,379</td>
<td>14,517</td>
<td>2,197</td>
<td>9,903</td>
</tr>
<tr>
<td># of patients in control group</td>
<td>17,443</td>
<td>4,307</td>
<td>5,901</td>
<td>31,970</td>
<td>7,265</td>
<td>1,056</td>
<td>4,948</td>
</tr>
<tr>
<td># of patients in screen group</td>
<td>17,390</td>
<td>4,255</td>
<td>5,951</td>
<td>48,409</td>
<td>7,252</td>
<td>1,141</td>
<td>4,955</td>
</tr>
<tr>
<td>Avg. age at randomization (yr)</td>
<td>62</td>
<td>63</td>
<td>60</td>
<td>60</td>
<td>62</td>
<td>61</td>
<td>62</td>
</tr>
<tr>
<td>PSA threshold (ng/mL)</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>4.0</td>
<td>4.0</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Screening frequency (yr)</td>
<td>4</td>
<td>4</td>
<td>3.0</td>
<td>4.0</td>
<td>4.0</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Participation rate</td>
<td>0.946</td>
<td>0.900</td>
<td>0.757</td>
<td>0.738</td>
<td>0.781</td>
<td>1.000</td>
<td>0.958</td>
</tr>
</tbody>
</table>

The total number of patients simulated for each center are shown in Figure 3.6. The number of patients simulated for each center is equal to the minimum of the numbers in the corresponding screen and control groups as listed in Table 3.11.

Figure 3.6: Number of patients simulated for each ERSPC study center. These values are equal to the minimums of the sizes of the screen and control groups reported for each center.

We obtained simulation estimates of NNS for each of the study centers. For each center, we simulated the minimum of the control- and screen-group sizes listed in Table 3.11. For the screened group in each center, we began screening at the average age at randomization listed in Table 3.11 and then screened at the PSA threshold and screening frequency listed in the same table. For each center, we estimated NNS
at patient follow-up periods of 9, 11, 13, and 15 years, and also with infinite follow-up.

Reports on mortality results from the ERSPC at 9, 11, and 13 years of follow-up [95, 96, 97] present the following estimates of the number needed to invite (NNI) for screening in order to avert 1 PCa death, respectively: 1410, 1055, and 781. NNI is different from NNS in that NNI assumes not everyone invited to participate in screening will in fact end up being screened and some patients may elect to terminate screening prior to the conclusion of the prevailing screening program, whereas NNS, as we are computing it, assumes complete participation and perfect adherence. To compare our results with the reported ERSPC NNI values, we used the reported percentage of patients in each study center who were screened at least once to model imperfect participation. For a given center, we used the reported percentage (divided by 100) as the probability that a given simulated patient will actually undergo screening. This probability is referred to as the participation rate in Table 3.11. Put simply, we used the complement of the participation rate to sample a subpopulation of the corresponding trial center’s population not to undergo any screening. By simulating with imperfect participation and computing estimates of NNS, we are really computing estimates of the related statistic NNI. These NNI estimates for the finite follow-up periods are plotted in Figure 3.7.
Figure 3.7: Simulation estimates of NNI in each ERSPC trial study center.

To illustrate the uncertainty around these NNI estimates, both due to the lengths of the follow-up periods and the simulated sample sizes, the NNI estimates from Figure 3.7 are reproduced separately by center with corresponding 95% CIs in Figures 3.8 through Figure 3.14.
Figure 3.8: Simulation estimates of NNI in the Netherlands study center of the ERSPC trial.

Figure 3.9: Simulation estimates of NNI in the Belgium study center of the ERSPC trial.
Figure 3.10: Simulation estimates of NNI in the Sweden study center of the ERSPC trial.

Figure 3.11: Simulation estimates of NNI in the Finland study center of the ERSPC trial.
Figure 3.12: Simulation estimates of NNI in the Italy study center of the ERSPC trial.

Figure 3.13: Simulation estimates of NNI in the Spain study center of the ERSPC trial.
Figure 3.14: Simulation estimates of NNI in the Switzerland study center of the ERSPC trial.

The CIs depicted for the Spain study center are unusually wide because of the relatively small size of the simulated population. A report from the Spanish arm of the ERSPC notes this limited sample size as an “obvious limitation” relative to the other centers [72].

In order to more meaningfully compare the NNI results from Figure 3.7 to the NNI values reported by the ERSPC, we aggregated all of the simulated study-center populations and estimated NNI for the aggregate population. The aggregate NNI estimates with 95% CIs at 9-, 11-, and 13-year follow-up periods are, respectively: 1025 [767, 1283]; 674 [540, 808]; and 531 [437, 625]. The ERSPC-reported values of NNI with 95% CIs at 9-, 11-, and 13-year follow-up periods are, respectively: 1410 [1142, 1721]; 1055 [645, 2894]; and 781 [490, 1929]. These ERSPC-reported point and CI estimates of NNI are given in Schröder et al. [95], Schröder et al. [97], and the supplementary appendix to Schröder et al. [96]. The simulated and ERSPC-reported NNI estimates are compared in Figure 3.15.

There are reasons to anticipate differences between our NNI estimates and those reported by the ERSPC. First, the ERSPC is based on European populations of patients, whereas our reference risk group is largely representative of US males. Second, our simulation of the ERSPC study centers relied upon broad statistical descriptions of each center’s population, and therefore can yield at best rough point and
Figure 3.15: Comparison of the aggregated NNI estimates with the NNI values reported by the ERSPC trial.

CI estimates of NNI for each study center or for the overall study. Nevertheless, Figure 3.15 shows there are no statistically significant differences between our model’s basecase results and the results of the ERSPC trial.

3.3.3 Numerical Results Using Base-case Parameter Values

The strategy of screening annually from age 50 to 75 using the PSA threshold 2.5 ng/mL (strategy #6) yielded the smallest estimated expected NNS for each of the risk groups. Figure 3.16 depicts these estimates by risk group. Table 3.12 contains estimates of the expected NNSs for strategies #2–#14 for each risk group. The confidence intervals in Figure 3.16 and Table 3.12 reflect the random variation between patients in the simulation model; they do not in any way reflect the real uncertainty in the population expected NNS induced by the variability in any of the model parameters. The risk groups to which the results variously pertain are the following:

(i) **Whites** – white US males with no family history of PCa and CCI=0.

(ii) **African Americans** – African American males with no family history of PCa and CCI=0.
Figure 3.16: The Estimated NNS to Avert 1 PCa Death for Each Risk Group (i)–(v). Also displayed are 95% CIs for expected NNS in each risk group based on a sample of 1 million patients. The NNS estimates and associated 95% CIs for risk groups (i)–(v) were, respectively as follows: 137 [134, 140]; 63 [62, 64]; 111 [109, 113]; 178 [173, 183]; 225 [218, 232].

(iii) **Family History** – white US males with a family history of PCa and CCI=0.

(iv) **CCI=1** – white US males with no family history of PCa and CCI=1.

(v) **CCI≥2** – white US males with no family history of PCa and CCI≥2.

Table 3.13 contains a selection of clinical statistics estimated for each of risk groups (i)–(v) by simulating 100,000 patients from the respective risk groups screened using strategy #6. In these tables, the number of screens is an estimate of the total lifetime number of PSA tests per 100,000 patients. The number of biopsies is an estimate of the number of prostate biopsies per 100,000 patients. The number of treatments is an estimate of the number of screen-detected PCa treatments per 100,000 patients. The number of PCa deaths averted is an estimate of the number of PCa-specific deaths averted by screening per 100,000 patients, relative to no screening. The number of overdiagnoses is an estimate of the number of patients treated for which neither clinical detection nor metastasis would have occurred in the absence of screening. The mean months saved is an estimate of the mean increase in life expectancy across all of
Table 3.12: Estimates of the expected NNS for screening in each of the compared risk groups using the screening strategies #1–#14; 95% confidence interval (CI) half-widths are shown in parentheses. Recall that strategy #1 is no screening.

<table>
<thead>
<tr>
<th></th>
<th>Whites</th>
<th>African Americans</th>
<th>Family History</th>
<th>CCI=1</th>
<th>CCI≥ 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>#2</td>
<td>230 (8)</td>
<td>100 (2)</td>
<td>187 (5)</td>
<td>298 (11)</td>
<td>371 (15)</td>
</tr>
<tr>
<td>#3</td>
<td>196 (6)</td>
<td>86 (2)</td>
<td>158 (4)</td>
<td>248 (8)</td>
<td>305 (11)</td>
</tr>
<tr>
<td>#4</td>
<td>204 (6)</td>
<td>88 (2)</td>
<td>164 (4)</td>
<td>253 (9)</td>
<td>303 (11)</td>
</tr>
<tr>
<td>#5</td>
<td>173 (5)</td>
<td>78 (1)</td>
<td>140 (4)</td>
<td>222 (7)</td>
<td>276 (10)</td>
</tr>
<tr>
<td>#6</td>
<td>137 (3)</td>
<td>63 (1)</td>
<td>111 (2)</td>
<td>178 (5)</td>
<td>225 (7)</td>
</tr>
<tr>
<td>#7</td>
<td>196 (6)</td>
<td>86 (2)</td>
<td>158 (4)</td>
<td>248 (8)</td>
<td>305 (11)</td>
</tr>
<tr>
<td>#8</td>
<td>173 (5)</td>
<td>78 (1)</td>
<td>140 (4)</td>
<td>222 (7)</td>
<td>276 (10)</td>
</tr>
<tr>
<td>#9</td>
<td>200 (6)</td>
<td>83 (2)</td>
<td>160 (4)</td>
<td>243 (8)</td>
<td>291 (10)</td>
</tr>
<tr>
<td>#10</td>
<td>159 (4)</td>
<td>72 (1)</td>
<td>128 (3)</td>
<td>203 (6)</td>
<td>254 (8)</td>
</tr>
<tr>
<td>#11</td>
<td>271 (10)</td>
<td>109 (2)</td>
<td>217 (7)</td>
<td>330 (13)</td>
<td>386 (16)</td>
</tr>
<tr>
<td>#12</td>
<td>1267 (99)</td>
<td>350 (14)</td>
<td>965 (65)</td>
<td>1332 (106)</td>
<td>1357 (109)</td>
</tr>
<tr>
<td>#13</td>
<td>720 (43)</td>
<td>260 (9)</td>
<td>581 (31)</td>
<td>805 (50)</td>
<td>880 (57)</td>
</tr>
<tr>
<td>#14</td>
<td>571 (30)</td>
<td>234 (8)</td>
<td>468 (22)</td>
<td>729 (42)</td>
<td>907 (58)</td>
</tr>
</tbody>
</table>

the simulated patients whose life expectancies were increased relative to no screening. The estimated clinical statistics across all of strategies #2–#14 are provided in Tables 3.14 through 3.18.
Table 3.13: Estimates of Clinical Statistics from Screening with Strategy #6 in Each of The Compared Risk Groups (i)–(v). Results are based on samples of 100,000 simulated patients from the corresponding risk groups. For mean months of life saved, the 95% CIs are reported.

<table>
<thead>
<tr>
<th></th>
<th>(i) Whites</th>
<th>(ii) African Americans</th>
<th>(iii) Family History</th>
<th>(iv) CCI=1</th>
<th>(v) CCI≥2</th>
</tr>
</thead>
<tbody>
<tr>
<td># PSA screens</td>
<td>2,035,796</td>
<td>1,635,368</td>
<td>1,996,171</td>
<td>1,953,910</td>
<td>1,898,023</td>
</tr>
<tr>
<td># prostate biopses</td>
<td>321,923</td>
<td>437,227</td>
<td>352,442</td>
<td>293,351</td>
<td>273,609</td>
</tr>
<tr>
<td># prostatectomies</td>
<td>16,583</td>
<td>28,221</td>
<td>19,961</td>
<td>15,275</td>
<td>14,277</td>
</tr>
<tr>
<td># PCa deaths averted</td>
<td>755</td>
<td>1,671</td>
<td>959</td>
<td>595</td>
<td>475</td>
</tr>
<tr>
<td>% overdiagnosis</td>
<td>43.3</td>
<td>50.8</td>
<td>43.2</td>
<td>54.1</td>
<td>61.2</td>
</tr>
<tr>
<td>Mean mo. of life saved</td>
<td>130 ± 6</td>
<td>132 ± 5</td>
<td>132 ± 6</td>
<td>113 ± 7</td>
<td>96 ± 7</td>
</tr>
</tbody>
</table>
Table 3.14: Clinical statistics for the whites risk group based on samples of 100,000 simulated patients under the corresponding screening strategies. CI HW means the 95% confidence-interval half width. Recall that strategy #1 is no screening.

<table>
<thead>
<tr>
<th>Strategy</th>
<th># screens</th>
<th># biopsies</th>
<th># treatments</th>
<th># PCa deaths averted</th>
<th>% overdiagnosis</th>
<th>Mean mo. life saved</th>
<th>Mean mo. life saved (CI HW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>#2</td>
<td>682307</td>
<td>34465</td>
<td>12307</td>
<td>437</td>
<td>32.3</td>
<td>127</td>
<td>9</td>
</tr>
<tr>
<td>#3</td>
<td>1051035</td>
<td>56201</td>
<td>13053</td>
<td>524</td>
<td>34.8</td>
<td>134</td>
<td>8</td>
</tr>
<tr>
<td>#4</td>
<td>2067909</td>
<td>83268</td>
<td>12781</td>
<td>519</td>
<td>34.6</td>
<td>137</td>
<td>8</td>
</tr>
<tr>
<td>#5</td>
<td>2061530</td>
<td>100261</td>
<td>14401</td>
<td>597</td>
<td>37.9</td>
<td>131</td>
<td>7</td>
</tr>
<tr>
<td>#6</td>
<td>2035796</td>
<td>166085</td>
<td>16583</td>
<td>755</td>
<td>43.3</td>
<td>130</td>
<td>6</td>
</tr>
<tr>
<td>#7</td>
<td>1249629</td>
<td>56212</td>
<td>13055</td>
<td>524</td>
<td>34.8</td>
<td>134</td>
<td>8</td>
</tr>
<tr>
<td>#8</td>
<td>2260125</td>
<td>100273</td>
<td>14403</td>
<td>597</td>
<td>37.9</td>
<td>131</td>
<td>7</td>
</tr>
<tr>
<td>#9</td>
<td>1221584</td>
<td>82707</td>
<td>11081</td>
<td>536</td>
<td>33.0</td>
<td>138</td>
<td>8</td>
</tr>
<tr>
<td>#10</td>
<td>1523761</td>
<td>124078</td>
<td>14923</td>
<td>666</td>
<td>39.5</td>
<td>128</td>
<td>7</td>
</tr>
<tr>
<td>#11</td>
<td>339991</td>
<td>27591</td>
<td>8464</td>
<td>377</td>
<td>27.6</td>
<td>146</td>
<td>9</td>
</tr>
<tr>
<td>#12</td>
<td>93728</td>
<td>4293</td>
<td>1954</td>
<td>76</td>
<td>19.5</td>
<td>172</td>
<td>23</td>
</tr>
<tr>
<td>#13</td>
<td>89251</td>
<td>8291</td>
<td>3496</td>
<td>155</td>
<td>20.7</td>
<td>158</td>
<td>15</td>
</tr>
<tr>
<td>#14</td>
<td>82178</td>
<td>12034</td>
<td>4806</td>
<td>177</td>
<td>24.1</td>
<td>124</td>
<td>11</td>
</tr>
</tbody>
</table>
Table 3.15: Clinical statistics for the African Americans risk group based on samples of 100,000 simulated patients under the corresponding screening strategies. CI HW means the 95% confidence-interval half width. Recall that strategy #1 is no screening.

<table>
<thead>
<tr>
<th>Strategy</th>
<th># screens</th>
<th># biopsies</th>
<th># treatments</th>
<th># PCa deaths averted</th>
<th>% overdiagnosis</th>
<th>Mean mo. life saved</th>
<th>Mean mo. life saved (CI HW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>#2</td>
<td>605565</td>
<td>38651</td>
<td>21407</td>
<td>1054</td>
<td>38.6</td>
<td>137</td>
<td>6</td>
</tr>
<tr>
<td>#3</td>
<td>873315</td>
<td>54113</td>
<td>23087</td>
<td>1250</td>
<td>42.4</td>
<td>139</td>
<td>5</td>
</tr>
<tr>
<td>#4</td>
<td>1689585</td>
<td>81845</td>
<td>22773</td>
<td>1201</td>
<td>42.1</td>
<td>143</td>
<td>5</td>
</tr>
<tr>
<td>#5</td>
<td>1683077</td>
<td>82165</td>
<td>24864</td>
<td>1376</td>
<td>45.2</td>
<td>137</td>
<td>5</td>
</tr>
<tr>
<td>#6</td>
<td>1635368</td>
<td>127461</td>
<td>28221</td>
<td>1671</td>
<td>50.8</td>
<td>132</td>
<td>5</td>
</tr>
<tr>
<td>#7</td>
<td>1070654</td>
<td>54152</td>
<td>23099</td>
<td>1250</td>
<td>42.4</td>
<td>139</td>
<td>5</td>
</tr>
<tr>
<td>#8</td>
<td>1880413</td>
<td>82201</td>
<td>24875</td>
<td>1376</td>
<td>45.2</td>
<td>137</td>
<td>5</td>
</tr>
<tr>
<td>#9</td>
<td>989534</td>
<td>72864</td>
<td>21491</td>
<td>1270</td>
<td>42.3</td>
<td>140</td>
<td>5</td>
</tr>
<tr>
<td>#10</td>
<td>1186169</td>
<td>98387</td>
<td>25703</td>
<td>1487</td>
<td>47.0</td>
<td>133</td>
<td>5</td>
</tr>
<tr>
<td>#11</td>
<td>290212</td>
<td>33603</td>
<td>17470</td>
<td>981</td>
<td>36.9</td>
<td>146</td>
<td>6</td>
</tr>
<tr>
<td>#12</td>
<td>88890</td>
<td>8758</td>
<td>5828</td>
<td>304</td>
<td>26.7</td>
<td>171</td>
<td>12</td>
</tr>
<tr>
<td>#13</td>
<td>80729</td>
<td>13621</td>
<td>8481</td>
<td>424</td>
<td>28.1</td>
<td>151</td>
<td>9</td>
</tr>
<tr>
<td>#14</td>
<td>69017</td>
<td>15763</td>
<td>9296</td>
<td>460</td>
<td>32.3</td>
<td>126</td>
<td>8</td>
</tr>
</tbody>
</table>
Table 3.16: Clinical statistics for the Family History risk group based on samples of 100,000 simulated patients under the corresponding screening strategies. CI HW means the 95% confidence-interval half width. Recall that strategy #1 is no screening.

<table>
<thead>
<tr>
<th>Strategy</th>
<th># screens</th>
<th># biopsies</th>
<th># treatments</th>
<th># PCa deaths averted</th>
<th>% overdiagnosis</th>
<th>Mean mo. life saved</th>
<th>Mean mo. life saved (CI HW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>#2</td>
<td>677168</td>
<td>36400</td>
<td>14847</td>
<td>560</td>
<td>32.0</td>
<td>131</td>
<td>8</td>
</tr>
<tr>
<td>#3</td>
<td>1036769</td>
<td>57150</td>
<td>15791</td>
<td>656</td>
<td>34.7</td>
<td>137</td>
<td>7</td>
</tr>
<tr>
<td>#4</td>
<td>2035395</td>
<td>82518</td>
<td>15437</td>
<td>653</td>
<td>34.4</td>
<td>140</td>
<td>7</td>
</tr>
<tr>
<td>#5</td>
<td>2027867</td>
<td>98557</td>
<td>17383</td>
<td>756</td>
<td>37.7</td>
<td>134</td>
<td>7</td>
</tr>
<tr>
<td>#6</td>
<td>1996171</td>
<td>160701</td>
<td>19961</td>
<td>959</td>
<td>43.2</td>
<td>132</td>
<td>6</td>
</tr>
<tr>
<td>#7</td>
<td>1235343</td>
<td>57163</td>
<td>15793</td>
<td>656</td>
<td>34.7</td>
<td>137</td>
<td>7</td>
</tr>
<tr>
<td>#8</td>
<td>2226441</td>
<td>98570</td>
<td>17385</td>
<td>756</td>
<td>37.7</td>
<td>134</td>
<td>7</td>
</tr>
<tr>
<td>#9</td>
<td>1201635</td>
<td>82105</td>
<td>13556</td>
<td>677</td>
<td>33.2</td>
<td>141</td>
<td>7</td>
</tr>
<tr>
<td>#10</td>
<td>1491246</td>
<td>121075</td>
<td>18049</td>
<td>838</td>
<td>39.4</td>
<td>131</td>
<td>6</td>
</tr>
<tr>
<td>#11</td>
<td>336955</td>
<td>29229</td>
<td>10426</td>
<td>475</td>
<td>27.9</td>
<td>147</td>
<td>9</td>
</tr>
<tr>
<td>#12</td>
<td>93708</td>
<td>4889</td>
<td>2465</td>
<td>110</td>
<td>19.6</td>
<td>178</td>
<td>20</td>
</tr>
<tr>
<td>#13</td>
<td>89093</td>
<td>9236</td>
<td>4358</td>
<td>190</td>
<td>20.7</td>
<td>158</td>
<td>13</td>
</tr>
<tr>
<td>#14</td>
<td>81690</td>
<td>13042</td>
<td>5864</td>
<td>214</td>
<td>24.1</td>
<td>123</td>
<td>10</td>
</tr>
</tbody>
</table>

73
Table 3.17: Clinical statistics for the CCI=1 risk group based on samples of 100,000 simulated patients under the corresponding screening strategies. CI HW means the 95% confidence-interval half width. Recall that strategy #1 is no screening.

<table>
<thead>
<tr>
<th>Strategy</th>
<th># screens</th>
<th># biopsies</th>
<th># treatments</th>
<th># PCa deaths averted</th>
<th>% overdiagnosis</th>
<th>Mean mo. life saved</th>
<th>Mean mo. life saved (CI HW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>#2</td>
<td>657204</td>
<td>30515</td>
<td>10946</td>
<td>350</td>
<td>41.9</td>
<td>116</td>
<td>9</td>
</tr>
<tr>
<td>#3</td>
<td>1012905</td>
<td>51722</td>
<td>11994</td>
<td>427</td>
<td>45.4</td>
<td>119</td>
<td>8</td>
</tr>
<tr>
<td>#4</td>
<td>1982761</td>
<td>77209</td>
<td>11815</td>
<td>425</td>
<td>45.2</td>
<td>121</td>
<td>8</td>
</tr>
<tr>
<td>#5</td>
<td>1977630</td>
<td>91191</td>
<td>13149</td>
<td>477</td>
<td>45.6</td>
<td>116</td>
<td>8</td>
</tr>
<tr>
<td>#6</td>
<td>1953910</td>
<td>152776</td>
<td>15275</td>
<td>595</td>
<td>54.1</td>
<td>113</td>
<td>7</td>
</tr>
<tr>
<td>#7</td>
<td>1211499</td>
<td>51733</td>
<td>11996</td>
<td>427</td>
<td>45.4</td>
<td>119</td>
<td>8</td>
</tr>
<tr>
<td>#8</td>
<td>2176225</td>
<td>91203</td>
<td>13151</td>
<td>477</td>
<td>45.6</td>
<td>116</td>
<td>8</td>
</tr>
<tr>
<td>#9</td>
<td>1209978</td>
<td>81285</td>
<td>10815</td>
<td>456</td>
<td>45.3</td>
<td>117</td>
<td>8</td>
</tr>
<tr>
<td>#10</td>
<td>1456851</td>
<td>114993</td>
<td>13892</td>
<td>531</td>
<td>51.0</td>
<td>112</td>
<td>7</td>
</tr>
<tr>
<td>#11</td>
<td>337951</td>
<td>27287</td>
<td>8324</td>
<td>330</td>
<td>38.8</td>
<td>120</td>
<td>9</td>
</tr>
<tr>
<td>#12</td>
<td>93728</td>
<td>4293</td>
<td>1954</td>
<td>76</td>
<td>20.3</td>
<td>136</td>
<td>23</td>
</tr>
<tr>
<td>#13</td>
<td>89251</td>
<td>8291</td>
<td>3496</td>
<td>142</td>
<td>27.6</td>
<td>130</td>
<td>14</td>
</tr>
<tr>
<td>#14</td>
<td>82178</td>
<td>12034</td>
<td>4806</td>
<td>148</td>
<td>38.8</td>
<td>108</td>
<td>11</td>
</tr>
</tbody>
</table>
Table 3.18: Clinical statistics for the CCI≥2 risk group based on samples of 100,000 simulated patients under the corresponding screening strategies. CI HW means the 95% confidence-interval half width. Recall that strategy #1 is no screening.

<table>
<thead>
<tr>
<th>Strategy</th>
<th># screens</th>
<th># biopsies</th>
<th># treatments</th>
<th># PCa deaths averted</th>
<th>% overdiagnosis</th>
<th>Mean mo. life saved</th>
<th>Mean mo. life saved (CI HW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>#2</td>
<td>640732</td>
<td>27956</td>
<td>10032</td>
<td>274</td>
<td>48.5</td>
<td>94</td>
<td>9</td>
</tr>
<tr>
<td>#3</td>
<td>986403</td>
<td>48592</td>
<td>11216</td>
<td>341</td>
<td>52.5</td>
<td>96</td>
<td>8</td>
</tr>
<tr>
<td>#4</td>
<td>1924394</td>
<td>73038</td>
<td>11127</td>
<td>343</td>
<td>52.7</td>
<td>98</td>
<td>8</td>
</tr>
<tr>
<td>#5</td>
<td>1920219</td>
<td>85086</td>
<td>12263</td>
<td>379</td>
<td>55.6</td>
<td>95</td>
<td>7</td>
</tr>
<tr>
<td>#6</td>
<td>1898023</td>
<td>143977</td>
<td>14277</td>
<td>475</td>
<td>61.2</td>
<td>96</td>
<td>7</td>
</tr>
<tr>
<td>#7</td>
<td>1184997</td>
<td>48603</td>
<td>11218</td>
<td>341</td>
<td>52.5</td>
<td>96</td>
<td>8</td>
</tr>
<tr>
<td>#8</td>
<td>2118814</td>
<td>85098</td>
<td>12265</td>
<td>379</td>
<td>55.7</td>
<td>95</td>
<td>7</td>
</tr>
<tr>
<td>#9</td>
<td>1201033</td>
<td>80301</td>
<td>10643</td>
<td>362</td>
<td>53.5</td>
<td>96</td>
<td>7</td>
</tr>
<tr>
<td>#10</td>
<td>1409941</td>
<td>108779</td>
<td>13083</td>
<td>420</td>
<td>58.3</td>
<td>93</td>
<td>7</td>
</tr>
<tr>
<td>#11</td>
<td>336338</td>
<td>27066</td>
<td>8200</td>
<td>261</td>
<td>46.1</td>
<td>97</td>
<td>9</td>
</tr>
<tr>
<td>#12</td>
<td>93728</td>
<td>4293</td>
<td>1954</td>
<td>72</td>
<td>21.3</td>
<td>120</td>
<td>20</td>
</tr>
<tr>
<td>#13</td>
<td>89251</td>
<td>8291</td>
<td>3496</td>
<td>117</td>
<td>31.7</td>
<td>112</td>
<td>14</td>
</tr>
<tr>
<td>#14</td>
<td>82178</td>
<td>12034</td>
<td>4806</td>
<td>104</td>
<td>48.1</td>
<td>78</td>
<td>11</td>
</tr>
</tbody>
</table>
3.3.4 Sensitivity Analysis

We performed one-way sensitivity analysis on the following model parameters: (a) the annual probability of other-cause death; (b) the annual probability of developing PCa; (c) the annual metastasis probability for patients in state C; and (d) the PCa-screening leadtime. These parameters were varied from their base-case levels to plausible low and high levels. We varied the annual probabilities of other-cause death by ±20%. For the low and high levels of the annual probability of developing PCa, we used the lower and upper bounds of the 95% confidence interval provided in the source for the base-case estimates [44]. Similarly, for the low and high levels of the annual metastasis probability for patients in state C, we used the lower and upper bounds of the 95% confidence interval provided in the source for the base-case estimates [62]. We varied the PCa-screening leadtime by ±2 years from the base-case value of 12.8 years.

At each experimental level, we estimated the expected NNS for each of strategies #2–#12 across all compared risk groups using samples of 1 million patients from the corresponding risk groups. The NNS estimates for each risk group at every experimental level in the sensitivity analysis are provided in Tables 3.19 through 3.22. The strategy of screening annually from age 50 to 75 using the PSA threshold 2.5 ng/mL (strategy #6) yielded the smallest estimated expected NNS at every experimental level. That is, the observation of strategy #6 yielding the smallest estimated expected NNS was robust to variation in the four model parameters in the sensitivity analysis.
Table 3.19: NNS estimates from sensitivity analysis over the annual probability of other-cause death ($d_t$). For each risk group, the headings “low” and “high” respectively denote the low and high levels of parameter $d_t$. The 95% CI estimates are based on simulated samples of 1 million patients.

<table>
<thead>
<tr>
<th></th>
<th>Whites</th>
<th>African American</th>
<th>Family History</th>
<th>CCI=1</th>
<th>CCI≥2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>#2</td>
<td>207 ± 6</td>
<td>255 ± 9</td>
<td>88 ± 2</td>
<td>116 ± 3</td>
<td>167 ± 5</td>
</tr>
<tr>
<td>#3</td>
<td>178 ± 5</td>
<td>212 ± 7</td>
<td>76 ± 1</td>
<td>99 ± 2</td>
<td>143 ± 4</td>
</tr>
<tr>
<td>#4</td>
<td>186 ± 5</td>
<td>220 ± 7</td>
<td>79 ± 1</td>
<td>101 ± 2</td>
<td>149 ± 4</td>
</tr>
<tr>
<td>#5</td>
<td>157 ± 4</td>
<td>189 ± 5</td>
<td>68 ± 1</td>
<td>90 ± 2</td>
<td>126 ± 3</td>
</tr>
<tr>
<td>#6</td>
<td>122 ± 3</td>
<td>150 ± 4</td>
<td>55 ± 1</td>
<td>73 ± 1</td>
<td>98 ± 2</td>
</tr>
<tr>
<td>#7</td>
<td>178 ± 5</td>
<td>212 ± 7</td>
<td>76 ± 1</td>
<td>99 ± 2</td>
<td>143 ± 4</td>
</tr>
<tr>
<td>#8</td>
<td>157 ± 4</td>
<td>189 ± 5</td>
<td>68 ± 1</td>
<td>90 ± 2</td>
<td>126 ± 3</td>
</tr>
<tr>
<td>#9</td>
<td>185 ± 5</td>
<td>214 ± 7</td>
<td>74 ± 1</td>
<td>95 ± 2</td>
<td>149 ± 4</td>
</tr>
<tr>
<td>#10</td>
<td>142 ± 4</td>
<td>173 ± 5</td>
<td>63 ± 1</td>
<td>83 ± 2</td>
<td>115 ± 3</td>
</tr>
<tr>
<td>#11</td>
<td>257 ± 9</td>
<td>293 ± 11</td>
<td>97 ± 2</td>
<td>123 ± 3</td>
<td>206 ± 6</td>
</tr>
<tr>
<td>#12</td>
<td>1170 ± 86</td>
<td>1272 ± 99</td>
<td>328 ± 13</td>
<td>371 ± 15</td>
<td>924 ± 61</td>
</tr>
<tr>
<td>#13</td>
<td>689 ± 40</td>
<td>735 ± 43</td>
<td>234 ± 8</td>
<td>291 ± 11</td>
<td>559 ± 29</td>
</tr>
<tr>
<td>#14</td>
<td>519 ± 26</td>
<td>638 ± 35</td>
<td>210 ± 6</td>
<td>279 ± 10</td>
<td>419 ± 19</td>
</tr>
</tbody>
</table>
Table 3.20: NNS estimates from sensitivity analysis over the annual probability of developing PCa ($w_t$). For each risk group, the headings “low” and “high” respectively denote the low and high levels of parameter $w_t$. The 95% CI estimates are based on simulated samples of 1 million patients.

<table>
<thead>
<tr>
<th></th>
<th>Whites</th>
<th>African American</th>
<th>Family History</th>
<th>CCI=1</th>
<th>CCI≥2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>#2</td>
<td>288 ± 11</td>
<td>229 ± 8</td>
<td>126 ± 3</td>
<td>102 ± 2</td>
<td>231 ± 8</td>
</tr>
<tr>
<td>#3</td>
<td>244 ± 8</td>
<td>198 ± 6</td>
<td>108 ± 2</td>
<td>87 ± 2</td>
<td>198 ± 6</td>
</tr>
<tr>
<td>#4</td>
<td>257 ± 9</td>
<td>205 ± 6</td>
<td>113 ± 3</td>
<td>90 ± 2</td>
<td>207 ± 6</td>
</tr>
<tr>
<td>#5</td>
<td>217 ± 7</td>
<td>175 ± 5</td>
<td>98 ± 2</td>
<td>79 ± 1</td>
<td>175 ± 5</td>
</tr>
<tr>
<td>#6</td>
<td>173 ± 5</td>
<td>138 ± 3</td>
<td>79 ± 1</td>
<td>64 ± 1</td>
<td>139 ± 3</td>
</tr>
<tr>
<td>#7</td>
<td>244 ± 8</td>
<td>198 ± 6</td>
<td>108 ± 2</td>
<td>87 ± 2</td>
<td>198 ± 6</td>
</tr>
<tr>
<td>#8</td>
<td>217 ± 7</td>
<td>175 ± 5</td>
<td>98 ± 2</td>
<td>79 ± 1</td>
<td>175 ± 5</td>
</tr>
<tr>
<td>#9</td>
<td>253 ± 9</td>
<td>204 ± 6</td>
<td>106 ± 2</td>
<td>85 ± 2</td>
<td>204 ± 6</td>
</tr>
<tr>
<td>#10</td>
<td>199 ± 6</td>
<td>160 ± 4</td>
<td>90 ± 2</td>
<td>73 ± 1</td>
<td>160 ± 4</td>
</tr>
<tr>
<td>#11</td>
<td>344 ± 14</td>
<td>278 ± 10</td>
<td>139 ± 3</td>
<td>111 ± 2</td>
<td>278 ± 10</td>
</tr>
<tr>
<td>#12</td>
<td>1724 ± 155</td>
<td>1271 ± 99</td>
<td>498 ± 24</td>
<td>362 ± 15</td>
<td>1362 ± 109</td>
</tr>
<tr>
<td>#13</td>
<td>928 ± 62</td>
<td>724 ± 43</td>
<td>339 ± 13</td>
<td>266 ± 9</td>
<td>741 ± 45</td>
</tr>
<tr>
<td>#14</td>
<td>712 ± 42</td>
<td>591 ± 32</td>
<td>282 ± 10</td>
<td>241 ± 8</td>
<td>574 ± 30</td>
</tr>
</tbody>
</table>

78
Table 3.21: NNS estimates from sensitivity analysis over the annual metastasis probability for patients in state C ($e_t$). For each risk group, the headings “low” and “high” respectively denote the low and high levels of parameter $e_t$. The 95% CI estimates are based on simulated samples of 1 million patients.

<table>
<thead>
<tr>
<th></th>
<th>Whites</th>
<th></th>
<th>African American</th>
<th></th>
<th>Family History</th>
<th></th>
<th>CCI=1</th>
<th></th>
<th>CCI≥2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>#2</td>
<td>1161 ±110</td>
<td>96 ±2</td>
<td>415 ±21</td>
<td>44 ±1</td>
<td>951 ±82</td>
<td>78 ±1</td>
<td>1414 ±240</td>
<td>124 ±3</td>
<td>1972 ±240</td>
</tr>
<tr>
<td>#3</td>
<td>971 ±81</td>
<td>82 ±2</td>
<td>352 ±16</td>
<td>38 ±0</td>
<td>789 ±60</td>
<td>66 ±1</td>
<td>1185 ±164</td>
<td>104 ±2</td>
<td>1575 ±164</td>
</tr>
<tr>
<td>#4</td>
<td>1046 ±92</td>
<td>84 ±2</td>
<td>378 ±18</td>
<td>39 ±0</td>
<td>849 ±67</td>
<td>68 ±1</td>
<td>1215 ±172</td>
<td>105 ±2</td>
<td>1621 ±172</td>
</tr>
<tr>
<td>#5</td>
<td>838 ±64</td>
<td>73 ±1</td>
<td>313 ±13</td>
<td>35 ±0</td>
<td>678 ±46</td>
<td>59 ±1</td>
<td>1028 ±132</td>
<td>95 ±2</td>
<td>1385 ±132</td>
</tr>
<tr>
<td>#6</td>
<td>618 ±38</td>
<td>59 ±1</td>
<td>243 ±9</td>
<td>29 ±0</td>
<td>497 ±27</td>
<td>48 ±1</td>
<td>782 ±84</td>
<td>77 ±1</td>
<td>1060 ±84</td>
</tr>
<tr>
<td>#7</td>
<td>971 ±81</td>
<td>82 ±2</td>
<td>352 ±16</td>
<td>38 ±0</td>
<td>789 ±60</td>
<td>66 ±1</td>
<td>1185 ±164</td>
<td>104 ±2</td>
<td>1575 ±164</td>
</tr>
<tr>
<td>#8</td>
<td>838 ±64</td>
<td>73 ±1</td>
<td>313 ±13</td>
<td>35 ±0</td>
<td>678 ±46</td>
<td>59 ±1</td>
<td>1028 ±132</td>
<td>95 ±2</td>
<td>1385 ±132</td>
</tr>
<tr>
<td>#9</td>
<td>971 ±80</td>
<td>83 ±2</td>
<td>337 ±15</td>
<td>37 ±0</td>
<td>779 ±57</td>
<td>67 ±1</td>
<td>1127 ±144</td>
<td>102 ±2</td>
<td>1468 ±144</td>
</tr>
<tr>
<td>#10</td>
<td>724 ±50</td>
<td>67 ±1</td>
<td>281 ±11</td>
<td>32 ±0</td>
<td>590 ±37</td>
<td>54 ±1</td>
<td>913 ±106</td>
<td>87 ±2</td>
<td>1218 ±106</td>
</tr>
<tr>
<td>#11</td>
<td>1414 ±146</td>
<td>113 ±2</td>
<td>460 ±24</td>
<td>47 ±1</td>
<td>1155 ±108</td>
<td>91 ±2</td>
<td>1582 ±259</td>
<td>136 ±3</td>
<td>2088 ±259</td>
</tr>
<tr>
<td>#12</td>
<td>9434 ±2950</td>
<td>481 ±22</td>
<td>2037 ±256</td>
<td>145 ±4</td>
<td>6944 ±1768</td>
<td>377 ±15</td>
<td>10000 ±3577</td>
<td>493 ±22</td>
<td>10638 ±3577</td>
</tr>
<tr>
<td>#13</td>
<td>4926 ±1131</td>
<td>291 ±10</td>
<td>1277 ±124</td>
<td>110 ±2</td>
<td>3891 ±775</td>
<td>234 ±7</td>
<td>4762 ±1373</td>
<td>327 ±12</td>
<td>5747 ±1373</td>
</tr>
<tr>
<td>#14</td>
<td>2667 ±391</td>
<td>240 ±8</td>
<td>907 ±68</td>
<td>106 ±2</td>
<td>2315 ±329</td>
<td>195 ±6</td>
<td>2899 ±682</td>
<td>305 ±11</td>
<td>4065 ±682</td>
</tr>
</tbody>
</table>

79
Table 3.22: NNS estimates from sensitivity analysis over the PCa-screening leadtime \((t_{LT})\). For each risk group, the headings “low” and “high” respectively denote the low and high levels of parameter \(t_{LT}\). The 95% CI estimates are based on simulated samples of 1 million patients.

<table>
<thead>
<tr>
<th></th>
<th>Whites</th>
<th>African American</th>
<th>Family History</th>
<th>CCI=1</th>
<th>CCI≥2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>#2</td>
<td>350 ± 15</td>
<td>172 ± 5</td>
<td>137 ± 3</td>
<td>81 ± 2</td>
<td>285 ± 11</td>
</tr>
<tr>
<td>#3</td>
<td>282 ± 11</td>
<td>151 ± 4</td>
<td>112 ± 3</td>
<td>71 ± 1</td>
<td>228 ± 8</td>
</tr>
<tr>
<td>#4</td>
<td>296 ± 11</td>
<td>157 ± 4</td>
<td>116 ± 3</td>
<td>73 ± 1</td>
<td>238 ± 8</td>
</tr>
<tr>
<td>#5</td>
<td>243 ± 8</td>
<td>137 ± 3</td>
<td>99 ± 2</td>
<td>65 ± 1</td>
<td>196 ± 6</td>
</tr>
<tr>
<td>#6</td>
<td>180 ± 5</td>
<td>112 ± 2</td>
<td>77 ± 1</td>
<td>55 ± 1</td>
<td>145 ± 4</td>
</tr>
<tr>
<td>#7</td>
<td>282 ± 11</td>
<td>151 ± 4</td>
<td>112 ± 3</td>
<td>71 ± 1</td>
<td>228 ± 8</td>
</tr>
<tr>
<td>#8</td>
<td>243 ± 8</td>
<td>137 ± 3</td>
<td>99 ± 2</td>
<td>65 ± 1</td>
<td>196 ± 6</td>
</tr>
<tr>
<td>#9</td>
<td>279 ± 10</td>
<td>157 ± 4</td>
<td>106 ± 2</td>
<td>70 ± 1</td>
<td>224 ± 7</td>
</tr>
<tr>
<td>#10</td>
<td>216 ± 7</td>
<td>127 ± 3</td>
<td>90 ± 2</td>
<td>61 ± 1</td>
<td>174 ± 5</td>
</tr>
<tr>
<td>#11</td>
<td>399 ± 18</td>
<td>206 ± 6</td>
<td>145 ± 4</td>
<td>89 ± 2</td>
<td>320 ± 13</td>
</tr>
<tr>
<td>#12</td>
<td>1946 ± 200</td>
<td>945 ± 63</td>
<td>498 ± 25</td>
<td>277 ± 10</td>
<td>1513 ± 137</td>
</tr>
<tr>
<td>#13</td>
<td>1125 ± 87</td>
<td>517 ± 25</td>
<td>371 ± 16</td>
<td>197 ± 6</td>
<td>910 ± 63</td>
</tr>
<tr>
<td>#14</td>
<td>945 ± 67</td>
<td>402 ± 17</td>
<td>333 ± 13</td>
<td>180 ± 5</td>
<td>772 ± 50</td>
</tr>
</tbody>
</table>

80
3.4 Discussion

There are varying recommendations for PCa screening, including those provided by the USPSTF [78], the ACS [98], the AUA [16], and the National Comprehensive Cancer Network (NCCN) [15], among others [6, 85]. The USPSTF recommends against routine screening for average-risk males [78]. The AUA recommends that screening from ages 55 to 69 should involve shared decision-making that considers family history, race, and comorbidities. Furthermore, the AUA acknowledges that race- or family-history-based subgroups may benefit from screening from age 40 to 54 [16]. Similarly, both the NCCN and the ACS recommend that the decision to initiate screening must consider race, family history, and life expectancy [15, 98].

In contrast to other simulation studies [43, 48], we compared PCa screening for different well-established risk groups, and we quantified the results in terms of the expected NNS to avert 1 PCa death. In its most common implementation, NNS has shortcomings, such as censoring [71]. However, we avoided censoring by using an unlimited follow-up period for each simulated patient. Our results suggest that a strategy with a relatively low PSA threshold (2.5 ng/mL compared with 3.0–6.5 ng/mL) might achieve the greatest mortality reduction with the fewest number of screens. Also, based on the limited but representative set of screening strategies we considered, our results suggest that a single “one-size-fits-all” strategy might be desirable for the risk groups we compared, from the perspective of minimizing NNS.

The lowest NNS was for the African American risk group, suggesting that screening in this subpopulation might have the most benefit with the fewest patients screened. On the other hand, the NNS values for white US males with CCI=1 or CCI≥2 were much higher, supporting the position that patients with significant competing risks from comorbidities might stand to gain less than other risk groups in the absence of risk-based customized PSA screening. Thus, the benefit from PSA screening did not appear to be uniformly distributed across risk groups; screening guidelines that fail to account for this heterogeneity might, in part, explain, or potentially exacerbate observed disparities across some risk groups.

A limitation of our study is that our model assumed all diagnoses of PCa were treated by radical prostatectomy. However, a systematic review of 18 RCTs and 473 observational studies failed to find a treatment modality that was superior in terms of survival outcomes [109]. Moreover, contemporary use
of active surveillance would likely improve postdetection quality-adjusted survival; thus our findings likely underestimate the benefits of screening. Even in light of this underestimation our study suggests the benefits of screening are higher than reported from the ERSPC trial. Another limitation of our study is that the annual probability of PCa-specific death for African American men was approximated by multiplying the annual probability of PCa-specific death for the reference subpopulation by an all-cause mortality hazard ratio. This was done because there is evidence to suggest that differences in outcomes between African American and white men may be due to inequalities in access to health care and/or differences in comorbidities [61]. Also, the African American risk group depends on more assumptions than do the other risk groups because the African American risk group involves the most parameter adjustments. In light of this, results from the African American risk group must be treated with heightened caution. And, finally, our conclusions are based on models of different risk groups derived from predominantly Caucasian US-male population parameters.

In spite of these limitations, there are a number of important insights that can be drawn from the results. First, for all risk groups, the same strategy (screening annually from age 50 to 75 using the PSA threshold 2.5 ng/mL) yielded the smallest estimated expected NNS. Second, across risk groups, NNSs differed significantly, with African American men having the lowest NNS, suggesting the benefits of screening may vary substantially among risk groups. Third, comorbidities were associated with substantially higher NNSs, suggesting that a patient’s individual health status should play an important role in the decision to have a PSA test. Given the evidence of increased potential benefit for patients in the Family History and African American risk groups, there could be benefits to additional research to further study the effect of screening in healthy African American men and men without a family history of PCa.
A Three-Stage Regularization Approach for the Optimization of PSA-Based Prostate Cancer Screening

4.1 Introduction

Finding an optimal strategy that trades off the benefits and harms of diagnosis and treatment is difficult. The more aggressive a strategy is, the more men that will be biopsied. So, an aggressive strategy may be very good at finding and treating PCa, but it may also result in unnecessary biopsies and treatments. And although there is a definite perceived harm caused by overtreatment, it is difficult in the context of a clinical study or an RCT to estimate accurately the actual harms of a given strategy so as to weigh them against the benefits in deciding whether a given strategy, though possibly very aggressive, is justified. Despite the limitations of simulation studies, they have the advantage of being able to estimate the performance of many alternative strategies that would otherwise be prohibitive to compare using RCTs. We capitalize on this advantage by extending the simulation model from Chapter 3 to find strategies that achieve large expected quality-adjusted survival (as measured in QALYs or a derivative thereof). This QALYs-extended simulation model is embedded in a complex random-search based metaheuristic
optimization method in order to search for a strategy for a given risk group that most favorably balances the benefits and harms of screening (i.e., that maximizes expected quality-adjusted survival). The use of a metaheuristic is justified because the underlying objective function being maximized is a complex, nonlinear, probably-nonconcave, stochastic function of many variables. This function, which estimates the expected quality-adjusted survival of a given screening strategy (solution), is really a discrete-event stochastic process.

The optimization approach we develop in this chapter combines global and local search methods. The global method (stage 1) is a GA. The GA is followed by a clean-up procedure (stage 2) that eliminates noise from the strategy produced by the GA. The local method (stage 3) is an implementation of the COMPASS procedure that is used to make local smoothing improvements to the strategy from the clean-up procedure, with the ultimate goal of rendering an improved screening strategy that is not characterized by a highly nonsmooth (erratic) time series of PSA thresholds. The GA identifies a PSA screening strategy with large expected QALYs, and then—after the clean-up procedure—our implementation of COMPASS tries to make smoothing improvements without substantial loss in expected QALYs. This combination of a GA (with a clean-up procedure) and COMPASS constitutes the three-stage simulation-optimization procedure we use to solve variants of the PCa screening problem introduced in Chapter 3. By combining the three stages, we identify screening strategies yielding large expected QALYs, but we can also provide strategies that are less noisy and less irregular (i.e., more smooth), and thus more amenable to actual clinical implementation and analysis.

In addition to carrying out simulation optimization of the PCa screening problem, we develop an analytical formulation of a simplified model of PCa screening. In this simplified model, there is a single screen event at some fixed age for all patients. The formulated optimization problem using this single-period PCa screening model is to find the optimal PSA threshold with which to screen all patients at some predefined age. In this research, we describe this single-period model and derive point and CI estimators for the objective function. We use this single-period model to further analyze the results from the three-stage simulation-optimization method.

Whereas previously we have solved a variant of the PCa screening problem using a GA with expected
QALYs as the performance measure of the simulation model [107], in this chapter we experiment with different variants of a GA and compare and validate them using a set of well known global-optimization test problems. We also motivate and develop a different performance measure, the NNSQ statistic, which represents the number needed to screen to produce a net gain of 1 QALY in the screened population. We use an estimator of expected QALYs gained (QG) relative to no screening, which is a more well-behaved functional transform of NNSQ, as the performance measure in our three-stage procedure. The expected value of QG is simply the reciprocal of NNSQ, and it expresses the expected gain in QALYs that is produced by using a given screening strategy relative to no screening. Computing this relative measure involves comparisons of paired (synchronized) simulations. This use of common random numbers (CRN) to estimate our stochastic performance measure permits us to make reliable comparisons of different screening strategies with fewer simulation replications (i.e., fewer simulated patients) than would be required to make similarly reliable comparisons using merely expected QALYs. Additionally, we explain how NNSQ, or a suitable transform thereof, could have significant utility for research into other disease-screening problems, independent of the performance benefit to simulation methodology.

While a GA can be highly effective at producing an improved screening strategy, the stochastic nature of GAs as simulation-optimization heuristics and the low sensitivity of expected QALYs to some individual PSA thresholds tends to produce PSA threshold–based screening strategies with highly variable PSA thresholds—some of which may be practically insignificant (i.e., noise)—over age (time). To eliminate practically insignificant PSA thresholds, we introduce a clean-up procedure as a secondary stage following the GA. Given that highly variable strategies may be difficult to accept and apply in practice, we introduce a regularizing (smoothing) objective function that incorporates a QALYs-based performance measure and a measure of the regularity (smoothness) of PSA thresholds over time. We integrate the regularizing objective function in a tertiary stage, after the clean-up procedure, in which a COMPASS implementation is used to make regularizing improvements to the cleaned-up optimized strategy from the GA.

Also, our simulation of clinical detection is an improvement upon the QALYs-based examination of PSA screening in our previous work in [107]. Some PCa cases that evade screen detection will ultimately
be detected clinically by virtue of the manifestation of physiological symptoms that lead the physician to suspect PCa. In some of these cases, the PCa that was not detected by PSA screening will in fact be treated. So, if clinical detection is not accounted for, the estimated benefit of screening relative to no screening is likely to be exaggerated. We discussed in Chapter 3 how the consideration of clinical detection renders our NNS estimates more reliable, and by the same logic our consideration of clinical detection renders our NNSQ (and QG) estimates more reliable.

The remainder of this chapter is organized as follows. In Section 4.2 we describe an analytical model for optimal single-period screening, modifications to the simulation model of the natural history of prostate cancer from Chapter 3, and the method for finding near-optimal screening strategies. In Section 4.3 we present estimates of the parameters of our model, including the plausible ranges used in sensitivity analysis, and results from the three-stage method, the analytical single-period model, and sensitivity analysis. Finally in Section 4.4 we analyze the results and make concluding remarks about this chapter.

4.2 Methodology

This subsection is organized as follows. In Section 4.2.1, we develop a simple analytical model of PCa screening as a vehicle for gaining insights into the PCa screening problem. In Section 4.2.2 we describe a QALYs parameterization of the model from Chapter 3. In Section 4.2.3 we develop the variance-reducing NNSQ statistic. In Section 4.2.4 we formalize the PCa screening process in order to formalize the optimization problem, which we then present formally in Section 4.2.5. In Section 4.2.6 we present a regularization approach that modifies the objective function in the optimization problem. Finally, in Section 4.2.7 we present a three-stage regularizing simulation-optimization method.

4.2.1 Single-Period Optimal Screening Analytical Model

Perhaps the simplest conceptual model for thinking about the PCa screening problem is that of a single point in time, \( t \), at which a PSA screening test occurs with some fixed PSA threshold. In this conceptual model, each patient is PSA-tested at that time (some age fixed for all patients). If the patient’s PSA level
is found to be greater than or equal to the fixed threshold, then a biopsy is performed. Upon biopsy, if PCa is detected, then the patient is treated; otherwise, the patient is not treated. In this conceptual model, we seek the optimal PSA threshold for biopsy, i.e., the unique PSA threshold that maximizes the patient’s expected net benefit (expressed in QALYs) at the given time $t$.

Before discussing the details of the simulation optimization approach, we begin by presenting an analytical formulation of this simplified conceptual model. This analytical model fundamentally involves the following pieces of information: the probability distribution of PSA dependent upon whether PCa is present; the expected benefit of biopsy; the expected disutility incurred by biopsy; and the probability of having PCa.

As for the probability distribution of PSA, we utilize the PSA growth model of Gulati et al. [40] from Chapter 3, which is

$$Z \equiv \log(\text{PSA}_j(t)) = \beta_{0,j} + \beta_{1,j} \cdot t + \beta_{2,j} (t - t_{0,j}) \cdot \mathbb{I}(t > t_{0,j})$$

$$+ \varepsilon_{t,j} \quad \text{for} \quad t \in T,$$

(4.1)

where for patient $j$ we have the patient-specific regression coefficients

$$\beta_{k,j} \sim \mathcal{N}(\mu_k, \sigma_k^2) \quad \text{for} \quad k = 0, 1, 2,$$

and the patient-specific noise term

$$\varepsilon_{t,j} \sim \mathcal{N}(0, \sigma_e^2);$$

recall that $\mathbb{I}(\cdot)$ is the indicator function and $t_{0,j}$ is the time (age) at PCa onset for patient $j$. Although Gulati et al. [40] state in their supplementary material that the noise term is independent of the patient-specific regression coefficients so that we have

$$\text{Corr}(\beta_{k,j}, \varepsilon_{t,j}) = 0 \quad \text{for} \quad k \in \{0, 1, 2\}, \quad t \in T,$$

the authors do not explicitly state that the regression coefficients are mutually independent. However, in
the absence of any explicitly defined model parameters specifying the correlations between the regression coefficients, we infer the assumption that

$$\text{Corr}(\beta_{k,j}, \beta_{\ell,j}) = 0 \text{ for } k \neq \ell \text{ and } k, \ell \in \{0, 1, 2\}. \quad (4.2)$$

If we fix patient $j$ and a time $t \in \mathcal{T}$ that is assumed to be earlier than clinical detection, then we have the events

$$\text{Ca} \equiv \{ \text{patient } j \text{ has PCa at time } t \},$$

$$\text{NC} \equiv \{ \text{patient } j \text{ does not have PCa at time } t \},$$

and the random variable

$$Y \equiv e^Z = \text{PSA}_j(t).$$

Note that our single-period analytical model does not account for clinical detection. It follows from Equations (4.1)–(4.2) that given event Ca, the random variable $Y$ has the conditional lognormal density function

$$f_{Y|\text{Ca}}(y) = \frac{1}{y\sigma_{Z|\text{Ca}}\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{\ln y - \mu_{Z|\text{Ca}}}{\sigma_{Z|\text{Ca}}} \right)^2 \right\} \text{ for } y \geq 0, \quad (4.3)$$

where

$$\mu_{Z|\text{Ca}} \equiv E[Z|\text{Ca}]$$

and

$$\sigma_{Z|\text{Ca}}^2 \equiv \text{Var}[Z|\text{Ca}].$$
Similarly given the no-cancer event NC, the variate $Y$ has the conditional lognormal density function

$$f_{Y|NC}(y) = \frac{1}{y \sigma_{Z|NC} \sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left( \frac{\ln y - \mu_{Z|NC}}{\sigma_{Z|NC}} \right)^2 \right\} \text{ for } y \geq 0,$$

(4.4)

where

$$\mu_{Z|NC} \equiv \mathbb{E}[Z|NC]$$

(4.5)

and

$$\sigma_{Z|NC}^2 \equiv \text{Var}[Z|NC].$$

(4.6)

Now, using the parameters of the model in Equation (4.1), which were described in Table 3.5, we define the mean and variance of patient PSA conditioned on whether the patient has PCa. Given that a patient has PCa, the mean PSA level is

$$\mu_{Z|Ca} = \mu_{0|Ca} + \mu_{1|Ca} \cdot t + \mu_{2|Ca} \cdot \left( t - \mathbb{E}\left[ t_{0,j}|Ca, t \right] \right),$$

(4.7)

where $t$ is the given time (age) for patient $j$, and $\mathbb{E}\left[ t_{0,j}|Ca, t \right]$ is the conditional expected PCa onset time taken over all patients who have PCa at time (age) $t$. Given that a patient has PCa, the variance of PSA is

$$\sigma_{Z|Ca}^2 = \sigma_{0|Ca}^2 + \sigma_{1|Ca}^2 + \sigma_{2|Ca}^2 \cdot \text{Var}\left[ t - t_{0,j}|Ca, t \right]$$

$$+ \mu_{2|Ca} \cdot \text{Var}\left[ t - t_{0,j}|Ca, t \right]$$

$$+ \left( \mathbb{E}\left[ t - t_{0,j}|Ca, t \right] \right)^2 \cdot \sigma_{2|Ca}^2 + \sigma_{\varepsilon}^2,$$

(4.8)

where $\text{Var}\left[ t - t_{0,j}|Ca, t \right]$ is the conditional variance of the time $t - t_{0,j}$ since PCa onset given the patient has PCa at the fixed time (age) $t$; and $\mathbb{E}\left[ t - t_{0,j}|Ca, t \right]$ is the corresponding conditional mean. Given
that a patient does not have PCa, the mean PSA level is

\[ \mu_{Z|NC} = \mu_{0|NC} + \mu_{1|NC} \cdot t \quad (4.9) \]

and the variance of PSA is

\[ \sigma_{Z|NC}^2 = \sigma_{0|NC}^2 + \sigma_{1|NC}^2 \cdot t^2 + \sigma_{\epsilon}^2, \quad (4.10) \]

where \( t \) is the fixed patient time (age).

We also define the events

\[ \text{Tr} \equiv \{ \text{patient } j \text{ is treated immediately after time } t \} , \]

\[ \text{NT} \equiv \{ \text{patient } j \text{ is not treated immediately} \} , \]

\[ \text{B} \equiv \{ \text{patient } j \text{ has biopsy at time } t \} , \]

and

\[ \text{NB} \equiv \{ \text{patient } j \text{ does not have biopsy at time } t \} . \]

Given the PSA threshold \( \phi \), we define the random variable

\[ \mathcal{R}(\phi) \equiv \begin{cases} \text{reward received by patient } j \text{ at time } t \text{ (i.e., just after any} & \\
\text{actions taken at time } t \text{ because } Y \geq \phi). \end{cases} \]

Therefore, the expected net benefit of biopsy and any subsequent treatment (if PCa is detected by screening) is

\[ \mathcal{B}(\phi) \equiv E[\mathcal{R}(\phi)|\text{Ca, B}] - E[\mathcal{R}(\phi)|\text{Ca, NB}]; \]
and the expected disutility of biopsy is

\[ \delta(\phi) \equiv E[\mathcal{R}(\phi)|\text{NC, NB}] - E[\mathcal{R}(\phi)|\text{NC, B}]. \]

We let

\[ \pi_{\text{Ca}} \equiv \Pr \{ \text{Ca} \} \]
\[ \pi_{\text{NC}} \equiv \Pr \{ \text{NC} \} = 1 - \pi_{\text{Ca}}. \]

We take

\[ B = \{ Y \geq \phi \}, \]

and

\[ \Pr \{ B \mid \text{Ca} \} = \Pr \{ Y \geq \phi \mid \text{Ca} \} = 1 - F_{Y|\text{Ca}}(\phi), \]

where \( F_{Y|\text{Ca}}(\cdot) \) is the distribution function corresponding to the density function in Equation (4.3); and similarly we have

\[ \Pr \{ B \mid \text{NC} \} = \Pr \{ Y \geq \phi \mid \text{NC} \} = 1 - F_{Y|\text{NC}}(\phi), \]

where \( F_{Y|\text{NC}}(\cdot) \) is the distribution function corresponding to the density function in Equation (4.4).

Although in general \( b \) and \( \delta \) may depend on \( \phi \), in the following analysis we assume that \( b \) and \( \delta \) are positive constants so that the analytical optimization described in the following equations can be achieved without requiring information on the functional forms of \( b(\phi) \) and \( \delta(\phi) \). This is reasonable because \( \delta(\phi) \) may be thought of as a one-time disutility like \( D_{\text{Biop}} \), which is reasonably assumed to be
independent of the screening threshold, and because attempting to characterize the functional form of $b(\phi)$ would require a compute-intensive random sampling from the non-single-period discrete-event stochastic process. In terms of the setup in Equations (4.1)–(4.4), the single period expected reward for screening is

$$
\vartheta(\phi) \equiv E[R(\phi)]
= \pi_{Ca} \Pr \{B|Ca\} E[R(\phi)|Ca, B] + \pi_{Ca} \Pr \{NB|Ca\} E[R(\phi)|Ca, NB]
+ \pi_{NC} \Pr \{B|NC\} E[R(\phi)|NC, B] + \pi_{NC} \Pr \{NB|NC\} E[R(\phi)|NC, NB]
= \pi_{Ca} \left(1 - F_{Y|Ca}(\phi)\right) E[R(\phi)|Ca, B] + \pi_{Ca} F_{Y|Ca}(\phi) E[R(\phi)|Ca, NB]
+ \pi_{NC} \left(1 - F_{Y|NC}(\phi)\right) E[R(\phi)|NC, B] + \pi_{NC} F_{Y|NC}(\phi) E[R(\phi)|NC, NB]
= \left(\pi_{Ca} E[R(\phi)|Ca, B] + \pi_{NC} E[R(\phi)|NC, B]\right)
- \pi_{Ca} F_{Y|Ca}(\phi) \left(E[R(\phi)|Ca, B] - E[R(\phi)|Ca, NB]\right)
+ \pi_{NC} F_{Y|NC}(\phi) \left(E[R(\phi)|NC, B] - E[R(\phi)|NC, NB]\right)
= \pi_{Ca} E[R(\phi)|Ca, B] + \pi_{NC} E[R(\phi)|NC, B] - \pi_{Ca} F_{Y|Ca}(\phi) b + \pi_{NC} F_{Y|NC}(\phi) \delta. \quad (4.11)
$$

If the function $\vartheta(\phi)$ attains a local (or global) maximum at $\phi = \phi^*$ and if we assume that $F_{Y|Ca}(\phi)$ and $F_{Y|NC}(\phi)$ are the only terms in Equation (4.11) that depend on $\phi$, then we must have

$$
\vartheta'(\phi^*) = \pi_{NC} f_{Y|NC}(\phi^*) \delta - \pi_{Ca} f_{Y|Ca}(\phi^*) b = 0 \quad (4.12)
$$

and

$$
\vartheta''(\phi^*) \leq 0. \quad (4.13)
$$

To solve the equation $\vartheta'(\phi) = 0$, we proceed as follows. If $\vartheta'(\phi) = 0$, then

$$
\pi_{NC} f_{Y|NC}(\phi) \delta = \pi_{Ca} f_{Y|Ca}(\alpha) b
$$
implies
\[
\frac{f_{Y|\text{Ca}}(\alpha)}{f_{Y|\text{NC}}(\phi)} = \left( \frac{\pi_{\text{NC}}}{\pi_{\text{Ca}}} \right) \times \left( \frac{\delta}{b} \right)
\]
by the assumption that \( \delta \) and \( b \) are positive constants. In view of Equations (4.3)–(4.6), we must have
\[
\left( \frac{\sigma_{Z|\text{NC}}}{\sigma_{Z|\text{Ca}}} \right) \exp \left\{ -\frac{1}{2} \left[ \left( \frac{\ln \phi - \mu_{Z|\text{Ca}}}{\sigma_{Z|\text{Ca}}} \right)^2 - \left( \frac{\ln \phi - \mu_{Z|\text{NC}}}{\sigma_{Z|\text{NC}}} \right)^2 \right] \right\} = \left( \frac{\pi_{\text{NC}}}{\pi_{\text{Ca}}} \right) \times \left( \frac{\delta}{b} \right);
\]
and if we take
\[
\gamma = -2 \ln \left[ \left( \frac{\sigma_{Z|\text{Ca}}}{\sigma_{Z|\text{NC}}} \right) \left( \frac{\pi_{\text{NC}}}{\pi_{\text{Ca}}} \right) \left( \frac{\delta}{b} \right) \right],
\]
then the desired root of \( \partial^\prime(\phi) = 0 \) is the solution of the following quadratic equation in the variable \( \zeta = \ln \phi \):
\[
\left( \frac{\zeta - \mu_{Z|\text{Ca}}}{\sigma_{Z|\text{Ca}}} \right)^2 - \left( \frac{\zeta - \mu_{Z|\text{NC}}}{\sigma_{Z|\text{NC}}} \right)^2 = \gamma,
\]
which is equivalent to
\[
\left( \sigma_{Z|\text{NC}}^2 - \sigma_{Z|\text{Ca}}^2 \right) \zeta^2 + 2 \left( \sigma_{Z|\text{Ca}\mu Z|\text{NC}}^2 - \sigma_{Z|\text{NC}\mu Z|\text{Ca}}^2 \right) \zeta
\]
\[
+ \sigma_{Z|\text{NC}\mu Z|\text{Ca}}^2 - \sigma_{Z|\text{Ca}\mu Z|\text{NC}}^2 - \sigma_{Z|\text{Ca}}^2 \sigma_{Z|\text{NC}}^2 = 0
\]
(4.15)
For the left-hand side of Equation (4.15) to define a quadratic (or linear) function that attains a bounded maximal value, we must assume that \( \left( \sigma_{Z|\text{NC}}^2 - \sigma_{Z|\text{Ca}}^2 \right) \leq 0 \) so we have
\[
\sigma_{Z|\text{NC}}^2 \leq \sigma_{Z|\text{Ca}}^2,
\]
which is reasonable because the presence of cancer in some sense corresponds to greater “pathological entropy” in the patient compared with the absence of cancer.

Therefore, the solution of Equation (4.15) is
\[
\zeta^\ast = - \left( \frac{\sigma_{Z|\text{Ca}\mu Z|\text{NC}}^2 - \sigma_{Z|\text{NC}\mu Z|\text{Ca}}^2}{\sigma_{Z|\text{NC}}^2 - \sigma_{Z|\text{Ca}}^2} \right) \pm \frac{\sqrt{\delta}}{2 \left( \sigma_{Z|\text{NC}}^2 - \sigma_{Z|\text{Ca}}^2 \right)},
\]
(4.16)
where

\[ D \equiv 4 \left( \sigma_{Z|Ca}^2 \mu_{Z|NC} - \sigma_{Z|NC}^2 \mu_{Z|Ca} \right)^2 + 4 \left( \sigma_{Z|Ca}^2 - \sigma_{Z|NC}^2 \right) \left( \sigma_{Z|NC}^2 \mu_{Z|Ca}^2 - \sigma_{Z|Ca}^2 \mu_{Z|NC}^2 - \sigma_{Z|Ca}^2 \sigma_{Z|NC}^2 \gamma \right); \]

and provided that Equation (4.16) does not have complex roots, we have

\[ \phi^* = \exp (\xi^*) > 0. \]

In Section 4.3.3 we will evaluate \( \pi_{Ca}, \delta, \) and \( \ell \) by simulation for selected values of \( t \) and \( \phi \) in order to evaluate \( \gamma \) and determine how (4.16) describes the behavior of the expected net benefit function \( \vartheta (\phi) \). In general we will see that if (4.16) has two complex roots, then \( \vartheta (\phi) \) is a strictly decreasing function of \( \phi \) for \( \phi \in (0, \infty) \). If (4.16) has two real roots, then one of the roots corresponds to a global maximum of \( \vartheta (\phi) \) and the other root corresponds to a global minimum.

We now present a large-sample-size CI estimator for \( \vartheta (\phi) \) when \( \ell = \ell (\phi) \). Given \( \alpha \in (0, 1) \), we seek to compute an asymptotically valid \( 100(1 - \alpha)\% \) CI for Equation (4.11) given a time (age) \( t \) and PSA screening level \( \phi \). Based on a large sample of \( m \) simulated patients, we obtain the estimate

\[ \hat{\pi}_{Ca} = m^{-1} \sum_{j=1}^{m} I (\text{Patient } j \text{ has PCa at time } t); \]  

(4.17)

and based on an independently seeded set of \( m \) simulated patients, we obtain the estimate

\[ \hat{\ell} (\phi) = m^{-1} \sum_{k=1}^{m} \ell (\phi), \]  

(4.18)

where \( \ell_k (\phi) \) is the net benefit of biopsy performed at time \( t \) and of any subsequent treatment if PCa is detected by screening with PSA threshold \( \phi \) at time \( t \).

It follows from Equation (4.17) that \( \text{Var} [\hat{\pi}_{Ca}] \) can be established by

\[ \text{Var} [\hat{\pi}_{Ca}] = \hat{\pi}_{Ca} (1 - \hat{\pi}_{Ca}) / m; \]

(4.19)
and it follows from Equation (4.18) that \( \text{Var} \left[ \hat{b}(\phi) \right] \) can be estimated by

\[
\hat{\text{Var}} \left[ \hat{b}(\phi) \right] = \frac{S^2_{\hat{b}(\phi)}}{m};
\]  

(4.20)

where

\[
S^2_{\hat{b}(\phi)} \equiv (m - 1)^{-1} \sum_{k=1}^{m} \left[ \hat{b}_k(\phi) - \hat{b}(\phi) \right]^2.
\]  

(4.21)

Based on another independently seeded set of \( m \) patients, we estimate \( \mathbb{E}[R(\phi) | \text{Ca}, B] \) as follows:

\[
G_\ell(\phi) \equiv \begin{cases} 
\text{reward received by patient } \ell \text{ at time } t \text{ because patient has} \\
\text{PCa and PSA}_\ell(t) > \phi \text{ so that patient has a biopsy at time } t,
\end{cases}
\]  

(4.22)

for \( \ell = 1, \ldots, m \) so that we estimate \( \mathbb{E}[R(\phi) | \text{Ca}, B] \) from the sample statistics

\[
\bar{G}(\phi) = m^{-1} \sum_{\ell=1}^{m} G_\ell(\phi)
\]  

(4.23)

and

\[
S^2_{\bar{G}(\phi)} = (m - 1)^{-1} \sum_{\ell=1}^{m} \left[ G_\ell(\phi) - \bar{G}(\phi) \right]^2.
\]  

(4.24)

Similarly based on yet another independently seeded set of \( m \) patients, we estimate \( \mathbb{E}[R(\phi) | \text{NC}, B] \) as follows:

\[
H_u(\phi) \equiv \begin{cases} 
\text{reward received by patient } u \text{ at time } t \text{ because patient has} \\
\text{PSA}_u(t) > \phi \text{ but patient does not have PCa so that} \\
\text{patient has unnecessary biopsy at time } t,
\end{cases}
\]  

(4.25)

for \( u = 1, \ldots, m \) so that we estimate \( \mathbb{E}[R(\phi) | \text{NC}, B] \) from the sample statistics

\[
\bar{H}(\phi) = m^{-1} \sum_{u=1}^{m} H_u(\phi)
\]  

(4.26)

and

\[
S^2_{\bar{H}(\phi)} = (m - 1)^{-1} \sum_{u=1}^{m} \left[ H_u(\phi) - \bar{H}(\phi) \right]^2.
\]  

(4.27)
Next we will discuss how to estimate $F_{Y|Ca}(\phi)$. Based on another independently seeded set of $m$ patients, we estimate $E[t - t_0|Ca]$ and $\text{Var}[t - t_0|Ca]$ as follows. For the $j$th patient who at time (age) $t$ has PCa, let $A_j$ denote the time delay since the onset of PCa,

$$A_j = t - t_{0,j} \quad \text{for} \quad j = 1, \ldots, m, \quad (4.28)$$

so that we have the sample statistics

$$\bar{A} = m^{-1} \sum_{j=1}^{m} A_j \quad (4.29)$$
$$S_A^2 = \frac{1}{m-1} \sum_{j=1}^{m} (A_j - \bar{A})^2 \quad (4.30)$$

and the corresponding estimators

$$\hat{\mu}_{Z|Ca} = \mu_{0|Ca} + \mu_{1|Ca} \cdot t + \mu_{2|Ca} \cdot \bar{A}, \quad (4.31)$$
$$\hat{\sigma}_Z^2|Ca = \sigma_{0|Ca}^2 + \sigma_{1|Ca}^2 \cdot t^2 + \sigma_{2|Ca}^2 \cdot S_A^2 + \mu_{2|Ca}^2 \cdot S_A^2 + \bar{A}^2 \cdot \sigma_{2|Ca}^2 + \sigma_e^2 \quad (4.32)$$

by results IR$_2$. We will use the multivariate delta method (Theorem 5.4.6 of Hogg et al. [52]) to derive the large-sample properties of the estimator

$$\hat{F}_{Y|Ca}(\phi) = \Phi \left[ \frac{\ln(\phi) - \hat{\mu}_{Z|Ca}}{\hat{\sigma}_Z|Ca} \right] \quad (4.33)$$

of the target distribution function

$$F_{Y|Ca}(\phi) = \Phi \left[ \frac{\ln(\phi) - \mu_{Z|Ca}}{\sigma_{Z|Ca}} \right], \quad (4.34)$$

where

$$\Phi(z) = \int_{-\infty}^{z} \varphi(u) \, du \quad (4.35)$$
is the standard normal distribution function and
\[ \varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \]  
(4.36)
is the standard normal density function. If \( m \) is large, then the multivariate delta method ensures that \( \hat{F}_{Y|Ca}(\phi) \) is approximately normal with mean
\[ E\left[ \hat{F}_{Y|Ca}(\phi) \right] = F_{Y|Ca}(\phi) \]  
(4.37)
and variance
\[ \text{Var}\left[ \hat{F}_{Y|Ca}(\phi) \right] = \nabla F_{Y|Ca}(\phi) \Sigma \left( \hat{\mu}_{Z|Ca}, \hat{\sigma}_{Z|Ca}^2 \right) \left[ \nabla F_{Y|Ca}(\phi) \right]^T, \]  
(4.38)
where we have the gradient vector of \( F_{Y|Ca}(\phi) \) with respect to the parameters \( \mu_{Z|Ca} \) and \( \sigma_{Z|Ca}^2 \).

\[ \nabla F_{Y|Ca}(\phi) = \begin{bmatrix} \frac{\partial F_{Y|Ca}(\phi)}{\partial \mu_{Z|Ca}} & \frac{\partial F_{Y|Ca}(\phi)}{\partial \sigma_{Z|Ca}^2} \end{bmatrix} = \begin{bmatrix} -\varphi \left( \frac{\ln (\phi) - \mu_{Z|Ca}}{\sigma_{Z|Ca}} \right) & -\varphi \left( \frac{\ln (\phi) - \mu_{Z|Ca}}{\sigma_{Z|Ca}} \right) \left( \frac{\ln (\phi) - \mu_{Z|Ca}}{\sigma_{Z|Ca}} \right) \frac{1}{2\sigma_{Z|Ca}} \end{bmatrix} \]  
(4.39)
and we have the covariance matrix of the random vector \( \left[ \hat{\mu}_{Z|Ca}, \hat{\sigma}_{Z|Ca}^2 \right] \).

\[ \Sigma \left( \hat{\mu}_{Z|Ca}, \hat{\sigma}_{Z|Ca}^2 \right) = \begin{bmatrix} \text{Var} \left[ \hat{\mu}_{Z|Ca} \right] & \text{Cov} \left[ \hat{\mu}_{Z|Ca}, \hat{\sigma}_{Z|Ca}^2 \right] \\ \text{Cov} \left[ \hat{\mu}_{Z|Ca}, \hat{\sigma}_{Z|Ca}^2 \right] & \text{Var} \left[ \hat{\sigma}_{Z|Ca}^2 \right] \end{bmatrix}. \]  
(4.40)
If we assume that \( A_j \sim N(\mu_A, \sigma_A^2) \), then it follows that \( \bar{A} \) and \( S_A^2 \) are independent (Theorem 3.6.1(b) of
Hogg et al. [52]); and from Equations (4.31) and (4.32), we see that

\[
\text{Var} \left[ \hat{\mu}_{Z|\text{Ca}} \right] = \mu^2_{2|\text{Ca}} \sigma^2_{\text{Ca}} / m \tag{4.41}
\]

\[
\text{Var} \left[ \delta^2_{Z|\text{Ca}} \right] = \left( \sigma^2_{2|\text{Ca}} + \mu^2_{2|\text{Ca}} \right)^2 \text{Var} \left[ S^2_{\text{Ca}} \right] + \sigma^4_{2|\text{Ca}} \text{Var} \left[ \bar{A}^2 \right]
+ 2 \left( \sigma^2_{2|\text{Ca}} + \mu^2_{2|\text{Ca}} \right) \sigma^2_{2|\text{Ca}} \text{Cov} \left[ S^2_{\text{Ca}}, \bar{A}^2 \right]. \tag{4.42}
\]

We consider the separate terms in Equation (4.42). We have

\[
\text{Var} \left[ S^2_{\text{Ca}} \right] = \frac{2\sigma^4_{\text{Ca}}}{m - 1} \approx \frac{2\sigma^4_{\text{Ca}}}{m} \tag{4.43}
\]

for large \( m \) by Theorem 3.6.1(c) of Hogg et al. [52]. We also have that \( S^2_{\text{Ca}} \) and \( \bar{A} \) are independent by Theorem 3.6.1(b) of Hogg et al. [52]; and therefore \( S^2_{\text{Ca}} \) and \( \bar{A}^2 \) are independent so that we have

\[
\text{Cov} \left[ S^2_{\text{Ca}}, \bar{A}^2 \right] = 0. \tag{4.44}
\]

Finally we obtain a large-sample expression for \( \text{Var} \left[ \bar{A}^2 \right] \) by the univariate delta method. In terms of the transformation

\[
\bar{A}^2 = g(\bar{A}) \tag{4.45}
\]

we have by the univariate delta method that

\[
\text{Var} \left[ \bar{A}^2 \right] = \left[ g'(\mu_{A}) \right]^2 \text{Var} \left[ \bar{A} \right]
= [2\mu_{A}]^2 \cdot \sigma^2_{\bar{A}} / m
= 4\mu^2_{A} \sigma^2_{\bar{A}} / m. \tag{4.46}
\]

Plugging (4.43), (4.44), and (4.46) into (4.42), we have

\[
\text{Var} \left[ \delta^2_{Z|\text{Ca}} \right] = 2 \left( \sigma^2_{2|\text{Ca}} + \mu^2_{2|\text{Ca}} \right)^2 \sigma^4_{\text{Ca}} / m + 4\sigma^4_{2|\text{Ca}} \mu^2_{A} \sigma^2_{\text{Ca}} / m. \tag{4.47}
\]
Thus our large-sample approximation to \( \text{Var} \left[ \hat{\sigma}^2_{Z|Ca} \right] \) has the form

\[
\text{Var} \left[ \hat{\sigma}^2_{Z|Ca} \right] = 2 \left( \sigma^2_{Z|Ca} + \mu^2_{Z|Ca} \right)^2 S^4_A / m + 4 \sigma^4_{Z|Ca} \bar{S}^2_A / m. (4.48)
\]

To finish the computation of \( \text{Var} \left[ \hat{F}_{Y|Ca}(\phi) \right] \) as specified by (4.38), we must compute \( \text{Cov} \left[ \hat{\mu}_{Z|Ca}, \hat{\sigma}^2_{Z|Ca} \right] \) as required in Equation (4.40). From (4.31) and (4.32) we have

\[
\text{Cov} \left[ \hat{\mu}_{Z|Ca}, \hat{\sigma}^2_{Z|Ca} \right] = \mu^2_{Z|Ca} \sigma^2_{Z|Ca} \text{Cov} \left[ \bar{A}, \bar{A}^2 \right] (4.49)
\]

because \( \text{Cov} \left[ \bar{A}, S^2_A \right] = 0 \) since \( \bar{A} \) and \( S^2_A \) are independent. We have

\[
\text{Cov} \left[ \bar{A}, \bar{A}^2 \right] = E \left[ \bar{A}^3 \right] - E \left[ \bar{A} \right] E \left[ \bar{A}^2 \right] (4.50)
\]

\[
\quad = E \left[ (\bar{A} - \mu_A)^3 \right] + 3 \left( (\bar{A} - \mu_A)^2 \right) E \left[ \bar{A} \right] + (E \left[ \bar{A} \right])^3 \quad (4.51)
\]

\[
\quad \quad - \mu_A (\sigma^2_A / m + \mu^2_A) \quad (4.52)
\]

\[
\quad = 3 \mu_A \sigma^2_A / m + \mu^3_A - \mu_A \sigma^2_A / m - \mu^3_A \quad (4.53)
\]

\[
\quad = 2 \mu_A \sigma^2_A / m, \quad (4.54)
\]

because we have

\[
E \left[ (\bar{A} - \mu_A)^3 \right] = 0 \quad (4.55)
\]

by the symmetry of the normal density function about its mean.

Putting together (4.38), (4.39), and (4.40), we finally obtain a large-sample formula for \( \text{Var} \left[ \hat{F}_{Y|Ca}(\phi) \right] \).

To simplify the resulting expressions, we write

\[
z_{\phi} \equiv \frac{\ln (\phi) - \mu_{Z|Ca}}{\sigma_{Z|Ca}} (4.56)
\]

so that we have

\[
\nabla F_{Y|Ca}(\phi) = - \frac{\exp \left( -\frac{1}{2} z_{\phi}^2 \right)}{\sigma_{Z|Ca} \sqrt{2\pi}} \left[ 1, \ z_{\phi} / (2\sigma_{Z|Ca}) \right] (4.57)
\]

99
and

\[
\Sigma \left( \tilde{\mu}_{Z|\text{Ca}}, \tilde{\sigma}_{Z|\text{Ca}}^2 \right) = \begin{bmatrix}
\mu_{2|\text{Ca}}^2 \sigma_A^2 / m \\
2 \mu_A \sigma_A^2 / m \\
2 \mu_A \sigma_A^2 / m \\
2 \left( \sigma_{2|\text{Ca}}^2 + \mu_{2|\text{Ca}}^2 \right)^2 \sigma_A^4 / m \\
+ 4 \sigma_{2|\text{Ca}}^2 \mu_A^2 \sigma_A^2 / m
\end{bmatrix}.
\] (4.58)

Thus we see from (4.57) and (4.58) that

\[
\text{Var} \left[ \hat{F}_{Y|\text{Ca}}(\phi) \right] = \frac{\exp(-z_{\phi}^2)}{m \sigma_{Z|\text{Ca}}^2} \times \left\{ \mu_{2|\text{Ca}}^2 \sigma_A^2 + 2 z_{\phi} \mu_A \sigma_A^2 / \sigma_Z|\text{Ca} \\
+ z_{\phi}^2 \left( \sigma_{2|\text{Ca}}^2 + \mu_{2|\text{Ca}}^2 \right)^2 \sigma_A^4 / \left( 2 \sigma_{Z|\text{Ca}}^2 \right) \\
+ z_{\phi}^2 \sigma_{2|\text{Ca}}^2 \mu_A^2 \sigma_A^2 / \sigma_Z|\text{Ca} \right\}.
\] (4.59)

In terms of the simplifying term

\[
\hat{z}_{\phi} = \frac{\ln (\phi) - \tilde{\mu}_{Z|\text{Ca}}}{\tilde{\sigma}_{Z|\text{Ca}}},
\] (4.62)

we have the following large-sample estimator for \( \text{Var} \left[ \hat{F}_{Y|\text{Ca}}(\phi) \right] \):

\[
\widehat{\text{Var}} \left[ \hat{F}_{Y|\text{Ca}}(\phi) \right] = \frac{\exp(-\hat{z}_{\phi}^2)}{m \widehat{\sigma}_{Z|\text{Ca}}^2} \times \left\{ \mu_{2|\text{Ca}}^2 S_A^2 + 2 \hat{z}_{\phi} \bar{A} S_A^2 / \tilde{\sigma}_{Z|\text{Ca}} \\
+ \hat{z}_{\phi}^2 \left( \sigma_{2|\text{Ca}}^2 + \mu_{2|\text{Ca}}^2 \right)^2 S_A^4 / \left( 2 \tilde{\sigma}_{Z|\text{Ca}}^2 \right) \\
+ \hat{z}_{\phi}^2 \sigma_{2|\text{Ca}}^2 \bar{A}^2 S_A^2 / \tilde{\sigma}_{Z|\text{Ca}} \right\}.
\] (4.63)

In the following derivation of a CI for \( \phi \) based on \( \hat{\pi}_{\text{Ca}}, \hat{\pi}_{\text{NC}}, \hat{b}(\phi), S_{\hat{b}(\phi)}^2, \bar{G}(\phi), S_{\bar{G}(\phi)}^2, \bar{H}(\phi), S_{\bar{H}(\phi)}^2, \bar{A}, \) and \( S_A^2 \), we will repeatedly use the following intermediate results.
RESULT IR₁: If $X, Y$ are independent random variables, then

$$\text{Var}[XY] = \text{Var}[X] \text{Var}[Y] + (E[X])^2 \text{Var}[Y] + (E[Y])^2 \text{Var}[X]. \quad (4.66)$$

RESULT IR₂: If $X, Y, Z$ are independent random variables, then

$$\text{Cov}[XY, (1 - X)Z] = -E[Y]E[Z] \text{Var}[X]. \quad (4.67)$$

Proof.

$$= E\left\{YZE[X(1 - X)] - YE[X]ZE[1 - X]\right\}$$
$$= E[X]E[Y]E[Z] - E[Y]E[Z]\left\{\text{Var}[X] + (E[X])^2\right\}$$
$$= -E[Y]E[Z] \text{Var}[X]. \quad (4.68)$$

RESULT IR₃: If $X, Y, Z$ are independent random variables, then

$$\text{Cov}[XY, XZ] = E[Y]E[Z] \text{Var}[X]. \quad (4.69)$$

Proof. The proof is similar to that of Result IR₂. ■

Based on Equation (4.11) we formulate the following point estimator of $\hat{\delta}(\phi)$:

$$\hat{\delta}(\phi) = \hat{\pi}_C\hat{G}(\phi) + \hat{\pi}_N\bar{H}(\phi) - \hat{\pi}_C\hat{F}_Y|_{Ca}(\phi)\hat{E}(\phi) + \hat{\pi}_N\hat{F}_Y|_{NC}(\phi)\delta. \quad (4.70)$$
We have

\[
\text{Var}[\hat{\theta}(\phi)] = \text{Var}[\hat{\pi}_\text{Ca}\hat{G}(\phi)] + \text{Var}[\hat{\pi}_\text{NC}\hat{H}(\phi)] + \text{Var}[\hat{\pi}_\text{Ca}\hat{F}_{Y|\text{Ca}}\hat{\delta}(\phi)] + \text{Var}[\hat{\pi}_\text{NC}\hat{F}_{Y|\text{NC}}(\phi)\delta] \\
+ 2\text{Cov}[\hat{\pi}_\text{Ca}\hat{G}(\phi), \hat{\pi}_\text{NC}\hat{H}(\phi)] - 2\text{Cov}[\hat{\pi}_\text{Ca}\hat{G}(\phi), \hat{\pi}_\text{Ca}\hat{F}_{Y|\text{Ca}}\hat{\delta}(\phi)] \\
+ 2\text{Cov}[\hat{\pi}_\text{NC}\hat{H}(\phi), \hat{\pi}_\text{NC}\hat{F}_{Y|\text{NC}}(\phi)\delta] - 2\text{Cov}[\hat{\pi}_\text{NC}\hat{H}(\phi), \hat{\pi}_\text{Ca}\hat{F}_{Y|\text{Ca}}\hat{\delta}(\phi), \hat{\pi}_\text{NC}\hat{F}_{Y|\text{NC}}(\phi)\delta].
\]

(4.71)

We evaluate expressions $\mathcal{V}_1$ through $\mathcal{V}_{10}$ separately. First we evaluate $\mathcal{V}_1$:

\[
\mathcal{V}_1 = \text{Var}[\hat{\pi}_\text{Ca}] \text{Var}[\hat{G}(\phi)] + (\mathbb{E} [\hat{\pi}_\text{Ca}])^2 \text{Var}[\hat{G}(\phi)] + (\mathbb{E} [\hat{G}(\phi)])^2 \text{Var}[\hat{\pi}_\text{Ca}].
\]

Therefore we estimate $\mathcal{V}_1$ by the following statistic,

\[
\hat{\mathcal{V}}_1 = \hat{\pi}_\text{Ca}(1 - \hat{\pi}_\text{Ca})S_{G(\phi)}^2/m^2 + \hat{\pi}_\text{Ca}^2 S_{G(\phi)}^2/m + [\hat{G}(\phi)]^2 \hat{\pi}_\text{Ca}(1 - \hat{\pi}_\text{Ca})/m.
\]

Similarly we see that

\[
\mathcal{V}_2 = \text{Var}[\hat{\pi}_\text{NC}] \text{Var}[\hat{H}(\phi)] + (\mathbb{E} [\hat{\pi}_\text{NC}])^2 \text{Var}[\hat{H}(\phi)] + (\mathbb{E} [\hat{H}(\phi)])^2 \text{Var}[\hat{\pi}_\text{NC}],
\]

so that we have the following variance estimator,

\[
\hat{\mathcal{V}}_2 = \hat{\pi}_\text{Ca}(1 - \hat{\pi}_\text{Ca})S_{H(\phi)}^2/m^2 + (1 - \hat{\pi}_\text{Ca})^2 S_{H(\phi)}^2/m + [\hat{H}(\phi)]^2 \hat{\pi}_\text{Ca}(1 - \hat{\pi}_\text{Ca})/m.
\]
Next we evaluate $\mathbb{V}_3$. Since $\hat{\pi}_{Ca}$, $\hat{F}_{Y|Ca}$, and $\hat{b}(\phi)$ are independent random variables, we have

$$
\mathbb{V}_3 = E\left[\hat{\pi}_{Ca}^2 \left[\hat{F}_{Y|Ca}(\phi)\right]^2 \left[\hat{b}(\phi)\right]^2\right] - \left(E\left[\hat{\pi}_{Ca} \hat{F}_{Y|Ca}(\phi) \hat{b}(\phi)\right]\right)^2
$$

$$
= E\left[\hat{\pi}_{Ca}^2\right] E\left\{\left[\hat{F}_{Y|Ca}(\phi)\right]^2\right\} E\left\{\left[\hat{b}(\phi)\right]^2\right\} - \left\{E\left[\hat{\pi}_{Ca}\right] E\left[\hat{F}_{Y|Ca}(\phi)\right] E\left[\hat{b}(\phi)\right]\right\}^2
$$

$$
= \left[\pi_{Ca}(1 - \pi_{Ca})/m + \pi_{Ca}^2\right] \times \left\{\text{Var}\left[\hat{F}_{Y|Ca}(\phi)\right] + \left[\hat{F}_{Y|Ca}(\phi)\right]^2\right\}
$$

$$
\times \left\{\text{Var}\left[\hat{b}(\phi)\right] + \left[\hat{b}(\phi)\right]^2\right\} - \left[\pi_{Ca} F_{Y|Ca}(\phi) \hat{b}(\phi)\right]^2
$$

so that we have the corresponding estimator of $\mathbb{V}_3$:

$$
\hat{\mathbb{V}}_3 = \left[\hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m + \hat{\pi}_{Ca}^2\right]
$$

$$
\times \left\{\text{Var}\left[\hat{F}_{Y|Ca}(\phi)\right] + \left[\hat{F}_{Y|Ca}(\phi)\right]^2\right\}
$$

$$
\times \left\{\text{Var}\left[\hat{b}(\phi)\right] + \left[\hat{b}(\phi)\right]^2\right\}
$$

$$
- \left[\hat{\pi}_{Ca} \hat{F}_{Y|Ca}(\phi) \hat{b}(\phi)\right]^2
$$

(4.75)

Next we evaluate $\mathbb{V}_4$. We have

$$
\mathbb{V}_4 = \left[F_{Y|NC}(\phi)\right]^2 \delta^2 \text{Var}[\hat{\pi}_{NC}]
$$

$$
= \left[\delta F_{Y|NC}(\phi)\right]^2 \pi_{NC}(1 - \pi_{NC})/m,
$$

so that the corresponding estimator of $\mathbb{V}_4$ is

$$
\hat{\mathbb{V}}_4 = \left[\delta F_{Y|NC}(\phi)\right]^2 \hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m.
$$

Next we evaluate $\mathbb{V}_5$. By Equation (4.67), we have

$$
\mathbb{V}_5 = -2E\left[\hat{G}(\phi)\right] E\left[\hat{H}(\phi)\right] \pi_{Ca}(1 - \pi_{Ca})/m
$$

103
so that the corresponding estimator of $\mathbb{V}_5$ is

$$\hat{\mathbb{V}}_5 = -2\bar{G}(\phi)\bar{H}(\phi)\hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m.$$ 

Next we evaluate $\mathbb{V}_6$. By Equation (4.69), we have

$$\mathbb{V}_6 = -2\mathbb{E}\left[\bar{G}(\phi)\right]\mathbb{E}\left[\hat{F}_{Y|Ca}(\phi)\right]\mathbb{E}\left[\hat{\delta}(\phi)\right]\text{Var}[\hat{\pi}_{Ca}]$$

$$= -2\mathbb{E}\left[\bar{G}(\phi)\right]F_{Y|Ca}(\phi)\mathbb{E}\left[\hat{\delta}(\phi)\right]\pi_{Ca}(1 - \pi_{Ca})/m$$

so that the corresponding estimator of $\mathbb{V}_6$ is

$$\hat{\mathbb{V}}_6 = -2\hat{F}_{Y|Ca}(\phi)\cdot \bar{G}(\phi)\hat{\delta}(\phi)\hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m.$$ 

Next we evaluate $\mathbb{V}_7$. By Equation (4.67), we have

$$\mathbb{V}_7 = -2\mathbb{E}\left[\bar{G}(\phi)\right]\cdot F_{Y|NC}(\phi)\delta \cdot \text{Var}[\hat{\pi}_{Ca}]$$

$$= -2F_{Y|NC}(\phi)\delta\mathbb{E}\left[\bar{G}(\phi)\right]\pi_{Ca}(1 - \pi_{Ca})/m,$$

so that the corresponding estimator of $\mathbb{V}_7$ is

$$\hat{\mathbb{V}}_7 = -2F_{Y|NC}(\phi)\delta\bar{G}(\phi)\hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m.$$ 

Next we evaluate $\mathbb{V}_8$. By Equation (4.67), we have

$$\mathbb{V}_8 = 2\mathbb{E}\left[\bar{H}(\phi)\right]\mathbb{E}\left[\hat{\delta}(\phi)\right]\mathbb{E}\left[F_{Y|Ca}(\phi)\right]\text{Var}[\hat{\pi}_{Ca}]$$

$$= 2F_{Y|Ca}(\phi)\mathbb{E}\left[\bar{H}(\phi)\right]\mathbb{E}\left[\hat{\delta}(\phi)\right]\pi_{Ca}(1 - \pi_{Ca})/m.$$
so that the corresponding estimator of $\mathbb{V}_8$ is

$$
\hat{\mathbb{V}}_8 = 2\hat{F}_{Y|Ca}(\phi)\bar{H}(\phi)\hat{b}(\phi)\hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m.
$$

Next we evaluate $\mathbb{V}_9$. By Equation (4.69), we have

$$
\mathbb{V}_9 = 2E[\bar{H}(\phi)] F_{Y|NC}(\phi) \delta \text{Var}[\hat{\pi}_{Ca}]
= 2\delta F_{Y|NC}(\phi) \cdot E[\bar{H}(\phi)] \pi_{Ca}(1 - \pi_{Ca})/m,
$$

so that the corresponding estimator of $\mathbb{V}_9$ is

$$
\hat{\mathbb{V}}_9 = 2\delta F_{Y|NC}(\phi)\bar{H}(\phi)\hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m.
$$

Finally we evaluate $\mathbb{V}_{10}$. By Equation (4.68), we have

$$
\mathbb{V}_{10} = -2\delta \cdot E\left[\hat{F}_{Y|Ca}(\phi)\right] F_{Y|NC}(\phi) \times \text{Cov}\left[\hat{\pi}_{Ca}\hat{b}(\phi), 1 - \hat{\pi}_{Ca}\right]
= -2\delta \cdot F_{Y|Ca}(\phi) F_{Y|NC}(\phi) \times \text{Cov}\left[\hat{\pi}_{Ca}\hat{b}(\phi), -\hat{\pi}_{Ca}\right]. \tag{4.76}
$$

We have

$$
\text{Cov}\left[\hat{\pi}_{Ca}\hat{b}(\phi), -\hat{\pi}_{Ca}\right] = -E\left[\hat{\pi}_{Ca}^2 \hat{b}(\phi)\right] + E\left[\hat{\pi}_{Ca} \hat{b}(\phi)\right] E[\hat{\pi}_{Ca}]
= -E\left[\hat{\pi}_{Ca}^2\right] E\left[\hat{b}(\phi)\right] + (E[\hat{\pi}_{Ca}])^2 E\left[\hat{b}(\phi)\right]
= -\left\{\pi_{Ca}(1 - \pi_{Ca})/m + \pi_{Ca}^2\right\} b(\phi) + \pi_{Ca}^2 b(\phi)
= -\pi_{Ca}(1 - \pi_{Ca}) b(\phi)/m. \tag{4.77}
$$

Plugging Equation (4.77) into (4.76), we get the following expression:

$$
\mathbb{V}_{10} = 2\delta F_{Y|Ca}(\phi) F_{Y|NC}(\phi) \pi_{Ca}(1 - \pi_{Ca}) \phi(\phi)/m.
$$
so that we have the following estimator of $\mathbb{V}_{10}$:

$$\hat{\mathbb{V}}_{10} = 2\delta \hat{F}_{Y|Ca}(\phi) F_{Y|NC}(\phi) \hat{b}(\phi) \hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m.$$ 

Putting all this together, we have the variance estimator

$$\widehat{\text{Var}} \left[ \hat{\theta}(\phi) \right] = \sum_{\ell=1}^{10} \hat{\mathcal{V}}_{\ell}$$

$$= \hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca}) S_{G(\phi)}^2 / m^2 + \hat{\pi}_{Ca}^2 / m + \left[ \hat{G}(\phi) \right]^2 \hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m$$

$$+ \hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca}) S_{H(\phi)}^2 / m^2 + (1 - \hat{\pi}_{Ca})^2 S_{H(\phi)}^2 / m + \left[ \hat{H}(\phi) \right]^2 \hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m$$

$$+ \left( \hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca}) / m + \hat{S}^2_{Ca} \right) \left( \text{Var} \left[ \hat{F}_{Y|Ca}(\phi) \right] + \left[ \hat{F}_{Y|Ca}(\phi) \right]^2 \right)$$

$$\times \left( \frac{S^2_{b(\phi)}}{m} + \left[ \hat{b}(\phi) \right]^2 \right)$$

$$- \left[ \hat{\pi}_{Ca} \hat{F}_{Y|Ca}(\phi) \hat{b}(\phi) \right]^2$$

$$+ \left[ \delta F_{Y|NC}(\phi) \right]^2 \hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m$$

$$- 2 \hat{G}(\phi) \hat{H}(\phi) \hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m$$

$$- 2 \hat{F}_{Y|Ca}(\phi) \cdot \hat{G}(\phi) \hat{b}(\phi) \hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m$$

$$- 2 F_{Y|NC}(\phi) \delta \hat{G}(\phi) \hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m$$

$$+ 2 \hat{F}_{Y|Ca}(\phi) \hat{H}(\phi) \hat{b}(\phi) \hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m$$

$$+ 2 \delta F_{Y|NC}(\phi) \hat{H}(\phi) \hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m$$

$$+ 2 \delta \hat{F}_{Y|Ca}(\phi) F_{Y|NC}(\phi) \hat{b}(\phi) \hat{\pi}_{Ca}(1 - \hat{\pi}_{Ca})/m.$$ 

Finally, our approximate $100(1 - \alpha)\%$ CI estimator for $\hat{\theta}(\phi)$ is

$$\hat{\hat{\theta}}(\phi) \pm \hat{\delta}^{-1/2} \sqrt{\widehat{\text{Var}} \left[ \hat{\theta}(\phi) \right]},$$

(4.78)
where \( \tilde{z}_{1-\alpha/2} \) is the \( 1 - \alpha/2 \) quantile of the standard normal distribution.

In summary, we have shown that the simple single-period conceptual model of PSA screening can be analyzed as an optimization problem, namely, the problem of finding

\[
\phi^* = \arg \max_\phi \vartheta(\phi),
\]

which requires only the distributions of PSA conditional on whether PCa is present (\( F_{Y|C_a}(\cdot) \) and \( F_{Y|NC}(\cdot) \)), the expected benefit of biopsy (\( b \)), the expected disutility of biopsy (\( \delta \)), and the probability of having PCa (\( \pi_{C_a} \)). And we have derived point and CI estimators for the expected reward function \( \vartheta(\phi) \) that quantify the variability in the underlying parameters of the single-period model for large parameter-estimation sample sizes. This single-period analytical model does not, of course, capture important complexities of screening strategies with multiple potential screen events, but we offer it as a potential tool with which to derive meaningful insight into the real PCa screening problem. As an obvious example, comparing results from the single-period model to results from the three-stage method described below may shed light on the value added by having repeat PSA screenings. Also, results depicting the effect of variation in certain key model parameters on the function \( \vartheta(\phi) \) may provide a clearer picture of how certain physiological processes affect optimal PSA screening than we could hope to see from less transparent results from a random-search based metaheuristic applied to a simulation optimization problem.

### 4.2.2 QALYs Parameterized Simulation Model

The simulation model is based on the underlying discrete-event stochastic process described in Chapter 3. Here we develop a QALYs-based parameterization of the simulation model from Chapter 3. We then define the optimization problem for finding improved screening strategies based on QALYs.

Simulated patients accumulate rewards in QALYs during each period they are alive based on their health state and the clinical decisions made or previously made. For each year that a patient is alive, the patient earns a reward of 1 appropriately reduced by the applicable disutilities (if there are any).

There are disutilities in our model associated with PSA screening, prostate biopsy, the diagnosis of
Table 4.1: QALY disutilities in the PCa screening simulation model.

<table>
<thead>
<tr>
<th>Disutility</th>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening</td>
<td>$D_{\text{Scr}}$</td>
<td>Instantaneous disutility for PSA screening</td>
</tr>
<tr>
<td>Biopsy</td>
<td>$D_{\text{Biop}}$</td>
<td>Instantaneous disutility for prostate biopsy</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>$D_{\text{Dia}}$</td>
<td>Instantaneous disutility for the diagnosis of PCa</td>
</tr>
<tr>
<td>Treatment</td>
<td>$D_{\text{Tre}}$</td>
<td>Instantaneous disutility for definitive treatment</td>
</tr>
<tr>
<td>Recovery</td>
<td>$D_{\text{Rec}}$</td>
<td>Annual disutility for recovery from definitive treatment applied during the posttreatment recovery period</td>
</tr>
<tr>
<td>Metastasis</td>
<td>$D_{\text{Met}}$</td>
<td>Annual disutility for metastatic PCa</td>
</tr>
</tbody>
</table>

PCa, definitive treatment, posttreatment recovery, and living with metastatic PCa; the disutilities are listed in Table 4.1. The disutilities for screening ($D_{\text{Scr}}$), biopsy ($D_{\text{Biop}}$), and diagnosis ($D_{\text{Dia}}$) are instantaneous disutilities; these disutilities are one-time quality-of-life decrements that correspond to a specific period. The posttreatment recovery disutility ($D_{\text{Rec}}$) is an annual quality-of-life decrement that is applied in any time period during the posttreatment recovery period, other than the first year, immediately following definitive treatment. The metastasis disutility ($D_{\text{Met}}$) is an annual quality-of-life decrement that is applied in any time period in which the patient has metastatic PCa (i.e. any period in which the patient is in state M).

4.2.3 Number Needed to Screen to Gain 1 QALY (NNSQ)

It is common to evaluate population screening and intervention programs in terms of expected QALYs [48, 63]. The benefit of using QALYs is the ability to account for a range of desirable and undesirable effects on patient life-experience stemming from comorbidities and from the side effects of medicine and the discomfort of various clinical procedures. Another common mortality measure is NNS. A chief advantage of this measure is that it expresses directly the number of patients that must be screened to achieve a unit reduction in disease mortality. In this way, it is a measure of the efficiency of the screening program at saving lives.

In this chapter we employ a relatively new measure (estimated but not developed in the context of variance reduction in Zhang et al. [112]) that combines the advantages of using expected QALYs and the advantages of using NNS. While NNS only indicates screening effort needed to effectuate a unit
reduction in mortality, by using QALYs we can take into account benefits not measured by mortality, such as the benefit of avoiding unnecessary prostate biopsies. We refer to this relatively new measure, NNSQ, as the number needed to screen to produce a net gain of 1 QALY. Similarly to NNS, NNSQ is a measure of efficiency: NNSQ expresses directly the amount of screening effort required to achieve a unit increase in QALYs.

To estimate NNSQ for a given screening strategy, we first simulate a population of patients who adhere to the screening strategy, and then simulate the exactly same population in the absence of screening. We can then compare the QALYs gained by each simulated patient under the given strategy and under no screening in the following manner.

From a simulated population of men following a given screening strategy, we simulate a random sample of \( m \) patients; and for the \( j \)th such patient, let

\[
Q_{S,j} \equiv \begin{cases} 
\text{Total QALYs accumulated for patient } j \\
\text{under screening (S),}
\end{cases}
\]

and

\[
Q_{NS,j} \equiv \begin{cases} 
\text{Total QALYs accumulated for patient } j \\
\text{under no screening (NS).}
\end{cases}
\]

Let \( Q \) be a random variable denoting the QG from screening under the given strategy compared with no screening. With the setup in Equations (4.79) and (4.80), we define the QALYs difference for patient \( j \) as

\[
Q_j \equiv Q_{S,j} - Q_{NS,j} \text{ for } j = 1, \ldots, m.
\]

The corresponding sample mean and variance are computed in the usual fashion:

\[
\bar{Q} = \frac{1}{m} \sum_{j=1}^{m} Q_j
\]

(4.82)

\[
S_{Q}^2 = \frac{1}{m-1} \sum_{j=1}^{m} (Q_j - \bar{Q})^2.
\]

(4.83)
Now, we formally define the expected NNSQ as

\[
NNSQ = \frac{1}{E[Q]}.
\]  

(4.84)

Our point estimate of NNSQ is

\[
\hat{NNSQ} = \frac{1}{\bar{Q}}.
\]  

(4.85)

and an asymptotically valid 100(1 − \(\alpha\))% CI for NNSQ is

\[
\left[ \frac{1}{\bar{Q}} \pm \left( \tilde{z}_{1-\alpha/2} \right) \frac{S_Q}{\bar{Q}^2 \sqrt{m}} \right],
\]  

(4.86)

where \(\tilde{z}_{1-\alpha/2}\) is the \(1 - \alpha/2\) quantile of the standard normal distribution. The CI in Equation (4.86) is derived in the same way that we derived the CI in Equation (3.7).

4.2.4 Prostate Cancer Screening Model

We use the model described in Chapter 3 to simulate patients going through a clinical process of screening for PCa. The model assumes that every positive prostate biopsy entails immediate definitive treatment by radical prostatectomy. The model also assumes that state M is fully observable because of the likely severity of the attendant symptoms of metastatic PCa; and therefore if a patient transitions into state M, then he is ineligible for further screening. Based on the results of the two stages of decision making, and on other factors such as the patient’s underlying health state, the patient will accrue rewards in QALYs over time as he ages. Each period, the patient makes a health state transition based on the underlying Markov model. The simulation of a single period for a given patient is illustrated in Figure 4.1.

Figure 4.1: Sequence of events in the simulation model in a given time period. PSA testing, prostate biopsy, and definitive PCa treatment take place at the beginning of the period.
After the two stages of decision making are resolved, sufficient information is known to compute the reward for the patient in the current period. The reward for a patient at age \( t \) as a function of the patient’s state \((s_t)\) and clinical actions in the first and second decision-making stages \((a_t)\) for a period of length 1 year is

\[
r(s_t, a_t) = 1 - \Delta_{\text{Scr}}(\rho_t) - \Delta_{\text{Biop}}(\eta_t) - \Delta_{\text{Dia}}(\eta_1, \ldots, \eta_t, s_{t-1}, s_t) - \Delta_{\text{Tre}}(s_{t-1}, s_t) - \Delta_{\text{Rec}}(s_1, \ldots, s_t) - \Delta_{\text{Met}}(s_{t-1}),
\]

(4.87)

where:

\[
\Delta_{\text{Scr}}(\rho_t) = \begin{cases} 
\mathbb{1}_{\text{Scr}} & \text{if the (first-stage) decision to PSA test is made in current period,} \\
0 & \text{otherwise;}
\end{cases}
\]

\[
\Delta_{\text{Biop}}(\eta_t) = \begin{cases} 
\mathbb{1}_{\text{Biop}} & \text{if the (second-stage) decision to biopsy is made in the current period,} \\
0 & \text{otherwise;}
\end{cases}
\]

\[
\Delta_{\text{Dia}}(\eta_1, \ldots, \eta_t, s_{t-1}, s_t) = \begin{cases} 
\mathbb{1}_{\text{Dia}} & \text{if the patient just transitioned into state T or M from state NC or C,} \\
0 & \text{otherwise;}
\end{cases}
\]

\[
\Delta_{\text{Tre}}(s_{t-1}, s_t) = \begin{cases} 
\mathbb{1}_{\text{Tre}} & \text{if the patient just underwent treatment (i.e., if the patient just transitioned into state T from state NC or C),} \\
0 & \text{otherwise;}
\end{cases}
\]

\[
\Delta_{\text{Rec}}(s_1, \ldots, s_t) = \begin{cases} 
\mathbb{1}_{\text{Rec}} & \text{if the patient underwent definitive treatment in any of the most recent periods (but not the current period) within the posttreatment recovery period,} \\
0 & \text{otherwise;}
\end{cases}
\]
and

$$\Delta_{\text{Met}}(s_{t-1}, s_t) = \begin{cases} D_{\text{Met}} & \text{if the patient spent the most-recent previous period in state M,} \\ 0 & \text{otherwise.} \end{cases}$$

The reward function in Equation (4.87) defines how QALYs are accumulated for a given patient under a given screening strategy. In particular, the observed values of $Q_{S,j}$ and $Q_{NS,j}$ from Equations (4.79) and (4.80), respectively, are obtained by evaluation of the reward function for a patient $j$ under a given screening strategy and in the absence of screening, respectively.

Functionally, $\Delta_{\text{Scr}}$ depends on $\rho_t$ because screening has occurred at age $t$ if and only if $\rho_t$ does not have the “no PSA result” outcome. That is, the disutility $D_{\text{Scr}}$ applies if a PSA test is performed at age $t$. Similarly, $\Delta_{\text{Biop}}$ depends on $\eta_t$ because a prostate biopsy was performed at age $t$ if and only if $\eta_t$ does not have the outcome “no biopsy result.” That is, the disutility $D_{\text{Biop}}$ applies if a biopsy is performed at age $t$.

The dependence of $\Delta_{\text{Dia}}$ on $\eta_1, \ldots, \eta_t, s_{t-1}, s_t$ is slightly more complex. Diagnosis can occur because of screen detection, metastasis of hitherto undetected PCa, or clinical detection resulting in treatment. (Although in the real world it is possible to be diagnosed through a positive biopsy and then to decline treatment, since we assume in our model that all screen-detected PCa cases are treated immediately, this is not a scenario that we consider.) These three ways in which diagnosis can occur can be collectively captured by the following condition:

$$s_{t-1} \in \{\text{NC, C}\} \text{ and } s_t \in \{\text{T, M}\}.$$ 

We assume that transitioning to the metastasis state results in diagnosis if diagnosis has not yet occurred because the symptoms of metastatic PCa are severe and noticeable.

Both $\Delta_{\text{Tre}}$ and $\Delta_{\text{Rec}}$ are related to definitive treatment of PCa. The random variable $\Delta_{\text{Tre}}$ depends on $s_{t-1}, s_t$ because we can determine whether a patient has undergone definitive treatment during the period commencing at age $t$ if and only if the patient transitioned from a nontreatment state at age $t - 1$ to the
treatment state (T) at age $t$. In our particular health-state-space definition, this transition will necessarily be a transition from $s_{t-1} = C$ to $s_t = T$ (refer to the model assumptions described in Chapter 3). The disutility $D_{\text{tre}}$ represents the immediate and short-term discomfort, possible complications, and side effects of radical prostatectomy. On the other hand, the disutility $D_{\text{rec}}$ represents the longer-term recovery interval during which there may be lingering discomfort or side effects such as urinary and sexual dysfunction. In Heijnsdijk et al. [48], this posttreatment recovery interval is defined as the first 9 years after the year in which treatment was performed. So, by examining $s_1, \ldots, s_t$, we can determine whether there exists some $t' \in \{t - 9, t - 8, \ldots, t - 1\}$ such that $s_{t'} = T$ and $s_{t' - 1} \neq T$; if such a $t'$ exists, then the patient is within the posttreatment recovery interval at age $t$, and the disutility $D_{\text{rec}}$ applies. Note that because the earliest age at which a patient can possibly be screened in our simulation model is by assumption age 40 (described below in Section 4.2.5), there is no need to check boundary conditions: $t' - 1$ will always correspond to a valid age in the patient’s life.

The dependence of $\Delta_{\text{met}}$ on $s_{t-1}$ and $s_t$ is simple. If $s_{t-1} = C$ and $s_t = M$, then the disutility $D_{\text{met}}$ is incurred at time $t$. That is, for any period during which the patient lives in state $M$, the disutility for metastasis applies. This disutility represents the array of potential harms and discomforts that can accompany metastatic PCa.

In summary, the simulation model in Chapter 3 has been parameterized to account for QALYs and to estimate expected QG (and expected NNSQ). The modified version of Algorithm 1 that accounts for QALYs is detailed in pseudocode in Algorithm 2 below, which in turn utilizes Algorithm 3.
Algorithm 2: \texttt{SIMULATIONQALYS}(m)

\textbf{Input}: Number $m$ of patients to simulate.

1 \textbf{begin}
2 \textbf{foreach} patient $j \in \{1, \ldots, m\}$ \textbf{do}
3 \hfill \textbf{foreach} subsim in \{Screening, No Screening\} \textbf{do}
4 \hfill \hfill Set age $\leftarrow 40$;
5 \hfill \hfill Set state $\leftarrow \text{NC}$;
6 \hfill \hfill Set totalQALYs $\leftarrow 0$;
7 \hfill \hfill Set ageAtTreatment $\leftarrow -1$;
8 \hfill \hfill Set screen $\leftarrow \text{false}$;
9 \hfill \hfill Set biopsy $\leftarrow \text{false}$;
10 \hfill \textbf{while} state $\neq D$ \textbf{do}
11 \hfill \hfill Compute patient PSA level;
12 \hfill \hfill Check whether to start leadtime clock;
13 \hfill \hfill Set lastState $\leftarrow$ state;
14 \hfill \hfill \textbf{if} Clinical detection conditions met \textbf{then}
15 \hfill \hfill \hfill Set state $\leftarrow T$;
16 \hfill \hfill \textbf{else}
17 \hfill \hfill \hfill \textbf{if} Screen this period \textbf{then}
18 \hfill \hfill \hfill \hfill Set screen $\leftarrow \text{true}$;
19 \hfill \hfill \hfill \hfill \textbf{if} PSA level $\geq$ threshold \textbf{then}
20 \hfill \hfill \hfill \hfill \hfill Set biopsy $\leftarrow \text{true}$;
21 \hfill \hfill \hfill \hfill \hfill Transition state using $P^B$;
22 \hfill \hfill \hfill \hfill \textbf{else}
23 \hfill \hfill \hfill \hfill \hfill Transition state using $P^{NB}$;
24 \hfill \hfill \textbf{end}
25 \hfill \hfill \textbf{else}
26 \hfill \hfill \hfill Transition state using $P^{NB}$;
27 \hfill \textbf{end}
28 \hfill \hfill \textbf{end}
29 \hfill \textbf{end}
30 \hfill \textbf{end}
31 \hfill \textbf{end}
32 \hfill \textbf{if} state$\neq D$ \textbf{then}
33 \hfill \hfill \textbf{if} lastState$\neq T$ and state$=T$ \textbf{then}
34 \hfill \hfill \hfill Set ageAtTreatment $\leftarrow$ age;
35 \hfill \hfill \textbf{end}
36 \hfill \hfill Set totalQALYs $\leftarrow$ totalQALYs +
37 \hfill \hfill \texttt{PERIODQALYs}(screen, biopsy, age, ageAtTreatment, state, lastState);
38 \hfill \hfill Set age $\leftarrow$ age$+1$;
39 \hfill \textbf{end}
40 \hfill \textbf{end}
41 \textbf{end}
42 \textbf{end}
43
44 \textbf{end}
45
46 114
Algorithm 3: **PERIODQALYs**

**Input:** Whether PSA test performed this period; whether biopsy performed this period; age during period; age when treatment was performed; state transitioned to during this period; and the state transitioned from during this period.

1. begin
2. | Set periodQALYs ← 1;
3. | if screen then
4. | | Set periodQALYs ← periodQALYs−DScr;
5. | end
6. if lastState=C and (state=T or state=M) then
7. | Set periodQALYs ← periodQALYs−DDia;
8. end
9. if biopsy then
10. | Set periodQALYs ← periodQALYs−DBiop;
11. end
12. if age=ageAtTreatment then
13. | Set periodQALYs ← periodQALYs−DTre;
14. else
15. | if state=T and (age−ageAtTreatment ≤ 9) then
16. | | Set periodQALYs ← periodQALYs−DRec;
17. end
18. end
19. if state=M then
20. | Set periodQALYs ← periodQALYs−DMet;
21. end
22. return periodQALYs;
23. end

### 4.2.5 Optimization Model

As a result of making decision \( a_t \in A \) for a patient in state \( s_t \in S \) at age (epoch) \( t \), the patient receives the reward \( r(s_t, a_t) \) in QALYs, as defined in Equation (4.87). We define the \( N \)-element vector of decision variables

\[
x = [x_1, x_2, \ldots, x_N],
\]

where

\[
x_t \in \{0, 1, \ldots, 21\} \text{ for } t = 1, \ldots, N.
\]

In this model, we assume that screening may take place as frequently as annually, and that screening may begin no earlier than at age 40 years and may continue no later than at age 80 years. Thus, \( N = 41 \), and there is a decision variable \( x_t \) for each of ages 40, 41, \ldots, 80 years; in other words, the decision variable
Let \( \Theta \) denote the set of all \( x \) consistent with Equations (4.88) and (4.89) and the assumption that \( N = 41 \). Each \( x_t \) defines whether to PSA-screen eligible patients at epoch \( t \); and, if so, \( x_t \) also defines the PSA threshold to be used at that epoch. (By eligible patients, we simply mean patients who are still subject to PSA screening in the simulation model; for example, a patient whose PCa metastasizes at age 55 years would be ineligible for PSA screening at all later ages.) Any of the feasible values 0, 1, \ldots, 20 for \( x_t \) signify that PSA screening does occur at epoch \( t \). For any of these feasible values, the corresponding PSA threshold is equal to \( (x_t/2) \) ng/mL, the midpoint of the PSA interval. Consequently, if \( x_t = 0 \), all patients who reach (live through) epoch \( t \) without having previously undergone prostate biopsy or metastasis will undergo prostate biopsy at epoch \( t \) because all PSA test results will trivially satisfy the 0 ng/mL threshold. We let the value \( x_t = 21 \) denote the no-screen event (i.e., no PSA test is performed) at age \( t \). The screening-rule definitions for all feasible values of \( x_t \) are summarized in Table 4.2.
Table 4.2: Decision-variable ($x_t$) value to PSA-screening threshold rule mappings.

<table>
<thead>
<tr>
<th>$x_t$</th>
<th>PSA screening rule at epoch $t$ for eligible patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>PSA threshold 0.0 ng/mL (always biopsy)</td>
</tr>
<tr>
<td>1</td>
<td>PSA threshold 0.5 ng/mL</td>
</tr>
<tr>
<td>2</td>
<td>PSA threshold 1.0 ng/mL</td>
</tr>
<tr>
<td>3</td>
<td>PSA threshold 1.5 ng/mL</td>
</tr>
<tr>
<td>4</td>
<td>PSA threshold 2.0 ng/mL</td>
</tr>
<tr>
<td>5</td>
<td>PSA threshold 2.5 ng/mL</td>
</tr>
<tr>
<td>6</td>
<td>PSA threshold 3.0 ng/mL</td>
</tr>
<tr>
<td>7</td>
<td>PSA threshold 3.5 ng/mL</td>
</tr>
<tr>
<td>8</td>
<td>PSA threshold 4.0 ng/mL</td>
</tr>
<tr>
<td>9</td>
<td>PSA threshold 4.5 ng/mL</td>
</tr>
<tr>
<td>10</td>
<td>PSA threshold 5.0 ng/mL</td>
</tr>
<tr>
<td>11</td>
<td>PSA threshold 5.5 ng/mL</td>
</tr>
<tr>
<td>12</td>
<td>PSA threshold 6.0 ng/mL</td>
</tr>
<tr>
<td>13</td>
<td>PSA threshold 6.5 ng/mL</td>
</tr>
<tr>
<td>14</td>
<td>PSA threshold 7.0 ng/mL</td>
</tr>
<tr>
<td>15</td>
<td>PSA threshold 7.5 ng/mL</td>
</tr>
<tr>
<td>16</td>
<td>PSA threshold 8.0 ng/mL</td>
</tr>
<tr>
<td>17</td>
<td>PSA threshold 8.5 ng/mL</td>
</tr>
<tr>
<td>18</td>
<td>PSA threshold 9.0 ng/mL</td>
</tr>
<tr>
<td>19</td>
<td>PSA threshold 9.5 ng/mL</td>
</tr>
<tr>
<td>20</td>
<td>PSA threshold 10.0 ng/mL</td>
</tr>
<tr>
<td>21</td>
<td>No screening</td>
</tr>
</tbody>
</table>
To represent the stochastic progression of a patient through health states, PSA test results, and prostate biopsy results, we let

$$W = (w_1, w_2, \ldots, w_N) \equiv (s_1, \rho_1, \eta_1), (s_2, \rho_2, \eta_2), \ldots, (s_N, \rho_N, \eta_N)$$

denote the complete sample path of outcomes for all decision epochs, so that

$$W \in W = \{ \delta \times \mathcal{P} \times \mathcal{E} \}^N.$$  

The $j$th replication of the simulation model for strategy $x$ yields the quantity $Q_j(x)$ defined in Equation (4.79) for the $j$th simulated patient. Let

$$G(x; W_j) \equiv Q_j(x)$$

denote the QALYs gained by screening patient $j$ whose stochastic sample path is characterized by $W_j$ under screening strategy $x$. Our goal is to find the strategy yielding the smallest expected NNSQ, that is, strategy

$$x^* = \arg\min_{x \in \Theta} E[NNSQ(x)] = \arg\min_{x \in \Theta} (E_W[G(x; W)])^{-1}.$$  

In order to avoid the difficulties that can arise with an objective function like $\frac{1}{Q}$ when $Q$ is very close to 0, we formulate the equivalent objective of maximizing expected $QG$. That is, our objective is to find the optimal strategy $x^*$ such that

$$x^* = \arg\max_{x \in \Theta} E[Q(x)] = \arg\max_{x \in \Theta} E_W[G(x; W)].$$  

(4.90)

4.2.6 Regularizing Objective Function

Recall that a screening strategy is represented by a vector $x = (x_1, \ldots, x_N)$ with each $x_t \in \{0, 1, \ldots, 21\}$ for $t = 1, 2, \ldots, N$. The feasible value 21 for $x_t$ is characteristically different from feasible values 0, 1, \ldots, 20 in that value 21 represents no screening rather than a finite PSA threshold level. Now, we
introduce a function $y(\cdot)$ which yields a vector $y(x)$ from an input vector $x$. We define $y(x)$ to be the vector consisting of the elements of $x$ that are not equal to 21, where the ordering of the remaining elements is maintained. For example, if $x = (10, 21, 21, 8, 21)$, then $y(x) = (10, 8)$. With this definition in hand, we present the first of two smoothness terms:

$$R_{\text{PSA}}(x) \equiv \frac{1 + r_1(y(x))}{2} \quad \text{for } x \in \Theta, \quad (4.91)$$

where $r_1(v)$ is the lag-one autocorrelation coefficient of the time series $v = \{v_1, v_2, \ldots, v_{N'}\}$ formed by the $N'$ elements of vector $v$ with order maintained, and is calculated as

$$r_1(v) = \begin{cases} 
1, & \text{if } N' \leq 2 \text{ or } \sum_{i=1}^{N'} (v_i - \bar{v})(v_{i+1} - \bar{v}) = 0, \\
\frac{N' - 1}{N'} \sum_{i=1}^{N'} (v_i - \bar{v})^2, & \text{otherwise},
\end{cases} \quad (4.92)$$

where

$$\bar{v} \equiv \begin{cases} 
0, & \text{if } N' = 0, \\
\frac{1}{N'} \sum_{i=1}^{N'} v_i, & \text{if } N' > 0.
\end{cases}$$

In defining the lag-one autocorrelation coefficient, we use the notation $r_1(v)$ instead of $r_1(y(x))$ because $y(x)$ is not the only time series for which we will need to compute the lag-one autocorrelation coefficient.

Our motivation for the formulation (4.91) of $R_{\text{PSA}}(x)$ is the following. We seek a measure of the “smoothness” or “regularity” of a time series of PSA thresholds that takes values in $[0, 1]$, with the value 0 indicating a “nonsmooth” or “irregular” series of PSA thresholds, and the value 1 indicating a “smooth” or “perfectly regular” series of PSA thresholds.

In the context of continuous-time stochastic processes, antipersistence of sample paths means that a positive increment (increase) in the process is more likely to be followed by a negative increment (decrease) in the next nonoverlapping time interval and vice versa; and this tendency of the process to
turn back on itself results in sample paths with a very rough structure. The prototypical antipersistent process is fractional Brownian motion with Hurst exponent $H \in [0, 0.5]$ so that its adjacent increments are negatively correlated [1, 74].

On the other hand, persistence of sample paths in a continuous-times process means that successive nonoverlapping increments of the process are more likely to have the same sign; and smoother sample paths result from this tendency of the process to persist in its current direction of movement. The prototypical persistent process is fractional Brownian motion with Hurst exponent $H \in (0.5, 1)$ so that its adjacent increments are positively correlated.

In the context of discrete-time stochastic processes, the analogue of antipersistence is a negative lag-one correlation; and the analogue of persistence is a positive lag-one correlation. Since lag-one correlations are constrained to the interval $[-1, +1]$, with $-1$ indicating perfect negative lag-one correlation and $+1$ indicating perfect positive lag-one correlation, Equation (4.91) is the appropriate transformation of the sample lag-one correlation of a series of PSA thresholds to the interval $[0, 1]$ so that the lower limit 0 indicates a highly irregular series while the upper limit 1 indicates a highly regular series.

If the time series $\{v_1, v_2, \ldots, v_N\}$ has zero variance, then this situation corresponds to a screening strategy with totally uniform thresholds and thus perfect regularity; and in this case we define $r_1(v)$ to be equal to 1. We also define $r_1(v)$ to be equal to 1 in the trivial cases where $v$ has less than or equal to 2 elements; such cases occur when either the strategy is the no-screening strategy or the strategy consists of a single PSA screening event. When $v$ has 2 elements, we want to define $r_1(v)$ because otherwise the reasonable configuration depicted by Figure 4.2 would yield the undesirable result

$$r_1(v) = -1 \Rightarrow \frac{1 + r_1(v)}{2} = 0.$$
It is not unreasonable that a PSA screening strategy could contain only 2 PSA screening events, each at a different PSA threshold. For this reason, strategies with only 2 elements, such as depicted by Figure 4.2, are taken to be perfectly smooth. Contrarily, a “perfectly unsmooth” screening strategy would contain highly-alternating PSA thresholds such as depicted in Figure 4.3. From Equation (4.92), the PSA thresholds vector $v$ depicted by Figure 4.3 would entail

$$r_1(v) = -1 \Rightarrow \frac{1 + r_1(v)}{2} = 0,$$

which denotes perfect unsmoothness. This contrast between the smoothness represented by Figure 4.2 and the unsmoothness represented by Figure 4.3 is the motivation for using

$$\frac{1}{2} \left[ 1 + r_1(v) \right]$$

as the measure of smoothness for a series of PSA thresholds.
Next, let the function $I: \mathbb{Z}^N \mapsto \mathbb{Z}^N$ be defined so that the vector

$$I(x) \equiv \left[ I(x_i \neq 21) \right]_{i=1,\ldots,N}$$

where $I(\cdot)$ is the standard indicator function. Thus, the vector $I(x) = (I(x_1), I(x_2), \ldots, I(x_N)) \in \mathbb{Z}^N$, where $I(x_i) = 0$ if $x_i = 21$, and $I(x_i) = 1$ otherwise. With this definition in hand, we now present the second of two smoothness terms:

$$R_{\text{FRQ}}(x) \equiv \left| r_1(I(x)) \right|. \quad (4.93)$$

The purpose of $R_{\text{FRQ}}(x)$ is to measure the degree of regularity or consistency in the frequency at which PSA screening occurs. The value $R_{\text{FRQ}}(x) = 0$ corresponds to the theoretical situation in which the ages at which to screen are uniformly randomly chosen. We take the absolute value in Equation (4.93) because the value $r_1(I(x)) = -1$ corresponds to a sequence of alternating 0’s and 1’s, and such a sequence describes a screening strategy of consistently screening every other year, which is perfectly regular.

Before combining the regularity measures $R_{\text{PSA}}(x)$ and $R_{\text{FRQ}}(x)$ with the expected QG estimator
\( \bar{Q}(x) \), we first introduce the logistic function mapping the real number line \( \mathbb{R} \) to the unit interval:

\[
\psi(u) = \frac{1}{1 + e^{-u}}, \quad \text{for } -\infty < u < \infty.  
\]  

(4.94)

The function \( \psi : \mathbb{R} \mapsto \mathbb{R} \) is strictly increasing, and its image is \((0, 1)\). We now define the updated optimization problem using the regularization-modified objective function:

\[
\max_{x \in \Theta} \left\{ (1 - \omega) \psi(\mathbb{E}[\bar{Q}]) + (\omega) \left( \mathcal{R}_{PSA}(x) \cdot \mathcal{R}_{FRQ}(x) \right) \right\},  
\]

(4.95)

where the smoothing coefficient \( \omega \in [0, 1] \) sets the desired balance between the objective of producing a strategy that yields large expected QG and the (possibly, but not necessarily, competing) objective of producing a strategy with sufficient regularity to be suitable for clinical implementation. In terms of the random variable

\[
Z \equiv \psi\left[ \bar{Q} \right],  
\]

(4.96)

the objective function in Equation (4.95) is estimated by

\[
g(x) \equiv (1 - \omega) Z + (\omega) \left( \mathcal{R}_{PSA}(x) \cdot \mathcal{R}_{FRQ}(x) \right).  
\]

(4.97)

The quantity \( Z \) is easily seen to fall in the unit interval. Because \( r_1(y(x)) \in [-1, 1] \), the quantities \( \mathcal{R}_{PSA}(x) \) and \( \mathcal{R}_{FRQ}(x) \) both fall in the unit interval; and therefore the product of the two, \( \mathcal{R}_{PSA}(x) \cdot \mathcal{R}_{FRQ}(x) \), also falls in the unit interval. Therefore, \( g(x) \) defined in Equation (4.97) falls in the unit interval.

A valid \( 100(1 - \alpha)\% \) confidence-interval estimator for \( g(x) \) is

\[
\left[ (1 - \omega)Z + (\omega) \left( \mathcal{R}_{PSA}(x) \cdot \mathcal{R}_{FRQ}(x) \right) \right] 
\pm \left[ \tilde{z}_{1-\alpha/2} \cdot (1 - \omega) \left( \frac{S_{\bar{Q}(x)} \cdot e^{-\bar{Q}(x)}}{\left( 1 + e^{-\bar{Q}(x)} \right)^2 \cdot \sqrt{m}} \right) \right],  
\]

(4.98)

where \( m \) is the number of replications in the simulated sample, \( \tilde{z}_{1-\alpha/2} \) is the \( 1 - \alpha/2 \) quantile of the
standard normal distribution, and $S_{Q(x)}$ is the sample standard deviation as defined in Equation (4.83).

To derive the CI estimator in Equation (4.98), we observe that

$$\psi'(u) = \frac{e^{-u}}{(1 + e^{-u})^2} > 0 \quad \text{for} \quad -\infty < u < \infty$$

If $Q = \frac{1}{m} \sum_{j=1}^{m} Q_j$, then $Z = \psi'(Q) = \frac{1}{1 + e^{-Q}}$. From Hogg et al. [52], we know that the distribution of $Z$ as $m \to \infty$ is asymptotically normal so that we have the approximate result

$$Z \sim \mathcal{N} \left( \psi'(\mu_Q), \frac{\operatorname{Var}[Q]}{m} \right). \quad (4.99)$$

provided $m$ is sufficiently large. We compute our estimate

$$\widehat{\operatorname{Var}}[Z] \approx \left[ \psi'(Q) \right]^2 \cdot \frac{S_Q^2}{m}$$

where $S_Q^2 = \frac{1}{m-1} \sum_{i=1}^{m} (Q_i - \bar{Q})^2$ by the following asymptotically valid approximation:

$$\widehat{\operatorname{Var}}[Z] \approx \frac{e^{-2\bar{Q}}}{\left(1 + e^{-\bar{Q}}\right)^2} \cdot \frac{S_Q^2}{m}. \quad (4.100)$$

Using the standard error estimate

$$\widehat{\operatorname{SE}}[Z] \approx \frac{e^{-\bar{Q}}}{\left(1 + e^{-\bar{Q}}\right)^2} \cdot \frac{S_Q}{\sqrt{m}}$$

we have the following asymptotically valid $100(1 - \alpha)$% CI for $E[Z]$ as $m \to \infty$:

$$Z \pm \tilde{z}_{1-\alpha/2} \widehat{\operatorname{SE}}[Z]. \quad (4.101)$$
which expands to
\[
\left( \frac{1}{1 + e^{-\bar{Q}}} \right) \pm \left( \frac{\sqrt{1 - \alpha/2}}{\sqrt{1 + e^{-\bar{Q}}}^2} \right) \left( \frac{S_{\bar{Q}}}{\sqrt{m}} \right). \tag{4.102}
\]

Since both the second summand in Equation (4.97) and the quantity \( \omega \) are constant with respect to \( x \), the CI estimator in Equation (4.102) is easily adapted to the objective function estimate in Equation (4.97), yielding the adapted CI estimator in Equation (4.98).

The values of the independent regularity measures \( R_{PSA}(x) \) and \( R_{FRQ}(x) \) approach 0 in cases of extreme irregularity and 1 in cases of extreme regularity as \( N' \) increases. Given that independent value is placed on both kinds of regularity, an ideal screening strategy is one that yields large expected QG and also exhibits high regularity in both the magnitudes of screening thresholds and the frequency of screening. For the composite regularity measure \( R_{PSA}(x) \cdot R_{FRQ}(x) \), the largest value appears when both independent regularity measures are close to 1, and, contrariwise, the smallest value appears when both independent measures are close to 0. Thus, the composite regularity measure, as the product of two independent regularity measures, is a meaningful measure of the overall regularity of a given screening strategy.

### 4.2.7 Three-Stage Simulation-Optimization Method

Since \( E_W [G(x; W)] \) is a discrete, nonconvex, nonlinear function of strategy \( x \), we use a GA, a clean-up procedure, and an implementation of COMPASS in a three-stage simulation-optimization method. We estimate \( E_W [G(x; W)] \) by executing \( m \) simulation replications at strategy \( x \) and computing the sample mean of \( m \) observations of \( G(x; W) \); this is precisely the estimate \( \bar{Q}(x) \).

In the first stage of our simulation-optimization method, the GA (described below) is used without any consideration for regularity (letting \( \omega = 0 \)) to find an initial strategy \( \bar{x}^{(0)} \) with expected QG larger than the expected QG for the strategies that constituted the initial population of the GA. Then, in the second stage, \( \bar{x}^{(0)} \) is transformed into \( \bar{x}'^{(0)} \) by CleanUp (described below) to eliminate PSA thresholds that are purely artifacts of the stochastic crossover and mutation operators and that do not produce statistically-significant
positive contributions to the estimate $\bar{Q}(\bar{x})$. Finally, $\hat{x}''$ is used as an initial trial point for the third stage. In the third stage, an implementation of COMPASS (described below) using a nonzero value for $\omega$ starts with $\hat{x}''$ and explores the solution neighborhood around $\hat{x}''$ in order to produce an $\hat{x}^*$ that yields both large expected QG and high regularity. As illustrated in Figure 4.4, the GA in the first stage explores the feasible region for new improved strategies, the clean-up procedure in the second stage removes noise from the GA strategy, and the COMPASS implementation in the third stage exploits the region around the solution from the second stage in order to produce an even further improved strategy.

![Flowchart of three-stage simulation-optimization method](image)

Figure 4.4: Flowchart of three-stage simulation-optimization method. First, the GA broadly explores the feasible region to find a strategy with large expected QG. Then, the clean-up procedure systematically evaluates the strategy from the GA and discards practically insignificant PSA thresholds. Finally, the COMPASS method tries to make regularizing (smoothing) improvements to the cleaned-up strategy without a substantial loss of expected QG.
Genetic Algorithm

The implemented GA is a single-population globally parallel GA. We repeatedly modify a single population of screening strategies, and the evaluation by simulation of each strategy is carried out in parallel. Given the high computational costs of simulating (evaluating) each screening strategy relative to the computational costs of the various GA mechanisms (operators), this is a classic example of an embarrassingly parallel problem. During each iteration (or generation) of the GA, selection, crossover, and mutation operators are repeatedly applied to modify the current population of strategies. The repeated application of the selection, crossover, and mutation operators to construct a new population is referred to as recombination in this dissertation. The mechanics of these operators are such that new strategies probabilistically tend to be configured similarly to current strategies with relatively large expected QG. This probabilistic tendency produces improved screening strategies over the course of the GA. Figure 4.5 illustrates the flow of the GA at a high level.

Figure 4.5: High-level flowchart of the GA. Beginning with a population (set) of initial screening strategies, the GA simulates (evaluates) the population. Then, until some predetermined stopping criterion is met, the GA repeatedly constructs a new population by application of the selection, crossover, and mutation operators, and then simulates the new population.

Pseudocode for the GA is shown in Algorithm 4. The GA calls the RECOMBINE function, which in
turn calls the Selection, Crossover, and Mutation functions. Because there are multiple versions of the recombination, selection, and crossover operators, we use the abstract functions Recombine, Selection, and Crossover to refer to any version of the corresponding operators. The abstract Recombine function may refer to any of the following functions: RecombineStandard, RecombineDeterministicCrowding, or RecombineMeritocracy. The abstract Selection function may refer to any of the following functions: SelectRouletteWheel, SelectTournament, or SelectUniform. The abstract Crossover function may refer to any of the following functions: CrossoverOnePoint, CrossoverTwoPoint, or CrossoverUniform. Figure 4.6 illustrates the organization of the algorithms.

Algorithm 4: GeneticAlgorithm($X$)

Input: Population, $X$, of strategies to be modified.  
Output: Best strategy found by the algorithm.  
Data: The number $\text{NUMGENERATIONS}$ of desired generations of the GA. The number of patients, $\text{SAMPLESIZE}$, to simulate when evaluating any strategy. Function Recombine that takes a population of strategies as its first argument, takes the ordinal of the current GA generation at its second argument, and returns a recombined (modified) population.

1  begin
2      GENERATION $\leftarrow$ 0;
3      while GENERATION $\leq$ NUMGENERATIONS do
4          foreach strategy $X[i] \in X$ do
5              Evaluate $X[i]$ and record estimate, $g(X[i])$, from simulation using samples of size SAMPLESIZE;
6          end
7          BESTSTRATEGY $\leftarrow$ any strategy $X[i]$ such that $g(X[i]) \leq g(X[j]) \ \forall X[j] \in X$;
8          if GENERATION $< \text{NUMGENERATIONS}$ then
9              $X \leftarrow$ Recombine($X$, GENERATION);
10           end
11          GENERATION $\leftarrow$ GENERATION + 1;
12      end
13  return BESTSTRATEGY;
14  end
We experimented with 3 different recombination strategies: standard, deterministic crowding, and meritocracy. The term recombination is often used to refer to only the crossover and mutation operators, but because we develop and implement 3 different overall strategies for combined application of selection, crossover, and mutation, we use the term recombination to also include the selection operator. Standard recombination simply repeatedly calls the selection, crossover, and mutation operators to build a new population. We refer to this as “standard” recombination insofar as the general approach of using selection, crossover, and mutation operators to replace the entire population—excepting possibly a small number of so called elite members—has been referred to as canonical [82] and has been widely used with various minor modifications. Pseudocode for our implementation of standard recombination is shown in Algorithm 5.

Deterministic crowding principally differs from standard recombination in that it attempts to maintain diversity in the new population by replacing parent strategies in the old population by the most similar children produced by crossover, where similarity is measured by the Euclidean distance between the
decision variable vectors representing the parent strategies. Parents are replaced by the most similar children because doing so has the effect of isolating genetic diversity. And this isolation, in contrast to allowing the genetic features to be spread out into the population, tends to retain pockets of relatively-distinct genetic features, i.e., a diverse population. The replacement of a parent with a most-similar child is only made when that child is an improvement on that parent. Deterministic crowding also promotes diversity by selecting parents from the population without replacement, which reduces the probability of causing two identical parents from reproducing and thus forming two identical children; standard recombination, on the other hand, selects parents with replacement. Our implementation of deterministic crowding is closely based on the development in Mahfoud [73]. Pseudocode for our implementation of deterministic crowding is shown in Algorithm 6.

Meritocracy principally differs from standard recombination in that it discourages child strategies formed by crossover from replacing their corresponding parents if the parents are superior. If the selection, crossover, and mutation process during meritocratic recombination attempts to replace a superior parent strategy by an inferior child strategy, the replacement is disallowed with probability \((1 - \text{INFERIORITY RATE})\). This probability is varied over the course of the GA such that inferior replacements are permitted toward the beginning of the GA but are virtually eliminated toward the end of the GA. In meritocracy, we explicitly control the proportion of the population at each generation that we will permit to possibly become inferior to the corresponding parents. This, of course, allows us to limit the degree to which the GA performance can suffer as a result of generating children that are worse than their parents. Whereas our use of deterministic crowding is intended primarily to introduce a mechanism for maintaining population diversity, our development and use of meritocracy is intended primarily to introduce a mechanism whereby an increasingly smaller percentage of the population is permitted to be replaced by inferior solutions. Meritocracy is new research. Pseudocode for our implementation of meritocracy is shown in Algorithm 7.
Algorithm 5: \textsc{Recombine\_Standard}(X, \textsc{Generation})

\textbf{Input}: Population, $X$, of strategies to be modified; the second parameter, \textsc{Generation}, is ignored in this algorithm but included for uniform parameterization; number, \textsc{NumElites}, of best strategies not subject to recombination.

\textbf{Output}: Recombined population of strategies.

\textbf{Data}: The \textsc{Size}(-) operator returns the number of strategies in a given population.

1 \hspace{1em} \textbf{begin}
2 \hspace{2em} \textsc{oldPop} $\leftarrow X$;
3 \hspace{2em} \textsc{newPop} $\leftarrow \emptyset$;
4 \hspace{2em} Sort \textsc{oldPop} in descending order by $g(\cdot)$ estimate (i.e., sort from best to worst);
5 \hspace{2em} \textbf{for} $i = 0$ to \textsc{NumElites} $- 1$ \textbf{do}
6 \hspace{3em} \textsc{newPop} $\leftarrow$ \textsc{newPop} $\cup \{ \textsc{oldPop}[i] \}$;
7 \hspace{2em} \textbf{end}
8 \hspace{2em} \textbf{while} \textsc{Size}\textsc{newPop} $< \textsc{Size}\textsc{oldPop}$ \textbf{do}
9 \hspace{3em} \textsc{parent1} $\leftarrow$ \textsc{Selection}\textsc{oldPop}$;
10 \hspace{3em} \textsc{parent2} $\leftarrow$ \textsc{Selection}\textsc{oldPop}$;
11 \hspace{3em} \{ \textsc{child1}, \textsc{child2} \} $\leftarrow$ \textsc{Crossover}(\textsc{parent1}, \textsc{parent2});
12 \hspace{3em} Discard \textsc{child2};
13 \hspace{3em} \textsc{newStrategy} $\leftarrow$ \textsc{child1};
14 \hspace{3em} \textsc{Mutate}(\textsc{newStrategy});
15 \hspace{3em} \textsc{newPop} $\leftarrow$ \textsc{newPop} $\cup \{ \textsc{newStrategy} \}$;
16 \hspace{2em} \textbf{end}
17 \hspace{2em} \textbf{return} \textsc{newPop};
18 \hspace{1em} \textbf{end}
Algorithm 6: RECOMBINE\_DETERMINISTIC\_CROWDING($X$, GENERATION)

**Input:** Population, $X$, of strategies to be modified; the second parameter, GENERATION, is ignored in this algorithm but included for uniform parameterization.

**Output:** Recombined population of strategies.

**Data:** The Size() operator returns the number of strategies in a given population; the function $d(x, x') \equiv \|x - x'\|_2$ computes the usual $\ell_2$ norm of the vector $x - x' \in \mathbb{R}^N$; and for every MEMBER $\in X$ there is an associated $x = x$(MEMBER) $\in \Theta$, so that for example $d$(PARENT1, CHILD1) = $\|x$(PARENT1) $- x$(CHILD1)$\|_2$.

```
begin
    OLDPOP $\leftarrow X$;
    NEWPOP $\leftarrow \emptyset$;
    while Size(OLDPOP) $\geq$ 2 do
        PARENT1 $\leftarrow$ SELECTION(OLDPOP);
        Remove selected PARENT1 from OLDPOP;
        PARENT2 $\leftarrow$ SELECTION(OLDPOP);
        Remove selected PARENT2 from OLDPOP;
        { CHILD1, CHILD2 } $\leftarrow$ Crossover(PARENT1, PARENT2);
        MUTATE(CHILD1);
        MUTATE(CHILD2);
        if $d$(PARENT1, CHILD1) + $d$(PARENT2, CHILD2) $\leq$
            $d$(PARENT1, CHILD2) + $d$(PARENT2, CHILD1) then
            if $g$(CHILD1) $>$ $g$(PARENT1) then
                NEWPOP $\leftarrow$ NEWPOP $\cup$ { CHILD1 };
            else
                NEWPOP $\leftarrow$ NEWPOP $\cup$ { PARENT1 };
            end
            if $g$(CHILD2) $>$ $g$(PARENT2) then
                NEWPOP $\leftarrow$ NEWPOP $\cup$ { CHILD2 };
            else
                NEWPOP $\leftarrow$ NEWPOP $\cup$ { PARENT2 };
            end
        else
            if $g$(CHILD1) $>$ $g$(PARENT2) then
                NEWPOP $\leftarrow$ NEWPOP $\cup$ { CHILD1 };
            else
                NEWPOP $\leftarrow$ NEWPOP $\cup$ { PARENT2 };
            end
            if $g$(CHILD2) $>$ $g$(PARENT1) then
                NEWPOP $\leftarrow$ NEWPOP $\cup$ { CHILD2 };
            else
                NEWPOP $\leftarrow$ NEWPOP $\cup$ { PARENT1 };
            end
        end
        NEWPOP $\leftarrow$ NEWPOP $\cup$ OLDPOP;
    return NEWPOP;
end
```
**Algorithm 7: RECOMBINE\_MERITOCRACY($X$, GENERATION)**

**Input:** Population, $X$, of strategies to be modified; number, GENERATION, of the current generation of the GA used to compute INFERIORITY RATE.

**Output:** Recombined population of strategies.

**Data:** The SIZE() operator returns the number of strategies in a given population; the probability, INFERIORITY CONSTANT, used to compute INFERIORITY RATE; the number, NUMELITES, of best strategies not subject to recombination.

```
1 begin
2   OLDPOP ← X;
3   NEWPOP ← Ø;
4   Sort OLDPOP in descending order by $g()$ estimate (i.e., sort from best to worst);
5   CANDIDATES ← Ø;
6   for $i = 0$ to NUMELITES − 1 do
7       NEWPOP ← NEWPOP ∪ {OLDPOP[$i$]};
8   end
9   Remove first NUMELITES elements from OLDPOP;
10  BENCHPOP ← OLDPOP;
11  while SIZE(OLDPOP) > 0 do
12     if SIZE(OLDPOP) > 1 then
13         PARENT1 ← SELECT(OLDPOP);
14         Remove selected PARENT1 from OLDPOP;
15         PARENT2 ← SELECT(OLDPOP);
16         Remove selected PARENT2 from OLDPOP;
17         {CHILD1, CHILD2} ← CROSSOVER(PARENT1, PARENT2);
18         MUTATE(CHILD1);
19         MUTATE(CHILD2);
20         CANDIDATES ← CANDIDATES ∪ {CHILD1, CHILD2};
21     else
22         CANDIDATES ← CANDIDATES ∪ OLDPOP;
23     end
24   end
25   foreach $C$ ∈ CANDIDATES do
26       Evaluate $C$ and record estimate, $g(C)$, from simulation;
27   end
28   INFERIORITY RATE ← INFERIORITY CONSTANT \* GENERATION;
29   for $i = 0$ to SIZE(CANDIDATES) − 1 do
30       if $g($CANDIDATES[$i$]) > $g($BENCHPOP[$i$]) then
31           NEWPOP ← NEWPOP ∪ {CANDIDATES[$i$]};
32       else
33           $u$ ← uniform random variate ∈ [0, 1];
34           if $u$ ≤ INFERIORITY RATE then
35              NEWPOP ← NEWPOP ∪ {CANDIDATES[$i$]};
36           else
37              NEWPOP ← NEWPOP ∪ {BENCHPOP[$i$]};
38       end
39   end
40   return NEWPOP;
41 end
```
Each recombination strategy utilizes the underlying core mechanism of successive application of the selection, crossover, and mutation operators. An illustration of the combined operation of these operators in construction of a new strategy is shown in Figure 4.7.

![Diagram of selection, crossover, and mutation operators](image)

**Figure 4.7:** Illustration of the selection, crossover, and mutation operators in the GA. To construct a single new screening strategy, the GA selects two existing strategies which are then transformed by application of the crossover operator into a single strategy, which is further modified by application of the mutation operator.

We experimented with 3 different, commonly-used selection operators: roulette wheel, tournament, and uniform selection. Both roulette wheel and tournament selection are described in Sastry et al. [89] and Zhong et al. [113]. Uniform selection is recommended for use with deterministic crowding in Mahfoud [73], but could be used in the other recombination strategies. Roulette wheel selection randomly selects a strategy from the given population with a probability proportionate to that strategy’s fitness relative to the cumulative fitness of the entire population of strategies. Pseudocode for our implementation of roulette wheel selection is shown in Algorithm 8. Tournament selection uniformly randomly selects two strategies from the given population, and chooses the strategy with higher estimated fitness. Pseudocode for our implementation of tournament selection is shown in Algorithm 9. Uniform selection uniformly randomly selects a strategy from the given population without regard for the fitness of any strategy. Pseudocode for our implementation of uniform selection is shown in Algorithm 10.
Algorithm 8: `SELECT_ROLETWHEEL(X)`

**Input**: Population, \( X \), of strategies to select from.

**Output**: Selected strategy.

**Data**: The \( \text{SIZE}() \) operator returns the number of strategies in a given population. The \( \text{SUM}() \) operator returns the sum of the values in an array.

1. begin
2. \[ \text{MINVALUE} \leftarrow g(X[i]) : g(X[i]) \leq g(X[j]) \quad \forall j \in \{0, \ldots, \text{SIZE}(X) - 1\}; \]
3. \[ \text{for } i = 0 \text{ to } \text{SIZE}(X) - 1 \text{ do} \]
4. \[ \quad \text{PROBDIST}[i] \leftarrow g(X[i]); \]
5. end
6. /* Ensure that all PROBDIST values are positive. */
7. if \( \text{MINVALUE} < 0 \) then
8. \[ \quad \text{for } i = 0 \text{ to } \text{SIZE}(X) - 1 \text{ do} \]
9. \[ \quad \quad \text{PROBDIST}[i] \leftarrow \text{PROBDIST}[i] - \text{MINVALUE}; \]
10. end
11. /* Make PROBDIST array cumulative. */
12. \[ \text{for } i = 1 \text{ to } \text{SIZE}(X) - 1 \text{ do} \]
13. \[ \quad \text{PROBDIST}[i] \leftarrow \text{PROBDIST}[i - 1] + \text{PROBDIST}[i]; \]
14. end
15. /* Normalize to form a valid discrete probability distribution. */
16. \[ \text{SUMVALUES} \leftarrow \text{SUM}(\text{PROBDIST}); \]
17. \[ \text{for } i = 0 \text{ to } \text{SIZE}(X) - 1 \text{ do} \]
18. \[ \quad \text{PROBDIST}[i] \leftarrow \text{PROBDIST}[i] / \text{SUMVALUES}; \]
19. end
20. \( u \leftarrow \text{uniform random variate } \in [0, 1]; \)
21. \[ \text{for } i = 0 \text{ to } \text{SIZE}(X) - 1 \text{ do} \]
22. \[ \quad \text{if } u \leq \text{PROBDIST}[i] \text{ then} \]
23. \[ \quad \quad \text{return } X[i]; \]
24. end
25. end
Algorithm 9: SELECT_TOURNAMENT(\(X\))

**Input:** Population, \(X\), of strategies to select from.

**Output:** Selected strategy.

**Data:** The \(\text{SIZE}()\) operator returns the number of strategies in a given population.

1 begin
2 \hspace{1em} if \(\text{SIZE}(X) = 1\) then
3 \hspace{2em} return \(X[0]\); \hspace{1em} end
4 \hspace{1em} CANDIDATE1 \(\leftarrow\) random integer \(\in [0, \text{SIZE}(X) - 1]\);
5 \hspace{1em} repeat
6 \hspace{2em} CANDIDATE2 \(\leftarrow\) random integer \(\in [0, \text{SIZE}(X) - 1]\);
7 \hspace{2em} until \(\text{CANDIDATE1} \neq \text{CANDIDATE2}\);
8 \hspace{1em} if \(g(X[\text{CANDIDATE1}]) > g(X[\text{CANDIDATE2}])\) then
9 \hspace{2em} return \(X[\text{CANDIDATE1}]\); \hspace{1em} else
10 \hspace{2em} return \(X[\text{CANDIDATE2}]\); \hspace{1em} end
11 \hspace{1em} end

Algorithm 10: SELECT_UNIFORM(\(X\))

**Input:** Population, \(X\), of strategies to select from.

**Output:** Selected strategy.

**Data:** The \(\text{SIZE}()\) operator returns the number of strategies in a given population.

1 begin
2 \hspace{1em} INDEX \(\leftarrow\) random integer \(\in [0, \text{SIZE}(X) - 1]\);
3 \hspace{1em} return \(X[\text{INDEX}]\);
4 end
We experimented with 3 different, commonly used crossover operators: one-point, two-point, and uniform crossover. These crossover operators are described and compared in Spears and Anand [99] and Baluja and Caruana [11]. One-point crossover uniformly randomly chooses one point between two adjacent decision variables by which to partition the decision variables in two parent strategies into two segments, and then produces two children by swapping the first segments between each of the parents. Pseudocode for one-point crossover is shown in Algorithm 11. Two-point crossover uniformly randomly chooses two different points between adjacent decision variables by which to partition the decision variables in two parent strategies into three segments, and then produces two children by swapping the middle segments between each of the parents. Pseudocode for two-point crossover is shown in Algorithm 12. Uniform crossover cycles through each decision variable index and with some defined probability swaps the decision variable values between parents. Pseudocode for uniform crossover is shown in Algorithm 13.

Algorithm 11: CROSSOVER_ONEPOINT(PARENT1, PARENT2)

**Input:** Parent strategies to crossover.
**Output:** New child strategies formed by crossover.
**Data:** NUMVARS is the number of decision variables in a screening strategy.

```
1 begin
2    CHILD1 ← PARENT1;
3    CHILD2 ← PARENT2;
4    CXPOINT ← random integer ∈ [1, NUMVARS – 1];
5    for i = 0 to CXPOINT do
6        temp ← CHILD1[i];
7        CHILD1[i] ← CHILD2[i];
8        CHILD2[i] ← temp;
9    end
10   return {CHILD1, CHILD2};
11 end
```
Algorithm 12: Crossover_TwoPoint(PARENT1, PARENT2)

Input: Parent strategies to crossover.
Output: New child strategies formed by crossover.
Data: NUMVARS is the number of decision variables in a screening strategy.

1 begin
2 CHILD1 ← PARENT1;
3 CHILD2 ← PARENT2;
4 CXPOINT1 ← random integer ∈ [1, NUMVARS − 2];
5 CXPOINT2 ← random integer ∈ [CXPOINT1 + 1, NUMVARS − 1];
6 for i = CXPOINT1 + 1 to CXPOINT2 do
7   temp ← CHILD1[i];
8   CHILD1[i] ← CHILD2[i];
9   CHILD2[i] ← temp;
10 end
11 return {CHILD1, CHILD2};
12 end

Algorithm 13: Crossover_Uniform(PARENT1, PARENT2)

Input: Parent strategies to crossover.
Output: New child strategies formed by crossover.
Data: NUMVARS is the number of decision variables in a screening strategy.

1 begin
2 CHILD1 ← PARENT1;
3 CHILD2 ← PARENT2;
4 for i = 0 to NUMVARS − 1 do
5   u ← uniform random variate ∈ [0, 1];
6   if u ≤ 0.5 then
7     temp ← CHILD1[i];
8     CHILD1[i] ← CHILD2[i];
9     CHILD2[i] ← temp;
10 end
11 end
12 return {CHILD1, CHILD2};
13 end

For the mutation operator, the GA generates a random number (i.e., a sample from Uniform(0,1)) for each decision variable of a given strategy, and for each decision variable if the value of the corresponding random number is less than or equal to a predetermined mutation probability, then a new value for that decision variable is uniformly randomly chosen from the set of all feasible values for that decision.
variable. The mutation operator is described, and multiple mutation probabilities are tested and compared, in Baluja and Caruana [11] and Haupt [46]. Pseudocode for our implementation of the mutation operator is shown in Algorithm 14.

Algorithm 14: MUTATE(STRATEGY)

Input: Strategy to be mutated.
Output: Mutated strategy.

Data: NUMVARS is the number of decision variables in a screening strategy; MUTATION_RATE is the mutation rate; and $x_i(STRATEGY)$ is the $i$th decision variable in STRATEGY so that $x(STRATEGY) = [x_1(STRATEGY), \ldots, x_{NUMVARS}(STRATEGY)]$.

1 begin
2 MUTATEDSTRATEGY ← STRATEGY;
3 for $i = 1$ to NUMVARS do
4 \hspace{1em} u ← uniform random variate $\in [0, 1]$;
5 \hspace{1em} if $u \leq$ MUTATION_RATE then
6 \hspace{2em} $x_i$(MUTATEDSTRATEGY) ← uniformly randomly selected feasible value for the $i$th PSA-threshold decision variable;
7 \hspace{1em} end
8 end
9 return MUTATEDSTRATEGY;
10 end

Clean-Up Procedure

The clean-up procedure is motivated by the fact that the GA can produce highly fit strategies that have some screening thresholds that do not bring a significant positive contribution to estimated expected QG. For example, Figure 4.8 depicts a screening strategy that was produced for the White risk group. When we examine each threshold’s contribution to the estimated expected QG, we find that the two nonzero thresholds ($> 0$ ng/mL) in this case are not responsible for a significant amount of the expected total QG that this strategy yields. This can be seen visually in Figure 4.9, which plots all the positive $D_j$ observations grouped by age at biopsy (i.e., grouped by the screening threshold at which the biopsy that led to treatment was triggered) for a sample of 50,000 simulated patients.
Figure 4.8: Screening strategy with practically-insignificant residual DNA from the GA.
Figure 4.9: QALYs-gained observations (i.e., $D_j$ values) for all screen-detected patients who benefited from screening under the strategy depicted in Figure 4.8.
Because this kind of situation can occur, we now present a clean-up procedure to systematically eliminate thresholds in a strategy that do not significantly improve the estimated expected QG. Given a sample \( \{j\} \) of patients simulated under screening strategy \( x \) and under the no-screening strategy, let

\[
SD_t(x) \equiv \{ j : \text{patient } j \text{ was screen-detected at time } t \text{ under strategy } x \},
\]

(4.103)

and let

\[
K_t(x) \equiv \sum_{j \in SD_t(x)} D_j(x) = \sum_{j \in SD_t(x)} \left( Q_{S,j}(x) - Q_{NS,j}(x) \right).
\]

(4.104)

To discard screening thresholds that have no practical or statistically-significant benefit in terms of expected QG, we utilize \( K_t(x) \) in a simple procedure to systematically test the significance of each screening threshold, discarding insignificant ones. We test the thresholds \( \{ x_t : x_t \neq 21, \ t = 1, \ldots, N \} \) in increasing order of their individual contribution \( K_t(x) \) to the estimated \( E[QG] \) for strategy \( x \). As soon as a significant screening threshold is identified, the procedure terminates. When testing a given threshold for significance, we evaluate the modified strategy using \( m \) simulated patients and compare the estimated QG to the estimated QG obtained by simulating \( m \) patients under the original unmodified strategy. If these QG estimates are not statistically significantly different at the \( \alpha = 0.05 \) level, then we treat the tested threshold as practically insignificant and discard it. This clean-up procedure is defined formally in Algorithm 15. In this research, we always use \( m = 10^7 \) as the simulation sample size in the clean-up procedure.
Algorithm 15: CleanUp(x)

**Input:** Screening strategy x.

**Data:** Let m be the number of patients (sample size) of the simulations used to estimate the point and CI estimates of expected QG in this algorithm.

1. **begin**
2. Evaluate strategy x using m simulated patients;
3. Set $\text{LOWERBOUNDORIGINAL} \leftarrow \bar{Q}(x) - (2.24)S_{\bar{Q}(x)}/\sqrt{m}$;
4. Set CONTINUE $\leftarrow$ true;
5. **while** CONTINUE **do**
6. Set $t^* \leftarrow -1$;
7. Set $QG^* \leftarrow +\infty$;
8. **for** $t = 1$ to $N$ **do**
9. if $x[t] \neq 21$ then
10. if $K_t(x) < QG^*$ then
11. Set $t^* \leftarrow t$;
12. Set $QG^* \leftarrow K_t(x)$;
13. end
14. end
15. if $t^* \neq -1$ then
16. Set $t_{\text{last}} \leftarrow x[t^*]$;
17. Set $x[t^*] \leftarrow 21$;
18. Evaluate strategy x using m simulated patients;
19. if $\bar{Q}(x) + (2.24)S_{\bar{Q}(x)}/\sqrt{m} < \text{LOWERBOUNDORIGINAL}$ then
20. Set $x[t^*] \leftarrow t_{\text{last}}$;
21. Set CONTINUE $\leftarrow$ false;
22. end
23. else
24. Set CONTINUE $\leftarrow$ false;
25. end
26. end
27. **return** x;
28. **end**

Using basecase parameter settings for the reference risk group, Algorithm 15 transforms the strategy shown in Figure 4.8 to the strategy shown in Figure 4.10.
Figure 4.10: Screening strategy from Figure 4.8 after having the practically-insignificant thresholds removed by the clean-up procedure.
COMPASS

Our COMPASS implementation takes a single screening strategy as the initial solution. We let the stopping criterion be a predetermined number of iterations of the algorithm. Random sampling from the MPA is conducted by coordinate sampling [57]. In coordinate sampling, each sampled solution differs from the sample-best solution in only one coordinate. For each newly sampled solution, a coordinate is uniformly randomly chosen, and the value of that coordinate is uniformly randomly chosen from the set of possible values that would result in a solution inside or on the boundary of the MPA. Thus, each newly sampled solution is the sample-best solution shifted in a single dimension. In our particular usage of COMPASS, we wish to constrain COMPASS to only modifying strategies (solutions) by changing the values of decision variables that correspond to screen events (values not equal to 21) in the initial strategy. This is achieved by identifying all coordinates corresponding to decision variables in the initial strategy with value equal to 21, and exempting those coordinates from coordinate sampling. For a simplified illustrative example, if the initial strategy for COMPASS is

\[(x_1 = 21, x_2 = 10, x_3 = 21, x_4 = 21, x_5 = 9, x_6 = 8),\]

then the only coordinates eligible for varying by coordinate sampling are 2, 5, and 6. Consequently, in this example the final strategy returned by COMPASS is guaranteed to have the following property: \(x_1 = x_3 = x_4 = 21\). We refer to this process of coordinate sampling of the nonexempt coordinates as “restricted sampling.”

For the simulation allocation rule (SAR), our implementation of COMPASS utilizes an optimal computing budget allocation (OCBA) procedure to compute the number of replications required of all visited solutions in order to guarantee that the chosen sample-best solution is actually the best solution so far with a predetermined probability. Pseudocode for our implementation of COMPASS is shown in Algorithm 16.

145
Algorithm 16: COMPASS(INITIALSTRATEGY)

**Input:** Initial strategy.

**Output:** Best strategy found by the algorithm.

**Data:** \( \Theta \) is the feasible region of screening strategies; \texttt{INITIALREPLICATIONS} is the number of simulation replications for initially evaluating \texttt{INITIALSTRATEGY}; and \texttt{NUMITERATIONS} is the number of desired iterations of COMPASS.

```plaintext
1 begin
2 \hspace{1em} ITERATION \leftarrow 0;
3 \hspace{1em} VISITEDSET \leftarrow \{\texttt{INITIALSTRATEGY}\};
4 \hspace{1em} BESTSTRATEGY \leftarrow \texttt{INITIALSTRATEGY};
5 \hspace{1em} Identify coordinates \{ i : x_i(\texttt{INITIALSTRATEGY}) = 21 \} that are exempt from sampling;
6 \hspace{1em} Simulate \texttt{INITIALSTRATEGY} using \texttt{INITIALREPLICATIONS} replications;
7 \hspace{1em} MOSTPROMISINGAREA \leftarrow \Theta;
8 \hspace{1em} while ITERATION \leq \texttt{NUMITERATIONS} do
9 \hspace{2em} Randomly sample \texttt{SAMPLINGSIZE} new solutions from \texttt{MOSTPROMISINGAREA} using restricted sampling;
10 \hspace{2em} Add sampled solutions to \texttt{VISITEDSET};
11 \hspace{2em} Use OCBA to determine number of simulation replications \( m_\ell \) required for each \( X[\ell] \in \texttt{VISITEDSET} \);
12 \hspace{2em} foreach strategy \( X[\ell] \in \texttt{VISITEDSET} \) do
13 \hspace{3em} Evaluate \( X[\ell] \) and record estimate, \( g(X[\ell]) \), from simulation using \( m_\ell \) replications;
14 \hspace{2em} endforeach
15 \hspace{2em} BESTSTRATEGY \leftarrow any strategy \( X[\ell] \) such that \( g(X[\ell]) \geq g(X[j]) \) for all \( X[j] \in \texttt{VISITEDSET} \);
16 \hspace{2em} MOSTPROMISINGAREA \leftarrow \Theta;
17 \hspace{2em} foreach strategy \( X[\ell] \in \texttt{VISITEDSET} \) do
18 \hspace{3em} Construct hyperplane defined as the set of all points \( X \) satisfying
19 \hspace{4em} (BESTSTRATEGY \(- X[\ell])' \left( X - \frac{\text{BESTSTRATEGY} + X[\ell]}{2} \right) = 0
20 \hspace{3em} and bound the \texttt{MOSTPROMISINGAREA} using this hyperplane;
21 \hspace{2em} endforeach
22 \hspace{1em} ITERATION \leftarrow ITERATION + 1;
23 \hspace{1em} end
24 \hspace{1em} return BESTSTRATEGY;
25 end
```

Effort has been exerted to prove asymptotic convergence properties of GAs and the conditions under which such convergence properties apply [37, 92, 93]. Fogel [37] formulated the sequence of successive generations of a GA as a Markov chain with the states defined by all possible configurations of the
solutions (decision-variable vectors) in the population, and showed that although convergence to the global optimum will asymptotically occur when every decision variable is subject to mutation and the best solution at every generation is maintained (elitism), convergence may require far too much execution time to be of practical use. Nonetheless, GAs are widely used in fields such as operations research and management science [20]. The COMPASS procedure has been proven to asymptotically converge to a local optimum under mild conditions [54]. These convergence properties are important, but according to Hong and Nelson [54], most simulation-optimization software packages do not provide any convergence guarantees with their methods.

4.3 Results

4.3.1 Test Problems for the GA

To verify and compare the GA recombination strategies discussed above, we conducted optimization experiments on the following minimization test problems from Ali et al. [4]: Ackley’s Problem (ACKL); the Becker and Lago Problem (BELA); the Salomon Problem (SALO); and the Griewank Problem (GRWK). We use test problems with deterministic objective functions because the primary purpose of this testing is to independently (i.e., in isolation from the simulation model and the other methodologies in this thesis) verify the effectiveness of the GA recombination strategies.

**Ackley’s Problem**  Test problem ACKL has objective function

\[
f(x) = -20 \exp \left( -0.02 \sqrt{n^{-1} \sum_{i=1}^{n} x_i^2} \right) - \exp \left( n^{-1} \sum_{i=1}^{n} \cos (2\pi x_i) \right) + 20 + e \]

(4.105)

and global optimum

\[x^* = (0, \ldots, 0)\]

with corresponding objective function value

\[f(x^*) = 0,\]
and is bounded such that

$$-30 \leq x_i \leq 30 \quad \text{for} \quad i = 1, \ldots, n.$$  

**Becker and Lago Problem**  Test problem BELA has objective function

$$f(x) = \sum_{i=1}^{n} (|x_i| - 5)^2$$  

(4.106)

and global optimum

$$x^* = (\pm 5, \ldots, \pm 5)$$

with corresponding objective function value

$$f(x^*) = 0,$$

and is bounded such that

$$-10 \leq x_i \leq 10 \quad \text{for} \quad i = 1, \ldots, n.$$  

**Salomon Problem**  Test problem SALO has objective function

$$f(x) = 1 - \cos(2\pi \|x\|) + 0.1\|x\|$$  

(4.107)

and global optimum

$$x^* = (0, \ldots, 0)$$

with corresponding objective function value

$$f(x^*) = 0,$$

and is bounded such that

$$-100 \leq x_i \leq 100 \quad \text{for} \quad i = 1, \ldots, n.$$
**Griewank Problem**  Test problem GRWK has objective function

\[
f(x) = 1 + \frac{1}{4000} \sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos \left( \frac{x_i}{\sqrt{i}} \right) \tag{4.108}\]

and global optimum

\[x^* = (0, \ldots, 0)\]

with corresponding objective function value

\[f(x^*) = 0,\]

and is bounded such that

\[-600 \leq x_i \leq 600 \text{ for } i = 1, \ldots, n.\]

For each of the test problems and using each of the recombination strategies, we ran 50 independent runs of the GA, each with 200 generations, using two-point crossover, tournament selection (except in the case of Deterministic Crowding, which requires uniform selection), and a mutation rate of 0.01. Each of the 50 runs used a different randomly constructed initial population of solutions. Each test problem was set to have dimension 40, that is, \(\text{card}(x) = 40\). For a given pairing of test problem and recombination strategy, we computed descriptive statistics on the objective function value of the best solution at each GA generation over the sample of 50 runs.

The results from the comparison of Standard Recombination, Deterministic Crowding, and Meritocracy are presented by test problem. The results for test problems ACKL, BELA, SALO, and GRWK are shown, respectively, in Figures 4.11, 4.12, 4.13, and 4.14.
Figure 4.11: Comparison of the 3 recombination strategies for the GA on the ACKL test problem.
Figure 4.12: Comparison of the 3 recombination strategies for the GA on the BELA test problem.
Figure 4.13: Comparison of the 3 recombination strategies for the GA on the SALO test problem.
Figure 4.14: Comparison of the 3 recombination strategies for the GA on the GRWK test problem.
From these results, it is apparent that the GA tends to converge to the optimal solution neighborhood (values of \( f(x) \) close to 0) by 200 or fewer generations. Based on the tightness of the CIs and IQRs, the performance of the GA appears to be robust to randomness in the initial population. And, perhaps most importantly, we find no evidence in these test problems that any of the 3 recombination strategies is clearly superior.

### 4.3.2 Optimization Results for Basecase Parameter Values

Table 4.3 contains the basecase estimates for the QALY disutilities in the simulation model along with their sources. The basecase estimates for the other parameters of the reference risk group (whites) are listed along with their sources in Table 3.6. In all the results in this chapter, each simulated patient is permitted to undergo at most 3 biopsies, except where explicitly stated otherwise.

**Table 4.3: Basecase numerical estimates for QALY disutilities in the natural history model.**

<table>
<thead>
<tr>
<th>Disutility</th>
<th>Symbol</th>
<th>Estimate</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening</td>
<td>( D_{Scr} )</td>
<td>0.00019</td>
<td>[29, 23, 48]</td>
</tr>
<tr>
<td>Biopsy</td>
<td>( D_{Biop} )</td>
<td>0.00577</td>
<td>[23, 48]</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>( D_{Dia} )</td>
<td>0.01667</td>
<td>[65, 48]</td>
</tr>
<tr>
<td>Treatment</td>
<td>( D_{Tre} )</td>
<td>0.24667</td>
<td>[101, 14, 48]</td>
</tr>
<tr>
<td>Recovery</td>
<td>( D_{Rec} )</td>
<td>0.05</td>
<td>[87, 101, 48]</td>
</tr>
<tr>
<td>Metastasis</td>
<td>( D_{Met} )</td>
<td>0.4</td>
<td>[12]</td>
</tr>
</tbody>
</table>

For each of the compared risk groups, we ran the three-stage method to obtain an estimate of the optimal screening strategy. In the following three paragraphs, we describe the various settings of the algorithms in the three-stage method that were used to generate the results presented in the remainder of this chapter.

**GA settings.** In the first stage, the GA was run for 200 generations with a constant population size of 30, a mutation rate (\( MutationRate \)) of 0.01, and an elitism order (\( NumElites \)) of 1. The initial population for the GA was constructed as follows: 500 single-interval (only one age-span or interval
during which screening occurs), uniform-frequency (screening frequency unchanged during screening interval), and uniform-threshold (PSA threshold unchanged during screening interval) screening strategies were randomly generated; on the basis of estimated expected QG, the best 30 of the randomly generated strategies were used as the initial population for the GA. The GA was run using standard recombination, tournament selection, and two-point crossover. The coefficient of regularity, $\omega$, was constant at 0.

COMPASS settings. In the third stage, the best strategy found by the GA was used as the starting solution for COMPASS, which was run for 10 iterations, sampling 5 new strategies at each iteration. During the operation of COMPASS, the coordinate sampling method was restricted in sampling by moving along only those coordinates corresponding to the decision variables not equal to 21 (no screening) in the initial strategy. The coefficient of regularity, $\omega$, was constant at 0.5 in order to equally weight the QALYs and regularity terms in the objective function.

Using the basecase parameter values and the algorithm settings detailed above, we ran the three-stage method for each of the risk groups. The near-optimal strategy produced by this method for each risk group is shown in Figure 4.15.
Figure 4.15: Optimized strategies $\hat{x}_W^*$, $\hat{x}_{AA}^*$, $\hat{x}_{FH}^*$, $\hat{x}_{C1}^*$, and $\hat{x}_{C2}^*$ produced by the three-stage method for the White, African American, Family History, CCI=1, and CCI$\geq$2 risk groups, respectively, using the basecase parameter estimates.
To determine the extent to which the optimized strategies shown in Figures 4.15a through 4.15e are better than known screening strategies, we compared the optimized strategies with the strategies examined in Chapter 3. In that chapter, we examined screening strategies from the literature (see Table 3.4) on the basis of estimated expected NNS, which is a measure of mortality reduction solely. Here, however, we re-examine them on the basis of estimated expected QG, which is a measure that incorporates both mortality reduction and quality-of-life degradations incurred as a result of specific clinical tests and interventions. For each of the compared risk groups, we estimated the expected QG for the corresponding optimized strategy from Figures 4.15a through 4.15e and the strategies in Table 3.4. The numerical results of these comparisons, using basecase numerical estimates, is shown for each of the compared risk groups in Tables 4.4 through 4.8.

Table 4.4: Comparison of the screening strategies from Chapter 3 with the optimized strategy for the whites risk group, $\hat{x}_W^*$, based on estimated expected QG. Estimates are shown with 95% CIs and based on populations (samples) of $10^7$ independently simulated patients.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Estimated E [Q] 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.0000 ± 0.0000</td>
</tr>
<tr>
<td>#2</td>
<td>-0.0067 ± 0.0008</td>
</tr>
<tr>
<td>#3</td>
<td>0.0029 ± 0.0008</td>
</tr>
<tr>
<td>#4</td>
<td>0.0043 ± 0.0008</td>
</tr>
<tr>
<td>#5</td>
<td>0.0024 ± 0.0008</td>
</tr>
<tr>
<td>#6</td>
<td>0.0112 ± 0.0009</td>
</tr>
<tr>
<td>#7</td>
<td>0.0025 ± 0.0008</td>
</tr>
<tr>
<td>#8</td>
<td>0.0020 ± 0.0008</td>
</tr>
<tr>
<td>#9</td>
<td>0.0174 ± 0.0008</td>
</tr>
<tr>
<td>#10</td>
<td>0.0079 ± 0.0008</td>
</tr>
<tr>
<td>#11</td>
<td>0.0138 ± 0.0007</td>
</tr>
<tr>
<td>#12</td>
<td>0.0088 ± 0.0004</td>
</tr>
<tr>
<td>#13</td>
<td>0.0073 ± 0.0005</td>
</tr>
<tr>
<td>#14</td>
<td>0.0002 ± 0.0005</td>
</tr>
<tr>
<td>$\hat{x}_W^*$</td>
<td>0.0881 ± 0.0010</td>
</tr>
</tbody>
</table>
Table 4.5: Comparison of the screening strategies from Chapter 3 with the optimized strategy for the African American risk group, \( \hat{x}_{AA}^* \), based on estimated expected QG. Estimates are shown with 95% CIs and based on populations (samples) of \( 10^7 \) independently simulated patients.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Estimated E [Ω]</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.0000</td>
<td>± 0.0000</td>
</tr>
<tr>
<td>#2</td>
<td>0.0099</td>
<td>± 0.0011</td>
</tr>
<tr>
<td>#3</td>
<td>0.0278</td>
<td>± 0.0012</td>
</tr>
<tr>
<td>#4</td>
<td>0.0282</td>
<td>± 0.0012</td>
</tr>
<tr>
<td>#5</td>
<td>0.0251</td>
<td>± 0.0012</td>
</tr>
<tr>
<td>#6</td>
<td>0.0477</td>
<td>± 0.0013</td>
</tr>
<tr>
<td>#7</td>
<td>0.0246</td>
<td>± 0.0012</td>
</tr>
<tr>
<td>#8</td>
<td>0.0277</td>
<td>± 0.0012</td>
</tr>
<tr>
<td>#9</td>
<td>0.0431</td>
<td>± 0.0012</td>
</tr>
<tr>
<td>#10</td>
<td>0.0367</td>
<td>± 0.0012</td>
</tr>
<tr>
<td>#11</td>
<td>0.0314</td>
<td>± 0.0011</td>
</tr>
<tr>
<td>#12</td>
<td>0.0200</td>
<td>± 0.0007</td>
</tr>
<tr>
<td>#13</td>
<td>0.0128</td>
<td>± 0.0008</td>
</tr>
<tr>
<td>#14</td>
<td>0.0001</td>
<td>± 0.0007</td>
</tr>
<tr>
<td>( \hat{x}_{AA}^* )</td>
<td>0.1426</td>
<td>± 0.0014</td>
</tr>
</tbody>
</table>

Table 4.6: Comparison of the screening strategies from Chapter 3 with the optimized strategy for the family history risk group, \( \hat{x}_{FH}^* \), based on estimated expected QG. Estimates are shown with 95% CIs and based on populations (samples) of \( 10^7 \) independently simulated patients.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Estimated E [Ω]</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.0000</td>
<td>± 0.0000</td>
</tr>
<tr>
<td>#2</td>
<td>-0.0043</td>
<td>± 0.0008</td>
</tr>
<tr>
<td>#3</td>
<td>0.0073</td>
<td>± 0.0009</td>
</tr>
<tr>
<td>#4</td>
<td>0.0095</td>
<td>± 0.0009</td>
</tr>
<tr>
<td>#5</td>
<td>0.0075</td>
<td>± 0.0009</td>
</tr>
<tr>
<td>#6</td>
<td>0.0187</td>
<td>± 0.0010</td>
</tr>
<tr>
<td>#7</td>
<td>0.0069</td>
<td>± 0.0009</td>
</tr>
<tr>
<td>#8</td>
<td>0.0071</td>
<td>± 0.0009</td>
</tr>
<tr>
<td>#9</td>
<td>0.0242</td>
<td>± 0.0009</td>
</tr>
<tr>
<td>#10</td>
<td>0.0138</td>
<td>± 0.0009</td>
</tr>
<tr>
<td>#11</td>
<td>0.0190</td>
<td>± 0.0008</td>
</tr>
<tr>
<td>#12</td>
<td>0.0115</td>
<td>± 0.0005</td>
</tr>
<tr>
<td>#13</td>
<td>0.0095</td>
<td>± 0.0005</td>
</tr>
<tr>
<td>#14</td>
<td>0.0010</td>
<td>± 0.0005</td>
</tr>
<tr>
<td>( \hat{x}_{FH}^* )</td>
<td>0.1052</td>
<td>± 0.0011</td>
</tr>
</tbody>
</table>
Table 4.7: Comparison of the screening strategies from Chapter 3 with the optimized strategy for the CCI=1 risk group, $\hat{x}_{C1}$, based on estimated expected QG. Estimates are shown with 95% CIs and based on populations (samples) of $10^7$ independently simulated patients.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Estimated E [Q]</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.0000</td>
<td>± 0.0000</td>
</tr>
<tr>
<td>#2</td>
<td>-0.0269</td>
<td>± 0.0006</td>
</tr>
<tr>
<td>#3</td>
<td>-0.0236</td>
<td>± 0.0007</td>
</tr>
<tr>
<td>#4</td>
<td>-0.0221</td>
<td>± 0.0007</td>
</tr>
<tr>
<td>#5</td>
<td>-0.0259</td>
<td>± 0.0007</td>
</tr>
<tr>
<td>#6</td>
<td>-0.0229</td>
<td>± 0.0007</td>
</tr>
<tr>
<td>#7</td>
<td>-0.0240</td>
<td>± 0.0007</td>
</tr>
<tr>
<td>#8</td>
<td>-0.0263</td>
<td>± 0.0007</td>
</tr>
<tr>
<td>#9</td>
<td>-0.0143</td>
<td>± 0.0007</td>
</tr>
<tr>
<td>#10</td>
<td>-0.0240</td>
<td>± 0.0007</td>
</tr>
<tr>
<td>#11</td>
<td>-0.0120</td>
<td>± 0.0006</td>
</tr>
<tr>
<td>#12</td>
<td>0.0036</td>
<td>± 0.0004</td>
</tr>
<tr>
<td>#13</td>
<td>-0.0049</td>
<td>± 0.0004</td>
</tr>
<tr>
<td>#14</td>
<td>-0.0150</td>
<td>± 0.0004</td>
</tr>
<tr>
<td>$\hat{x}_{C1}$</td>
<td>0.0349</td>
<td>± 0.0008</td>
</tr>
</tbody>
</table>

Table 4.8: Comparison of the screening strategies from Chapter 3 with the optimized strategy for the CCI=2 risk group, $\hat{x}_{C2}$, based on estimated expected QG. Estimates are shown with 95% CIs and based on populations (samples) of $10^7$ independently simulated patients.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Estimated E [Q]</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.0000</td>
<td>± 0.0000</td>
</tr>
<tr>
<td>#2</td>
<td>-0.0346</td>
<td>± 0.0005</td>
</tr>
<tr>
<td>#3</td>
<td>-0.0350</td>
<td>± 0.0006</td>
</tr>
<tr>
<td>#4</td>
<td>-0.0338</td>
<td>± 0.0006</td>
</tr>
<tr>
<td>#5</td>
<td>-0.0382</td>
<td>± 0.0006</td>
</tr>
<tr>
<td>#6</td>
<td>-0.0384</td>
<td>± 0.0006</td>
</tr>
<tr>
<td>#7</td>
<td>-0.0354</td>
<td>± 0.0006</td>
</tr>
<tr>
<td>#8</td>
<td>-0.0386</td>
<td>± 0.0006</td>
</tr>
<tr>
<td>#9</td>
<td>-0.0297</td>
<td>± 0.0006</td>
</tr>
<tr>
<td>#10</td>
<td>-0.0383</td>
<td>± 0.0006</td>
</tr>
<tr>
<td>#11</td>
<td>-0.0244</td>
<td>± 0.0005</td>
</tr>
<tr>
<td>#12</td>
<td>0.0006</td>
<td>± 0.0003</td>
</tr>
<tr>
<td>#13</td>
<td>-0.0109</td>
<td>± 0.0004</td>
</tr>
<tr>
<td>#14</td>
<td>-0.0215</td>
<td>± 0.0003</td>
</tr>
<tr>
<td>$\hat{x}_{C2}$</td>
<td>0.0148</td>
<td>± 0.0005</td>
</tr>
</tbody>
</table>
For graphical comparison of the data in Tables 4.4 through 4.8, Figure 4.16 depicts the expected QG estimates for all risk groups for all of strategies #1–#14 from Chapter 3 and for strategies $\hat{x}_W^*$, $\hat{x}_{AA}^*$, $\hat{x}_{FH}^*$, $\hat{x}_{C1}^*$, and $\hat{x}_{C2}^*$.

From Figure 4.16, it is evident that the three-stage optimization method produces strategies that are statistically significantly better in terms of expected QG than those representative strategies from the literature listed in Table 3.4. The improvement from using the optimized strategies relative to using the best strategy from the literature is substantially greater for the African American risk group than for the others. Remarkably, each of the optimized strategies prescribes prostate biopsy irrespective of PSA level for all men at certain ages. It is difficult to infer meaningful differences between the optimized
strategies for the White, African American, and Family History risk groups. Considering the comorbidity risk groups, however, there does appear to be a trend of ending screening earlier in patients with more serious comorbidities.

Because it is more natural from a clinical perspective to compare screening strategies by NNSQ or NNS than by expected QG, we present in Figures 4.17 and 4.18 the estimated expected NNSQ and NNS, respectively, for each of the basecase optimized strategies in this chapter.

![Figure 4.17](image_url)

Figure 4.17: Comparison of the estimated expected NNSQ for the basecase optimized strategies depicted in Figures 4.15a through 4.15e based on samples of $10^7$ simulated patients from the corresponding risk groups. The 95% CIs reflect inter-patient stochastic variability; they do not reflect generalized uncertainty in any of the model parameters.
Figure 4.18: Comparison of the estimated expected NNS for the basecase optimized strategies depicted in Figures 4.15a through 4.15e based on samples of $10^7$ simulated patients from the corresponding risk groups. The 95% CIs reflect inter-patient stochastic variability; they do not reflect generalized uncertainty in any of the model parameters.

### 4.3.3 Single-Period Screening Model Analysis

In view of the fact that the basecase optimized strategy for each risk group consists of either 1 or 2 screen events (biopsies, in fact), we carry out some computational analysis of the $\vartheta$ ($\phi$) function from Equation 4.11 as a means of independently assessing the validity of the basecase results. In particular, we use the analytical single-period model to identify the theoretical optimal $\phi^*$ (within the range of assumed feasible values $\phi = 0, 0.5, \ldots, 10$ ng/mL). The robustness of this $\phi^*$ is examined by also computing the corresponding optima over varied levels of the following parameters: (i) the QALYs disutility of biopsy ($D_{\text{Biop}}$, or $\delta$ in the nomenclature of the single period model); (ii) the annual probability $e_t$ of metastasis of undetected PCa; (iii) the PCa leadtime $t_{LT}$; and (iv) the mean postonset slope-increment $\mu \beta_2$ of the PSA sampling model. We examine how variation in these parameters affects the theoretical optimal single-period threshold $\phi^*$ for men at age 53 in the reference risk group, based on the fact that, as shown in Figure 4.15a, the first (and thus likely most significant) of the two always-biopsy (0 ng/mL) thresholds for the reference risk group occurs at age 53. Also, for comparison, we examine how variation in these
parameters affects $\phi^*$ for men at age 69 in the reference risk group, because age 69 is a common age beyond which routine PSA-based screening is nearly universally discouraged.

In this analysis, rather than directly solving for $\phi^*$ using Equation (4.16), we compute and plot $\hat{\theta}(\phi)$ for $\phi = 0, 0.5, \ldots, 10$ ng/mL, so that the estimated optimum $\hat{\phi}^*$ for the given curve is taken to be the feasible value of $\phi$ for which the point estimate $\hat{\theta}(\phi)$ is largest. This method of analysis provides insight not only into the optimal single-period threshold, but also into the shape of the $\theta(\phi)$ function, which may provide further insight into the influence of the varied model parameters.

For each combination of experimental parameter level and $\phi$, we computed the quantities required by the CI estimator for $\theta(\phi)$ in Equation (4.78) in the following manner, where all nonvaried parameters are set to their basecase levels for the reference risk group. First, a sample of approximately 500,000 simulated patients was used to compute $\pi_{Ca}$. Second, a different independently-seeded sample of approximately 500,000 simulated patients was used to compute $\hat{b}(\phi)$ and $S^2_{\hat{b}(\phi)}$. Third, yet another independently seeded sample of approximately 500,000 simulated patients was used to compute $\bar{G}(\phi)$ and $S^2_{\bar{G}(\phi)}$. Fourth, yet another independently seeded sample of approximately 500,000 simulated patients was used to compute $\bar{H}(\phi)$ and $S^2_{\bar{H}(\phi)}$. And, finally, yet another independently seeded sample of approximately 500,000 simulated patients was used to compute sample-mean estimates of $E[t_{0,j}|Ca, t]$ and $\text{Var}[t_{0,j}|Ca, t]$.

The analysis of single-period screening at age 53 is shown in Figure 4.19, where each subfigure depicts the effect of one of parameters $\delta, e_t, t_{LT}$, or $\mu_{\beta_2}$ on the single-period expected reward. Figure 4.19a shows the $\hat{\theta}(\phi)$ curves for the SA over the biopsy disutility parameter, where $\delta = 0.005, 0.010, \ldots, 0.050$. At values of $\delta \leq 0.015$ (which includes the basecase level 0.00577), the estimated theoretical optimum $\hat{\phi}^*$ is 0 ng/mL (i.e., always biopsy). Figure 4.19b shows the $\hat{\theta}(\phi)$ curves for the SA over the annual probability of metastasis of undetected PCa ($e_t$), where $e_t$ is varied to the low and high levels described in the SA for Chapter 3 (see Section 3.3.4). Figure 4.19c shows the $\hat{\theta}(\phi)$ curves for the SA over the PCa leadtime ($t_{LT}$), where $t_{LT}$ is varied $\pm 2$ years from the basecase estimate of 12.8 years. Figure 4.19d shows the $\hat{\theta}(\phi)$ curves for the SA over the mean postonset slope-increment ($\mu_{\beta_2}$) of the PSA sampling model, where in addition to the basecase value $\mu_{\beta_2} = 0.1094$, we let $\mu_{\beta_2} = 0.0, 0.5,$ and $1.0$.  

163
Figure 4.19: Plots of the $\hat{\theta}$ ($\phi$) function showing the effect of variation in the certain model parameters on the single-period expected reward for screening 53-year-old men from the reference risk group (the White risk group) at different PSA thresholds ($\phi = 0.0, 0.5, \ldots, 10.0$ ng/mL).
The analysis on the QALYs disutility of biopsy in Figure 4.19a provides some support for the always-biopsy thresholds obtained by the three-stage method for the reference risk group (Figure 4.15a). For higher biopsy thresholds the optimal threshold appears to be close to 1 ng/mL, which is consistent with the 2015 NCCN guidelines [15] in that the guidelines treat PSA levels \( \geq 1 \) as warranting more frequent screening (owing to the patient being at higher risk). Although there is great overlap among the 95% CIs, just considering the sample-mean point estimates one can make some tentative observations. Interestingly, in distinction to the \( \hat{\theta}(\phi) \) curves for the various other parameters shown in Figures 4.19b through 4.19d, different levels of parameter \( \delta \) do affect not only the magnitude of the responses but also the characteristic shape of the response surface (curve). That is, we can see from Figure 4.19a that for values of \( \delta \) above some threshold level, the optimal single-period decision is no longer to always biopsy, independent of PSA. This is intuitive: eventually, for large enough values of \( \delta \), the expected future reward from biopsy will be so substantially offset by the one-time quality-of-life decrement from the biopsy procedure that such an aggressive always-biopsy strategy no longer is justified. For values \( \phi \geq 1 \) ng/mL, the \( \hat{\theta}(\phi) \) curves in Figure 4.19a are indistinguishable because \( \hat{\theta}(\phi) \) is an expected future reward across all patients, whereas the disutility \( \delta \) is a small one-time reward decrement that only pertains to those patients with PSA levels rising high enough to warrant biopsy. At the value \( \phi = 0 \), however, virtually all men will have their future reward decreased by \( \delta \), so that at \( \phi = 0 \) the effect of variation in \( \delta \) on \( \hat{\theta}(\phi) \) is far more pronounced. Estimates of disutility parameters, such as \( D_{\text{Biop}} \), are uncertain and somewhat subjective. So, it is noteworthy that the aggressive always-biopsy optimized strategy in the single-period analysis is robust to plausible variation in the disutility of biopsy.

In Figure 4.19b, the estimated theoretical optimum \( \phi^* \) at the basecase and high levels of \( e_t \) is 0 ng/mL (i.e., always biopsy), which is consistent with the basecase optimized strategy from the three-stage method. On the other hand, at the low level of \( e_t \), there does not appear to be a statistically significant difference between thresholds. This is consistent with, but does not conclusively establish, the hypothesis that at a very low level of \( e_t \), screening is not significantly beneficial relative to no screening. In contrast to the curves in Figure 4.19a, the curves in Figure 4.19b are quite distinguishable at all plotted values of \( \phi \). This is reasonable because, rather than a small one-time decrement, parameter \( e_t \) controls a major...
physiological process (prostate tumor growth) that highly impacts survival.

In Figure 4.19c, the estimated theoretical optimum $\hat{\phi}^*$ at each level of $t_{LT}$ is 0 ng/mL, though there is great overlap between many of the 95% CIs. This is consistent with the basecase optimized strategy from the three-stage method, though these curves taken together but apart from the other SA experiments do not constitute statistically-powerful support. However, at the high level of $t_{LT}$, there is a statistically significant difference between $\hat{\phi}(0)$ and $\hat{\phi}(\phi) \neq 0$. Based on the width of the CIs and the shape of the point-estimate time series, it seems that the optimal screening threshold $\phi^*$ may be fairly robust to changes in $t_{LT}$, but that changes in $t_{LT}$ may cause significant changes in the magnitude of $\hat{\phi}(\phi^*)$.

Given that the lower and upper bounds of the posterior 95% CI estimate of $\mu_{\beta_2}$ from Gulati et al. [40] are, respectively, 0.0919 and 0.1269, the nonbasecase levels in the present SA are quite wild. Nonetheless, because PSA growth is such a critical component of the PCa-screening optimization problem, the analysis in Figure 4.19d is helpful. And at each of the considered levels except for $\mu_{\beta_2} = 1.0$ (which is an order of magnitude larger than the upper bound source estimate), there is considerable evidence that the estimated theoretical optimum $\hat{\phi}^*$ is 0 ng/mL, though there is moderate CI overlap at $\mu_{\beta_2} = 5.0$ (which is nearly 4 times the upper bound source estimate).

The analysis of single-period screening at age 69 is shown in Figure 4.20, with subfigures corresponding to the subfigures of the analysis of screening at age 53 in Figure 4.19. As in the case of Figure 4.19, definitive analysis of the plots in Figure 4.20 is impeded by the overlapping confidence intervals. Comparing Figures 4.19a and 4.20a, we have the relatively strong suggestion that, unlike at age 53, single-period screening at age 69 is not optimal: the curves in Figure 4.20a seem to increase with PSA threshold, which suggests that no-screening may be optimal in single-period screening at age 69. The same trend shows up in Figure 4.20c, although it is less stark. Comparing Figures 4.19b and 4.20b, the same trend faintly appears at the low and basecase levels of $e_t$. At the high level of $e_t$, however, it seems that single-period screening still has benefit at age 69. Comparing Figures ?? and ??, this trend of screening no longer being beneficial at age 69 only possibly shows up at the two smallest levels of the postonset slope-increment mean.
Figure 4.20: Plots of the $\hat{\phi}$ ($\phi$) function showing the effect of variation in the certain model parameters on the single-period expected reward for screening 69-year-old men from the reference risk group (the White risk group) at different PSA thresholds ($\phi = 0.0, 0.5, \ldots, 10.0$ ng/mL).
4.3.4 Sensitivity Analysis on Maximum Number of Biopsies

In this chapter, we have been limiting patients to no more than 3 biopsies each. In order to test whether our results are biased by the allowance of more than 1 biopsy per patient, we ran the three-stage method at basecase parameter settings but permitted no patient to undergo more than 1 biopsy. The optimized strategies produced in this experiment for the various risk groups are shown in Figures 4.21a through 4.21e.

Comparing Figures 4.21a through 4.21e with Figures 4.15a through 4.15e, respectively, it seems that the allowance for an unlimited number of biopsies does not markedly bias the basecase results. There is some difference between the respective Family History optimized strategies, but it is unclear how meaningful those differences are. For each of the White, African American, and Family History risk groups, the optimized strategy in this experiment and in Section 4.3.2 contains two screening thresholds. Whereas in Section 4.3.2, the earlier screening thresholds are set to 0 ng/mL, in this experiment the earlier screening thresholds are set to values >0 ng/mL (but still relatively low). This is reasonable due to the fact that, in this experiment, if the earlier screening thresholds were in fact set to 0 ng/mL, then the later screening thresholds would be rendered meaningless, since all patients would have already used up their total allowed biopsies by the time the later screening comes around.
(a) Strategy for the White risk group.  
(b) Strategy for the African American risk group.  
(c) Strategy for the Family History risk group.  
(d) Strategy for the CCI=1 risk group.  
(e) Strategy for the CCI≥2 risk group. 

Figure 4.21: Optimized strategies produced by the three-stage method for the White, African American, Family History, CCI=1, and CCI≥2 risk groups, respectively, where no patient is permitted more than 1 prostate biopsy.
4.3.5 Sensitivity Analysis on Model Parameters

We performed one-way sensitivity analysis over the PCa leadtime, the QALY disutilities, and the following model parameters: the annual probability of metastasis for patients with undiagnosed PCa, \( e_t \); and the annual probability of other-cause death, \( d_t \). These probabilities, along with the other parameters used to compute the Markov state-transition probabilities in the natural history model, are repeated for convenience in Table 4.9.

Table 4.9: Definitions of parameters for transition probability matrices.

<table>
<thead>
<tr>
<th>Param.</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( w_t )</td>
<td>Annual probability of developing PCa</td>
</tr>
<tr>
<td>( d_t )</td>
<td>Annual probability of other-cause death</td>
</tr>
<tr>
<td>( b_t )</td>
<td>Annual probability of metastasis for patients in state T</td>
</tr>
<tr>
<td>( e_t )</td>
<td>Annual probability of metastasis for patients in state C</td>
</tr>
<tr>
<td>( z_t )</td>
<td>Annual probability of PCa-specific death for patients in state M</td>
</tr>
<tr>
<td>( f )</td>
<td>Prostate biopsy sensitivity</td>
</tr>
</tbody>
</table>

For the African American risk group, in place of parameter \( d_t \) we used the corresponding risk-adjusted parameter \( d_t^{AA} \). For the CCI=1 and CCI\( \geq 2 \) risk groups, in place of parameter \( d_t \) we used the corresponding risk-adjusted parameters \( d_t^{C1} \) and \( d_t^{C2} \), respectively.

**Estimates.** For parameters \( e_t \) and \( d_t \), we used the low and high levels (estimates) that were used in the Chapter 3 sensitivity analysis, which is described in Section 3.3.4. For PCa leadtime, we varied ±5 years the basecase estimate of 11.75 years. For the nonreference risk group parameters (\( d_t^{AA} \), \( d_t^{C1} \), and \( d_t^{C2} \)), we derived the low and high levels from the low and high levels of the corresponding reference risk-group parameters to derive the low and high levels of the nonreference risk group parameters. For the QALY disutilities, we used the low and high estimates from their respective sources as the low and high levels in the sensitivity analysis. The low and high levels of the QALY disutilities are shown in Table 4.10.
Table 4.10: Numerical estimates used as the low and high levels for QALY disutilities in the sensitivity analysis.

<table>
<thead>
<tr>
<th>Disutility</th>
<th>Symbol</th>
<th>Low level</th>
<th>High level</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening</td>
<td>$D_{Scr}$</td>
<td>0.0</td>
<td>0.00019</td>
<td>[29, 23, 48]</td>
</tr>
<tr>
<td>Biopsy</td>
<td>$D_{Biop}$</td>
<td>0.0</td>
<td>0.05</td>
<td>[23, 48]</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>$D_{Dia}$</td>
<td>0.01250</td>
<td>0.02083</td>
<td>[65, 48]</td>
</tr>
<tr>
<td>Treatment</td>
<td>$D_{Tre}$</td>
<td>0.09167</td>
<td>0.32333</td>
<td>[101, 14, 48]</td>
</tr>
<tr>
<td>Recovery</td>
<td>$D_{Rec}$</td>
<td>0.0</td>
<td>0.7</td>
<td>[87, 101, 48]</td>
</tr>
<tr>
<td>Metastasis</td>
<td>$D_{Met}$</td>
<td>0.1</td>
<td>0.7</td>
<td>[12]</td>
</tr>
</tbody>
</table>

At each parameter configuration (experiment), we ran the three-stage optimization method using the same settings that were used to generate the basecase results presented in Section 4.3.2. After each experiment, we estimated the expected QG of the optimized strategy for that experiment using a sample of $10^7$ independently simulated patients subject to the optimized strategy. For the reference risk group, Table 4.11 contains the 95% CI estimates for expected QG of the optimized strategy produced at each experiment. The corresponding results for the African American, Family History, CCI=1, and CCI≥2 risk groups are contained in Tables 4.12 through 4.15, respectively.

Table 4.11: Expected QG estimates from one-way sensitivity analysis over model parameters for the White risk group. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}$, $D_{Dia}$, $D_{Biop}$, $D_{Tre}$, and $D_{Met}$) were simultaneously varied to their lowest or highest plausible levels.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Low level</th>
<th>Basecase</th>
<th>High level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_t$</td>
<td>0.1063 ± 0.0011</td>
<td>0.0881 ± 0.0010</td>
<td>0.0651 ± 0.0009</td>
</tr>
<tr>
<td>$e_t$</td>
<td>0.0000 ± 0.0000</td>
<td>0.0881 ± 0.0010</td>
<td>0.2419 ± 0.0015</td>
</tr>
<tr>
<td>leadtime</td>
<td>0.0761 ± 0.0010</td>
<td>0.0881 ± 0.0010</td>
<td>0.0984 ± 0.0011</td>
</tr>
<tr>
<td>$D_*$</td>
<td>0.1380 ± 0.0011</td>
<td>0.0881 ± 0.0010</td>
<td>0.0365 ± 0.0010</td>
</tr>
</tbody>
</table>
Table 4.12: Expected QG estimates from one-way sensitivity analysis over model parameters for the African Americans risk group. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}$, $D_{Dia}$, $D_{Biop}$, $D_{Tre}$, and $D_{Met}$) were simultaneously varied to their lowest or highest plausible levels.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Low level</th>
<th>Basecase</th>
<th>High level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_t$</td>
<td>0.2291 ± 0.0017</td>
<td>0.1534 ± 0.0015</td>
<td>0.1123 ± 0.0013</td>
</tr>
<tr>
<td>$e_t$</td>
<td>0.0000 ± 0.0000</td>
<td>0.1534 ± 0.0015</td>
<td>0.5739 ± 0.0022</td>
</tr>
<tr>
<td>leadtime</td>
<td>0.1472 ± 0.0015</td>
<td>0.1534 ± 0.0015</td>
<td>0.1826 ± 0.0015</td>
</tr>
<tr>
<td>$D_*$</td>
<td>0.2458 ± 0.0015</td>
<td>0.1534 ± 0.0015</td>
<td>0.1032 ± 0.0015</td>
</tr>
</tbody>
</table>

Table 4.13: Expected QG estimates from one-way sensitivity analysis over model parameters for the Family History risk group. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}$, $D_{Dia}$, $D_{Biop}$, $D_{Tre}$, and $D_{Met}$) were simultaneously varied to their lowest or highest plausible levels.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Low level</th>
<th>Basecase</th>
<th>High level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_t$</td>
<td>0.1566 ± 0.0014</td>
<td>0.1052 ± 0.0012</td>
<td>0.0838 ± 0.0011</td>
</tr>
<tr>
<td>$e_t$</td>
<td>0.0000 ± 0.0000</td>
<td>0.1052 ± 0.0012</td>
<td>0.3053 ± 0.0016</td>
</tr>
<tr>
<td>leadtime</td>
<td>0.1011 ± 0.0011</td>
<td>0.1052 ± 0.0012</td>
<td>0.1249 ± 0.0012</td>
</tr>
<tr>
<td>$D_*$</td>
<td>0.1703 ± 0.0012</td>
<td>0.1052 ± 0.0012</td>
<td>0.0603 ± 0.0011</td>
</tr>
</tbody>
</table>

Table 4.14: Expected QG estimates from one-way sensitivity analysis over model parameters for the CCI=1 risk group. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}$, $D_{Dia}$, $D_{Biop}$, $D_{Tre}$, and $D_{Met}$) were simultaneously varied to their lowest or highest plausible levels.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Low level</th>
<th>Basecase</th>
<th>High level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_t$</td>
<td>0.0563 ± 0.0009</td>
<td>0.0349 ± 0.0008</td>
<td>0.0207 ± 0.0006</td>
</tr>
<tr>
<td>$e_t$</td>
<td>0.0000 ± 0.0000</td>
<td>0.0349 ± 0.0008</td>
<td>0.1666 ± 0.0012</td>
</tr>
<tr>
<td>leadtime</td>
<td>0.0284 ± 0.0007</td>
<td>0.0349 ± 0.0008</td>
<td>0.0392 ± 0.0008</td>
</tr>
<tr>
<td>$D_*$</td>
<td>0.0740 ± 0.0009</td>
<td>0.0349 ± 0.0008</td>
<td>0.0034 ± 0.0003</td>
</tr>
</tbody>
</table>
Table 4.15: Expected QG estimates from one-way sensitivity analysis over model parameters for the CCI≥2 risk group. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}$, $D_{Dia}$, $D_{Biop}$, $D_{Tre}$, and $D_{Met}$) were simultaneously varied to their lowest or highest plausible levels.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Low level</th>
<th>Basecase</th>
<th>High level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_t$</td>
<td>0.0270 $\pm$ 0.0007</td>
<td>0.0148 $\pm$ 0.0005</td>
<td>0.0074 $\pm$ 0.0004</td>
</tr>
<tr>
<td>$e_t$</td>
<td>0.0000 $\pm$ 0.0000</td>
<td>0.0148 $\pm$ 0.0005</td>
<td>0.0938 $\pm$ 0.0009</td>
</tr>
<tr>
<td>leadtime</td>
<td>0.0129 $\pm$ 0.0005</td>
<td>0.0148 $\pm$ 0.0005</td>
<td>0.0162 $\pm$ 0.0005</td>
</tr>
<tr>
<td>$D_*$</td>
<td>0.0481 $\pm$ 0.0006</td>
<td>0.0148 $\pm$ 0.0005</td>
<td>0.0013 $\pm$ 0.0002</td>
</tr>
</tbody>
</table>

The optimized strategies that were produced by the three-stage method in the SA and that are represented in Tables 4.11 through 4.15 are shown in Figures 4.22 through 4.26.
Figure 4.22: Optimized strategies produced by the three-stage method for the White risk group at different experiments in the sensitivity analysis. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}$, $D_{Dia}$, $D_{Biop}$, $D_{Tre}$, and $D_{Met}$) were simultaneously varied to their lowest or highest plausible levels.
Figure 4.23: Optimized strategies produced by the three-stage method for the African American risk group at different experiments in the sensitivity analysis. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}$, $D_{Dia}$, $D_{Biop}$, $D_{Tre}$, and $D_{Met}$) were simultaneously varied to their lowest or highest plausible levels.
Figure 4.24: Optimized strategies produced by the three-stage method for the Family History risk group at different experiments in the sensitivity analysis. The symbol $D_e$ denotes that all QALY disutilities ($D_{Scr}$, $D_{Dia}$, $D_{Biop}$, $D_{Tre}$, and $D_{Met}$) were simultaneously varied to their lowest or highest plausible levels.
Figure 4.25: Optimized strategies produced by the three-stage method for the CCI=1 risk group at different experiments in the sensitivity analysis. The symbol $\mathbb{D}_*$ denotes that all QALY disutilities ($\mathbb{D}_{\text{Scr}}, \mathbb{D}_{\text{Dia}}, \mathbb{D}_{\text{Biop}}, \mathbb{D}_{\text{Tre}}, \text{ and } \mathbb{D}_{\text{Met}}$) were simultaneously varied to their lowest or highest plausible levels.
Figure 4.26: Optimized strategies produced by the three-stage method for the CCI≥2 risk group at different experiments in the sensitivity analysis. The symbol $D_*$ denotes that all QALY disutilities ($D_{Sc}$, $D_{Dia}$, $D_{Biop}$, $D_{Tre}$, and $D_{Met}$) were simultaneously varied to their lowest or highest plausible levels.
To examine some of the most obvious or striking differences between this model and the model used in Underwood et al. [107], we conducted additional SA experiments over the following: (i) the preclinical dwelling time; (ii) the QALY disutility of biopsy; (iii) the per-patient maximum number of biopsies; and (iv) the PSA sampling method. In particular, to more closely approximate the model in [107], we set the disutility of biopsy to 0.05 QALYs, we set the per-patient maximum number of biopsies to 1, and we replaced the PSA sampling method of Gulati et al. [40] with the method in [107]. With these settings, we ran the three-stage method at each of the following levels of the preclinical dwelling time: 0.0, 0.4, 0.8, and 12.0 years. The optimized strategies produced by these experiments are shown in Figure 4.27. The only definite difference we observed between the basecase results for Whites in this research and the results from these SA experiments occurred at the lowest levels of the preclinical dwelling time. As the preclinical dwelling time decreased, the length of time between the age at initial screen and the age at final screen increased. Specifically, the age at initial screen is earlier when the preclinical dwelling time is eliminated (i.e., set to 0.0 years). This is reasonable because the shorter the preclinical dwelling time, the earlier PCa can metastasize. It is also important to note that the PSA thresholds did not substantially change even though an entirely different and independent PSA sampling method was used. That is, in each of Figures 4.27a through 4.27d, there is at least 1 always-biopsy threshold. This fact decreases the likelihood that the low PSA thresholds we observe in our results are an artifact of the Gulati et al. [40] sampling method.
Figure 4.27: Optimized strategies produced by the three-stage method for the reference risk group at different levels of the preclinical dwelling time ($t_{PD}$) parameter.

4.4 Discussion

4.4.1 PSA Screening

In Chapter 3, we saw that strategy #6 (screening annually from age 50 to 75 using PSA threshold 2.5 ng/mL) yielded the smallest estimated expected NNS for all the risk groups. In this chapter, the strategies produced by the three-stage method using basecase parameter values—the strategies in Figures 4.15a through 4.15e—are only slightly similar to strategy #6 in that they suggest the use of low PSA thresholds.
However, the strategies produced by optimization in this chapter are different in that they imply biopsying all men, regardless of PSA level. Furthermore, in some cases (White and CCI=1 risk groups) the recommendation is to biopsy twice due to the imperfect sensitivity of biopsy. The most prominent conclusion that can be inferred from the optimization results in this chapter is that a policy in which all men undergo prostate biopsy at some time in their lives maximizes expected QG. This inference finds further support from the analysis carried out in Section 4.3.3. In that analysis, we see somewhat-independent analytical confirmation of the always-biopsy rule that the strategy in Figure 4.15d embodies. The fact that this confirmation comes from the context of the CCI=1 risk group, which represents a relatively low-risk population, makes the inference to biopsying virtually all men in the White, African American, and Family History risk groups more reliable.

As Figures 4.17 and 4.18 clearly show, there are marked differences in the number of patients needed to be screened to make a unit gain either in mortality or in QALYs. The most striking differences lie between the two comorbidity risk groups, and between the comorbidity risk groups and the noncomorbidity risk groups. Although on the basis of PCa mortality (measured by NNS), the African American risk group appears to benefit significantly more than either the White or Family History risk group, on the basis of QALYs (measured by NNSQ) the disparity is less obvious or extreme.

The strategy in Figure 4.15d, biopsying all men in the CCI=1 risk group at age 51 and again at age 52 is a product of the imperfect sensitivity of the prostate biopsy procedure. This was confirmed by setting the biopsy sensitivity to $f = 1.0$ (perfect sensitivity, i.e., no false-negative results) and running the clean-up procedure again. With the sensitivity $f = 1.0$, the biopsy at age 52 was eliminated. That is, if the biopsy procedure had perfect sensitivity, only 1 biopsy would be necessary. This suggests that improvements in biopsy technology such as fusion biopsy, which uses magnetic resonance imaging to guide biopsy [77], could substantially reduce the number of biopsies in a near optimal screening strategy.

Comparing the optimized strategies in Figures 4.15a through 4.15e one with another, it is difficult to draw conclusions about meaningful differences in the characteristics of near-optimal screening in each risk group. For the White, African American, and Family History risk groups, the optimized strategies basically amount to biopsying all patients by some time between the ages of 55 and 60. Note that in
“one-off” post-optimization analysis, we found that the clean-up procedure eliminated the two thresholds 7.0 ng/mL from the strategy shown in Figure 4.15b. The strategy in Figure 4.15b demonstrates that the third-stage COMPASS method can, in a minority of cases, achieve higher regularity at the cost of introducing noise (i.e., practically insignificant thresholds) after the second-stage clean-up procedure is run.

With respect to the comorbidity risk groups, however, the nature of the optimized strategies does possess a distinguishing characteristic. For the CCI=1 risk group, the biopsy should take place earlier in life than for the noncomorbidity risk groups. Also, the biopsy should take place earlier in life for the CCI≥2 risk group than for the CCI=1 risk group. This is reasonable because, with the comorbidity risk groups, it is precisely those earlier-onset cancers that are likely to cause PCa-specific death, insofar as they may affect the patient before the existing comorbidities are able to cause death. By the same logic, it makes sense that the more severe comorbidity risk group (CCI≥2) would have an earlier biopsy than the more moderate comorbidity risk group (CCI=1).

From the sensitivity analysis in Section 4.3.5, it is clear that men in risk groups CCI=1 and CCI≥2 in general stand to garner a smaller benefit in QALYs from screening that do the noncomorbidity risk groups. Also, it is evident that QALYs are most sensitive to changes in parameter \( e_t \), the probability of metastasis of undetected PCa, and the PCa leadtime. In this sensitivity analysis, the only experimental levels that did not contain 0 ng/mL thresholds were the low level of \( e_t \) and the most unfavorable (high) level of \( D^* \). For each risk group, the optimized strategy at the low level of \( e_t \) was no screening. At the high level of \( D^* \), the optimized strategy for risk groups CCI=1 and CCI≥2 was to screen once at age 53 at PSA thresholds, respectively, 3.5 and 8.5 ng/mL. Given that varying \( e_t \) within a plausible range caused the optimized strategy to vary from an always-biopsy strategy to no screening, it is extremely important to improve our understanding of the true rates of metastasis of undetected PCa. It is also important to obtain better estimates of PCa leadtime. In our base-case analysis, we used 11.75 years as the estimate for the PCa leadtime (i.e., the time after PCa onset when the PCa will be detected clinically if it has not been detected by screening). Insofar as a larger leadtime causes PCa screening to appear more beneficial by delaying clinical detection, it is important to accurately understand the relationship of screen-detection
rates to clinical-detection rates.

Judging from the sensitivity analysis in Section 4.3.4, there does not appear to be a large difference in near-optimal screening whether patients are permitted to undergo more than 1 biopsy. The optimized strategies in Sections 4.3.2 and 4.3.4 are quite similar.

### 4.4.2 Disease-Screening Modeling and Methodology

In this chapter, although we explicitly use the estimated expected QG as our QALYs-based performance measure, maximizing this performance measure is mathematically identical to minimizing estimated expected NNSQ. And we show that there are important advantages to using NNSQ both for disease screening modeling and for our medical understanding of the target disease. Methodologically, NNSQ is calculated in a way that naturally lends itself to using CRN as a variance reduction technique for more accurate simulation-based estimation of the gain (or loss) in QALYs arising from a given screening strategy. With the application of CRN, significantly less simulation effort is required to compare alternative screening strategies on the basis of sample-mean estimates of expected QALYs gained. Furthermore, it is easily seen that minimizing expected NNSQ maximizes expected QALYs, and so nothing is lost in using expected NNSQ in place of expected QALYs as the QALYs-based performance measure.

In addition to the advantages NNSQ brings to simulation of disease screening, NNSQ has the potential to sharpen our medical understanding of screening for diseases such as prostate cancer by balancing the range of benefits and harms associated with disease screening that are not accounted for by measures such as NNS, which looks only at overall disease mortality reduction. In the case of prostate cancer screening, subjecting all men to a prostate biopsy exam at a certain age without regard to PSA test results or individual risk factors may well nearly maximally reduce the prostate cancer–specific mortality rate. However, such an aggressive screening program would have a large negative impact on quality of life for many men who really never needed to be screened. Using NNS, this quality-of-life impact is not considered. On the other hand, using NNSQ we can use QALY estimates of the discomfort and risk of infection associated with the prostate biopsy procedure to directly account for the negative impact of biopsy on a patient’s quality of life. Similar arguments could be made for prostate cancer treatment,
and even for the relatively innocuous PSA blood test. When researchers and physicians contemplate
the value of screening in light of RCT results, they, of course, are mindful of the fact that aggressive
screening strategies that strongly reduce disease mortality may result in harms due to overdiagnosis and
overtreatment. But NNS, unlike NNSQ, does not quantify these harms. Therefore, we contend that the use
of NNSQ may more quantitatively facilitate analysis of the benefit-to-harm trade-off that physicians know
exists. Additionally, NNS and NNSQ can be used together to compute an estimate of the quality-adjusted
life-expectancy gains by patients whom screening actually prevented from dying due to prostate cancer.

The regularizing objective function approach can be helpful in the optimization problems comprising
time-series based decision variables where there is an inherent advantage to maintaining some degree of
regularity in the variation of the decision variable values over time. As in the case of the present prostate
cancer screening problem, such advantages may not be easily captured by the objective function, if at
all. The regularization approach provides a mechanism by which the objective function can be explicitly
made to capture such advantages.

4.4.3 Limitations

One limitation of our model is that we do not account for patient adherence to screening strategies. In
the real world, some patients will elect to discontinue screening regardless of the prevailing screening
recommendations. Similarly, not every screened patient will begin screening precisely when the prevailing
recommendations say he should. But our model serves as an optimistic measure of the potential benefit
of screening. All simulated patients perfectly follow the given strategy. Thus our model simulates
patients with perfect adherence. In addition to the modeling limitations discussed in Section 3.4, there
are limitations relating to the addition of QALYs to the model. Estimates of QALYs are subjective,
and therefore represent at best an average disutility of screening for PCa. However, we used sensitivity
analysis to establish some degree of robustness with respect to variation in these estimates.

Being a heuristic optimization procedure, the three-stage method does not guarantee a truly optimal
strategy. However, we address this uncertainty both by testing the heuristic method on test problems with
known optima, and by experimenting with different variants of the GA (i.e., alternative recombination
strategies).

Although we use the analytical single-period model to provide independent confirmation of the results from the three-stage method, the analytical and heuristic methods are not truly independent. Key terms in the analytical model are estimated by the same simulation model used to evaluate screening strategies in the heuristic method. However, the mechanisms by which the heuristic arrives at the optimized strategies (i.e., the mechanics in the GA and COMPASS) are not part of the analytical method at all. Thus the near-optimality of the strategies produced by the heuristic is somewhat reinforced by the results from the heuristic, even though the methods are not technically independent in every relevant way.

Another limitation of the method in this chapter is the lack of a systematic way to analyze samples of optimized strategies and derive statistically-significant differences between the nature of near-optimal screening in different risk groups. We address this limitation head-on in Chapter 5.
5.1 Introduction

In this chapter, we present a binary encoding scheme for representing screening strategies that will allow more direct analysis of key features of near-optimal screening for the compared risk groups from Chapter 3, while also guaranteeing perfectly regular screening strategies.

We define a screening strategy by combinations of certain *defining characteristics*. These defining characteristics are the following: the age at which to begin screening; the age after which to stop screening; the frequency of PSA screens; the PSA threshold in ng/mL at the first PSA screen; and the PSA threshold in ng/mL at the final PSA screen. Given certain combinations of these characteristics, along with 3 assumptions, a screening strategy can be fully defined. The assumptions are the following: (i) there is a single time horizon over which screening occurs; (ii) if there are more than 2 screens, then the screening frequency is uniform within the screening time horizon; and (iii) if the PSA thresholds in ng/mL change with age inside the screening time horizon, then they change at a constant rate (linearly).
There are two key motivations for adopting the binary encoding scheme presented in this chapter: (i) it guarantees perfectly regular screening strategies; (ii) it permits direct analysis of the defining characteristics defined above; and (iii) it is generally faster insofar as it avoids a regularizing stage. The regularity guaranteed by the encoding scheme is immediate: all strategies are uniform in frequency and linear in PSA thresholds. And rather than merely running batteries of optimization experiments and ultimately proffering what emerges to be the “best” strategies for the various risk groups, this encoding scheme permits a more direct analysis of the defining characteristics of near-optimal screening as a function of the various simulation model parameters, such as those defining the compared risk groups. Direct calculation and analysis of these defining characteristics on the strategies produced by the GA in Chapter 4 is more likely to be frustrated by noisy thresholds. Using the method in this chapter, however, there is a far smaller likelihood that these defining characteristics are strongly influenced by noise. Finally, the binary encoding scheme completely avoids the preliminary stage of GA-based exploration of highly irregular (erratic) screening strategies that must then be followed by the clean-up and regularization procedures to yield screening strategies with seemingly higher face validity.

This is an alternative approach to the regularizing objective function from Chapter 4 to dealing with the problem of highly variant PSA-threshold time-series produced by GA-based simulation optimization. The downside to this approach is that it involves a few reasonable constraining assumptions on the screening strategies it produces, whereas the regularization approach does not make such assumptions. However, we mitigate the impact of this by drawing on the observed structure of the screening strategies obtained in Chapter 4 to design our binary encoding scheme. The main strengths of this approach are the following: it guarantees perfectly smooth strategies; and it is generally faster insofar as it avoids a regularizing stage.

The remainder of this chapter is organized as follows. In Section 5.2 we describe the binary encoding scheme used to encode perfectly regular screening strategies, and the method for finding near-optimal screening strategies. In Section 5.3 we present numerical results and sensitivity analysis from the optimization method. Finally in Section 5.4 we summarize the results and make concluding remarks about this chapter.
5.2 Methodology

This section is organized as follows. In Section 5.2.1 we describe the details of the encoding scheme. In Section 5.2.2 we provide examples of screening strategies encoded by the binary encoding scheme. In Section 5.2.3 we formalize the optimization problem over the encoded screening strategies. Finally, in Section 5.2.4 we describe the simulation optimization method used to find near-optimal binary-encoded screening strategies.

5.2.1 Encoding Scheme

The encoding scheme is based on the defining characteristics mentioned in Section 5.1. By fixing the ages and PSA thresholds for the first and final screenings, we construct a line segment connecting the age-threshold pairs for the first and final screen events. We then use the screening frequency to determine when screenings will occur after the first screen event. The PSA thresholds for screenings subsequent to the first screen are taken from the PSA threshold value corresponding to the respective ages of the screenings on the constructed line segment.

We first describe how the values of each of these defining characteristics are encoded as functions of ordinary decimal (base-10) quantities, before going on to describe the further binary (base-2) encoding that is used. Finally, we provide an algorithm for decoding a screening strategy from its binary encoding according to this scheme. We use binary encoding primarily because of the ease with which it can be used to define different characteristics of a screening strategy by recourse to decision variables that share the same set of feasible values (0 and 1). Secondarily, we adopt binary encoding in this chapter as a means of ensuring that there is no pathological shortcoming associated with the nonbinary encoding used in Chapter 4.

Age at initial screening. We may reasonably assume that the age at initial screening is between 40 and 71 years for the following reasons. In the optimization model in Chapter 4, we assumed an earliest age at initial screening of 40 years. There are no published guidelines that recommend screening earlier than age 40. In fact, contemporary guidelines recommend initiating screening at age 55 [16]. We are not
aware of any published PSA screening guideline that suggests screening at ages greater than 70 years. Moreover, nothing in the results from Chapters 3 or 4 indicates an imperative to consider an age at initial screening less than 40 years or greater than 71 years. Therefore, these assumptions on the possible age at initial screening is not unreasonable. We let the age at initial screening be $40 + \xi^{(1)}$ years. Then, we constrain $\xi^{(1)}$ as follows: $\xi^{(1)} \in \{0, 1, \ldots, 31\}$. We chose these constraint endpoints not only so as not to exclude published recommendations, but also because of the range of decimal values that various binary substrings (i.e., “bit strings”) of fixed length can represent. This is also true for the constraint endpoints in the definitions of the other defining characteristics below.

**Age at final screening.** We may reasonably assume that the age at final screening is between 49 and 80 years. The author is not aware of any contemporary published PSA screening recommendation that suggests either an age at final screening less than 49 years or greater than 80 years; there are recommendations that include baseline PSA measurements as early as 40 years with the possibility of no subsequent screens [15], but our model is not presently engineered to consider such adaptive strategies. And nothing in the results from Chapters 3 or 4 indicates an imperative to consider an age at final screening less than 40 years or greater than 71 years. Therefore, this assumption on the possible age at final screening is not unreasonable. Let the age at final screening be $49 + \xi^{(2)}$ years. Then, we constrain $\xi^{(2)}$ as follows: $\xi^{(2)} \in \{0, 1, \ldots, 31\}$.

**Frequency of PSA screenings.** We may reasonably assume that the frequency of PSA screenings ranges from annually (every 1 years) to octennially (every 8 years). We let PSA screening occur every $\xi^{(3)} + 1$ years. And we constrain $\xi^{(3)}$ as follows: $\xi^{(3)} \in \{0, 1, \ldots, 7\}$.

**PSA threshold in ng/mL at the initial PSA screen.** We assume that the PSA threshold at the initial screen is inclusively between 0 and 10 ng/mL. This range of PSA thresholds is consistent with the range of PSA thresholds considered in Chapter 3 (see Table 4.2), and it encompasses the different PSA threshold levels from the representative strategies from the literature that are shown in Table 3.4. Let the PSA threshold at the initial screen be $\xi^{(4)}/3$ ng/mL. Then, we constrain $\xi^{(4)}$ as follows: $\xi^{(4)} \in \{0, 1, \ldots, 31\}$.
The valid PSA threshold value corresponding to \( \xi^{(4)} = 31 \) is greater than our assumed upper bound of 10 ng/mL. So we let \( \xi^{(4)} = 31 \) denote “no screening.”

**PSA threshold in ng/mL at the final PSA screen.** We may assume that the PSA threshold at the final screen is inclusively between 0 and 10 ng/mL. This is reasonable for the same reasons that this range is reasonable for the PSA threshold at the initial PSA screen. Let the PSA threshold at the final screen be \( \xi^{(5)} / 3 \) ng/mL. Then, we constrain \( \xi^{(5)} \) as follows: \( \xi^{(5)} \in \{0, 1, \ldots, 31\} \). The valid PSA threshold value corresponding to \( \xi^{(5)} = 31 \) is greater than our assumed upper bound of 10 ng/mL. So we let \( \xi^{(5)} = 31 \) denote “no screening.”

Given the 5 variables \( \xi^{(1)}, \xi^{(2)}, \xi^{(3)}, \xi^{(4)}, \) and \( \xi^{(5)} \) representing the defining characteristics, a screening strategy may be represented by the integer-valued vector

\[
\tilde{\xi} = \left( \xi^{(1)}, \xi^{(2)}, \xi^{(3)}, \xi^{(4)}, \xi^{(5)} \right)_{10}
\]

where \( (\cdot)_{10} \) and \( (\cdot)_{2} \) denote, respectively, the decimal (base-10) and binary (base-2) numerical representation operators, with corresponding ranges

\[
\xi^{(j)} \in [0, 31] \cap \mathbb{Z}, \text{ for } j \in \{1, \ldots, 5\} \setminus \{3\},
\]

\[
\xi^{(3)} \in [0, 7] \cap \mathbb{Z}.
\]

This scheme for defining a screening strategy necessarily produces perfectly regular screening strategies. To more easily integrate this scheme with a GA, we transform the integer-vector representation stated in Equation (5.1) to the following equivalent binary-vector representation:

\[
\tilde{\xi} = \left( \xi^{(1)}_{0} \ldots \xi^{(1)}_{4} | \xi^{(2)}_{0} \ldots \xi^{(2)}_{4} | \xi^{(3)}_{0} \ldots \xi^{(3)}_{2} | \xi^{(4)}_{0} \ldots \xi^{(4)}_{4} | \xi^{(5)}_{0} \ldots \xi^{(5)}_{4} \right)_{2}
\]

190
with corresponding ranges

\[ \xi^{(j)}_{i} \in \{0, 1\} \quad i = 0, \ldots, 4, \text{ for } j \in \{1, \ldots, 5\} \setminus \{3\}, \]

\[ \xi^{(3)}_{i} \in \{0, 1\} \quad i = 0, \ldots, 2. \]

That is, the correspondence between Equations (5.1) and (5.2) is based on the following component-wise relations:

\[
\begin{align*}
\left( \xi^{(j)} \right)_{10} &= \left( \xi^{(j)}_{0} \xi^{(j)}_{1} \xi^{(j)}_{2} \xi^{(j)}_{3} \xi^{(j)}_{4} \right)_{2}, \quad j \in \{1, \ldots, 5\} \setminus \{3\}, \\
\left( \xi^{(3)} \right)_{10} &= \left( \xi^{(3)}_{0} \xi^{(3)}_{1} \xi^{(3)}_{2} \right)_{2}.
\end{align*}
\]

Algorithm 17 decodes a 23-bit binary string into a screening strategy according to the scheme just described.
Algorithm 17: DECODESCHEME(BITS)

Input: Binary vector BITS containing 23 binary elements in the following order:
\((\xi_0^{(1)} \ldots \xi_4^{(1)} | \xi_0^{(2)} \ldots \xi_4^{(2)} | \xi_0^{(3)} \ldots \xi_4^{(3)} | \xi_0^{(4)} \ldots \xi_4^{(4)} | \xi_0^{(5)} \ldots \xi_4^{(5)})\).

Output: PSA thresholds (ng/mL) in the vector THRESHOLD indexed by age.

Data: No screening occurs at any age \(t\) for which \(\text{THRESHOLD}[t] = 31/3\).

1 begin
2 \(\text{THRESHOLD}[t] \leftarrow 31/3\) for all ages \(t\) in the model;
3 \(\text{STARTAGE} \leftarrow 40 + (\xi_0^{(1)} \xi_1^{(1)} \xi_2^{(1)} \xi_3^{(1)} \xi_4^{(1)})_2\);
4 \(\text{STOPAGE} \leftarrow 49 + (\xi_0^{(2)} \xi_1^{(2)} \xi_2^{(2)} \xi_3^{(2)} \xi_4^{(2)})_2\);
5 \(\text{FREQSCREEN} \leftarrow 1 + (\xi_0^{(3)} \xi_1^{(3)} \xi_2^{(3)})_2\);
6 \(\text{STARTPSA} \leftarrow (1/3) \times (\xi_0^{(4)} \xi_1^{(4)} \xi_2^{(4)} \xi_3^{(4)} \xi_4^{(4)})_2\);
7 \(\text{STOPPSA} \leftarrow (1/3) \times (\xi_0^{(5)} \xi_1^{(5)} \xi_2^{(5)} \xi_3^{(5)} \xi_4^{(5)})_2\);
8 if \(\text{STOPAGE} = \text{STARTAGE}\) then
9 \(\text{SLOPEPSA} \leftarrow 0\);
10 end
11 else
12 \(\text{SLOPEPSA} \leftarrow (\text{STOPPSA} - \text{STARTPSA}) / (\text{STOPAGE} - \text{STARTAGE})\);
13 \(\text{t} \leftarrow \text{STARTAGE}\);
14 while \(\text{t} \leq \text{STOPAGE}\) do
15 if \((\text{t} - \text{STARTAGE}) \bmod \text{FREQSCREEN} = 0\) then
16 if \(\text{STARTPSA} + (\text{t} - \text{STARTAGE}) \times \text{SLOPEPSA} \leq 10\) then
17 \(\text{THRESHOLD}[t] \leftarrow \text{STARTPSA} + (\text{t} - \text{STARTAGE}) \times \text{SLOPEPSA}\);
18 end
19 end
20 \(\text{t} \leftarrow \text{t} + 1\);
21 end
22 return \(\text{THRESHOLD}\);
23 end

5.2.2 Examples of Encoded Screening Strategies

To further illustrate how screening strategies are encoded by the binary encoding scheme, we present 3 examples. Example 1 is the strategy of screening every 3 years, starting at age 45 with PSA threshold 1.0 ng/mL and stopping at age 61 with PSA threshold 5.7 ng/mL. Example 2 is the strategy of screening once at age 49 with PSA threshold 5.3 ng/mL. Example 3 is the no screening strategy; note that there are many
possible \( \tilde{\xi} \) that map to no screening. Although Example 3 is an odd way to express no screening, we present it to underscore that there are many configurations of \( \tilde{\xi} \) that the stochastic simulation-optimization method can encounter and that represent the strategy of no screening.

**Example 1**  Suppose we have the following binary-encoded solution:

\[
\tilde{\xi}_1 = (00101|01100|010|00011|10001|010)
\]

During the execution of DECODESCHEME(\( \tilde{\xi}_1 \)), we obtain the following values for the relevant defining characteristics:

- STARTAGE \( \leftarrow 40 + (00101)_2 = 45 \)
- STOPAGE \( \leftarrow 49 + (01100)_2 = 61 \)
- FREQSCREEN \( \leftarrow 1 + (010)_2 = 3 \)
- STARTPSA \( \leftarrow \frac{1}{3} \cdot (00011)_2 = 1.0 \)
- STOPPSA \( \leftarrow \frac{1}{3} \cdot (10001)_2 = 5 \frac{2}{3} \)

Figure 5.1 depicts the decoded PSA-threshold–based screening strategy returned by DECODESCHEME(\( \tilde{\xi}_1 \)).
Figure 5.1: PSA-threshold–based screening strategy decoded from $\bar{\xi}_1$. 
Example 2  Suppose we have the following binary-encoded solution:

$$\tilde{\xi}_2 = (01001|00101|110|10000|01101)$$

During the execution of \textsc{DecodeScheme}(\tilde{\xi}_2), we obtain the following values for the relevant defining characteristics:

- \text{STARTAGE} \leftarrow 40 + (01001)_2 = 49
- \text{STOPAGE} \leftarrow 49 + (00101)_2 = 54
- \text{FREQSCREEN} \leftarrow 1 + (110)_2 = 7
- \text{STARTPSA} \leftarrow \frac{1}{3}(10000)_2 = 5\frac{1}{3}
- \text{STOPPSA} \leftarrow \frac{1}{3}(01101)_2 = 4\frac{1}{3}

Figure 5.2 depicts the decoded PSA-threshold–based screening strategy returned by \textsc{DecodeScheme}(\tilde{\xi}_2).
Example 3  Suppose we have the following binary-encoded solution:

\[ \tilde{\xi}_3 = (11001|01111|000|10111|00100) \]

During the execution of \textsc{DecodeScheme}(\tilde{\xi}_3), we obtain the following values for the relevant defining characteristics:

- \textsc{StartAge} \leftarrow 40 + (11001)_2 = 65
- \textsc{StopAge} \leftarrow 49 + (01111)_2 = 64
- \textsc{FreqScreen} \leftarrow 1 + (000)_2 = 1
- \textsc{StartPSA} \leftarrow \frac{1}{3}(10111)_2 = 7.8
- \textsc{StopPSA} \leftarrow \frac{1}{3}(00100)_2 = 1.3

Since the age at final screen is before the age at initial screen, this encoding implies no screening.

5.2.3 Optimization Model

Having described the method of encoding screening strategies used in this chapter, we now formalize the corresponding optimization method, which utilizes the \( E[QG] \) estimator \( \mathcal{Q}(x) \) from Section 4.2.3. Let the function \( \mathcal{L} : \tilde{\xi} \mapsto x \) denote the operation of decoding a binary-encoded screening strategy according to Algorithm 17. We let \( \Xi \) denote the set of all 23-element binary vectors; that is, let

\[ \Xi = \left\{ \left[ v_i \right]_{i=1}^{23} : v_i \in \{0, 1\} \text{ for } i = 1, \ldots, 23 \right\}. \]

Note that every element of the set \( \Xi \) is a valid, feasible, binary-encoded strategy, although there are a large number that decode to no screening.

Using the binary encoding scheme, we seek to find the encoded screening strategy that maximizes expected QG. The underlying approach, which is used to estimate \( E[QG] \) for a given strategy, is the model presented in Section 4.2.2 through Section 4.2.4. Whereas in Equation (4.90), we defined the optimization
problem as
\[ x^* = \arg \max_{x \in \Theta} E[Q(x)]. \] (5.3)

here we define the optimization problem as
\[ \tilde{\xi}^* = \arg \max_{\tilde{\xi} \in \Xi} \mathbb{E} \left[ Q \left( \mathcal{L} \left( \tilde{\xi} \right) \right) \right]. \] (5.4)

By comparing Equations (5.3) and (5.4), it is evident that the fundamental simulation optimization problems in Chapters 4 and 5 are quite similar. Here, in Chapter 5, however, we do not need the regularizing objective-function approach from Section 4.2.6 because every strategy \( \mathcal{L} \left( \tilde{\xi} \right) \) is by construction perfectly regular.

### 5.2.4 Simulation Optimization Method

We used an adaptation of the GA and clean-up procedure presented in Section 4.2.7 to solve the simulation-optimization problem stated in Equation (5.4). Algorithm 18 describes the GA applied to the present simulation optimization problem. The function RECOMBINE in Algorithm 18 works exactly as in Section 4.2.7.
Algorithm 18: GeneticsAlgorithm2(X)

Input: Population, X, of binary-encoded strategies to be modified.
Output: Best strategy found by the algorithm.
Data: The number NUMGENERATIONS of desired generations of the GA. The number of patients, SAMPLESIZE, to simulate when evaluating any strategy. Function RECOMBINE that takes a population of strategies as its first argument, takes the index of the current GA generation at its second argument, and returns a recombined (modified) population.

1 begin
2 GENERATION ← 0;
3 while GENERATION ≤ NUMGENERATIONS do
4     for i = 0 to SIZE(Y) - 1 do
5         Evaluate Y[i] and record estimate, \( \tilde{Q} (Y[j]) \), from simulation;
6     end
7     BESTSTRATEGY ← any strategy such that \( \tilde{Q} (Y[i]) \) ≥ \( \tilde{Q} (Y[j]) \) \( \forall Y[j] \in Y \);
8     if GENERATION < NUMGENERATIONS then
9         Y ← RECOMBINE(Y, GENERATION);
10    end
11    GENERATION ← GENERATION + 1;
12 end
13 return BESTSTRATEGY;
14 end

After running the GA, we run a version of the clean-up procedure modified (slightly) for use on a binary-encoded strategy. Specifically, the only modification is that the input strategy is first decoded. This clean-up procedure is presented in Algorithm 19.
**Algorithm 19: CLEANUP2(STRATEGY)**

**Input:** Binary-encoded screening strategy STRATEGY.

**Data:** Let $m$ be the number of patients (sample size) of the simulations used to estimate the point and CI estimates of expected QG in this algorithm.

1. **begin**
2. Set $x \leftarrow \mathcal{L}(\text{STRATEGY});$
3. Set random-number stream seed $\mathbb{H} \leftarrow 1;$
4. Evaluate strategy $x$ using $m$ simulated patients with seed $\mathbb{H};$
5. Set $\text{LOWERBOUNDORIGINAL} \leftarrow \bar{Q}(x) - (2.24)S_{\bar{Q}(x)}/\sqrt{m} ;$
6. Set $\text{CONTINUE} \leftarrow \text{true};$
7. **while** $\text{CONTINUE} \leftarrow \text{true};$
8. **do**
9. Set $t^* \leftarrow -1;$
10. Set $QG^* \leftarrow +\infty;$
11. **for** $t = 1$ to $N$ **do**
12. **if** $x[t] \neq 21$ **then**
13. **if** $K_t(x) < QG^*$ **then**
14. Set $t^* \leftarrow t;$
15. Set $QG^* \leftarrow K_t(x);$**
16. **end**
17. **end**
18. **if** $t^* \neq -1$ **then**
19. Set $t_{\text{last}} \leftarrow x[t^*];$
20. Set $x[t^*] \leftarrow 21;$
21. Set $\mathbb{H} \leftarrow \mathbb{H} + 1;$
22. Evaluate strategy $x$ using $m$ simulated patients with seed $\mathbb{H};$
23. **if** $\bar{Q}(x) + (2.24)S_{\bar{Q}(x)}/\sqrt{m} < \text{LOWERBOUNDORIGINAL}$ **then**
24. Set $x[t^*] \leftarrow t_{\text{last}};$
25. Set $\text{CONTINUE} \leftarrow \text{false};$
26. **end**
27. **else**
28. Set $\text{CONTINUE} \leftarrow \text{false};$
29. **end**
30. **end**
31. **return** $x;$
32. **end**

### 5.3 Results

In Section 5.3.1, we describe the basecase parameter values and the GA settings used to generate basecase results for this section. We then present optimized strategies found by using the GA and clean-up procedure with the binary encoding scheme and compare them to strategies in Chapters 3 and 4.
5.3.1 Optimization Results for Basecase Parameter Values

The basecase estimates for the QALY disutilities and the other model parameters used to generate results in this section are the same as those used to generate results in Section 4.3.2. And, as in the basecase results in Chapter 4, the results in this section are based on a maximum of 3 biopsies per patient.

**GA settings.** For each optimized strategy, the GA was run for 50 generations with a constant population size of 30, a mutation rate (\texttt{MUTATION\_RATE}) of 0.01, and an elitism order (\texttt{ELITISM}) of 1. The initial population for the GA consisted of the 30 best of 500 randomly generated screening strategies, where each randomly generated strategy was constructed by uniformly randomly sampling values from \{0, 1\} for each decision variable. The GA was run using standard recombination, tournament selection, and two point crossover, which are described in Chapter 4 in Algorithms 5, 9, and 12, respectively.

**Random number seeds.** Instead of generating only 1 optimized strategy for a given risk-group and parameter-levels configuration, we generated samples of 5 optimized strategies. Consequently, for the basecase optimization results in this section, we executed 5 independent runs of the GA for each risk group. Each GA run was independent due to utilizing a different random number stream for the initial population construction and crossover and mutation operators. Specifically, at the beginning of a given run of the GA, the underlying random number generator was set to a given integer seed, and then the corresponding random number stream was used first to construct a random initial population and then to drive the crossover and mutation operators for the entirety of the GA run.

Using the basecase parameter values and the algorithm settings detailed above, we executed 5 independent runs of the binary-encoding optimization method for each of the compared risk groups. The resultant samples of optimized strategies are provided in Table 5.1. Instead of reporting the PSA thresholds at the initial and final screens, we report for each optimization run the average PSA threshold over the PSA threshold time-series constructed by decoding the corresponding optimized strategy.
Table 5.1: Samples of optimized strategies produced by the binary-encoding–scheme method for each of the compared risk groups using the basecase parameter estimates. Each QG estimate is based on a sample of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. PSA thresholds have units ng/mL, and ages and screening frequency have units years. The screening frequency means the time between screen events.

<table>
<thead>
<tr>
<th>Risk group</th>
<th>Seed</th>
<th>Age at initial screening</th>
<th>Age at final screening</th>
<th>Avg PSA threshold</th>
<th>Screening frequency</th>
<th>Estimated E[QG] (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whites</td>
<td>1</td>
<td>47</td>
<td>59</td>
<td>0.2</td>
<td>4</td>
<td>0.0882 ± 0.0011</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>53</td>
<td>58</td>
<td>0.3</td>
<td>1</td>
<td>0.0877 ± 0.0011</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>53</td>
<td>55</td>
<td>0.0</td>
<td>2</td>
<td>0.0877 ± 0.0010</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>51</td>
<td>58</td>
<td>0.3</td>
<td>1</td>
<td>0.0877 ± 0.0010</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>51</td>
<td>57</td>
<td>0.2</td>
<td>2</td>
<td>0.0879 ± 0.0011</td>
</tr>
<tr>
<td>African Americans</td>
<td>1</td>
<td>52</td>
<td>56</td>
<td>0.3</td>
<td>1</td>
<td>0.1728 ± 0.0015</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>52</td>
<td>56</td>
<td>0.3</td>
<td>1</td>
<td>0.1728 ± 0.0015</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>48</td>
<td>58</td>
<td>0.3</td>
<td>2</td>
<td>0.1704 ± 0.0015</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>53</td>
<td>55</td>
<td>0.1</td>
<td>1</td>
<td>0.1697 ± 0.0015</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>52</td>
<td>56</td>
<td>0.3</td>
<td>1</td>
<td>0.1728 ± 0.0015</td>
</tr>
<tr>
<td>Family History</td>
<td>1</td>
<td>47</td>
<td>59</td>
<td>0.2</td>
<td>4</td>
<td>0.1154 ± 0.0012</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>52</td>
<td>58</td>
<td>0.3</td>
<td>2</td>
<td>0.1136 ± 0.0012</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>52</td>
<td>56</td>
<td>0.1</td>
<td>2</td>
<td>0.1148 ± 0.0012</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>52</td>
<td>55</td>
<td>0.0</td>
<td>3</td>
<td>0.1137 ± 0.0012</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>52</td>
<td>58</td>
<td>0.3</td>
<td>1</td>
<td>0.1159 ± 0.0012</td>
</tr>
<tr>
<td>CCI=1</td>
<td>1</td>
<td>52</td>
<td>52</td>
<td>0.0</td>
<td>1</td>
<td>0.0332 ± 0.0007</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>52</td>
<td>53</td>
<td>0.0</td>
<td>1</td>
<td>0.0353 ± 0.0008</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>52</td>
<td>52</td>
<td>0.0</td>
<td>1</td>
<td>0.0332 ± 0.0007</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>52</td>
<td>52</td>
<td>0.0</td>
<td>1</td>
<td>0.0332 ± 0.0007</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>52</td>
<td>52</td>
<td>0.0</td>
<td>1</td>
<td>0.0332 ± 0.0007</td>
</tr>
<tr>
<td>CCI≥2</td>
<td>1</td>
<td>49</td>
<td>49</td>
<td>0.0</td>
<td>1</td>
<td>0.0148 ± 0.0005</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>48</td>
<td>48</td>
<td>0.0</td>
<td>1</td>
<td>0.0147 ± 0.0005</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>48</td>
<td>48</td>
<td>0.0</td>
<td>1</td>
<td>0.0147 ± 0.0005</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>48</td>
<td>48</td>
<td>0.0</td>
<td>1</td>
<td>0.0147 ± 0.0005</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>48</td>
<td>48</td>
<td>0.0</td>
<td>1</td>
<td>0.0147 ± 0.0005</td>
</tr>
</tbody>
</table>
The samples in Table 5.1 are described in terms of sample-mean point and CI estimators by risk group in Table 5.2.

Table 5.2: Sample statistics on the optimized strategies produced by the binary-encoding–scheme method for each of the compared risk groups using the basecase parameter estimates. The CIs are estimated using the Student-t distribution. PSA thresholds have units ng/mL, and ages and screening frequency have units years. The screening frequency means the time between screen events.

<table>
<thead>
<tr>
<th>Risk group</th>
<th>95% CI estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age at initial screening</td>
</tr>
<tr>
<td>Whites</td>
<td>51.0 ± 0.6</td>
</tr>
<tr>
<td>African Americans</td>
<td>51.4 ± 0.5</td>
</tr>
<tr>
<td>Family History</td>
<td>51.0 ± 0.5</td>
</tr>
<tr>
<td>CCI=1</td>
<td>52.0 ± 0.0</td>
</tr>
<tr>
<td>CCI≥2</td>
<td>48.2 ± 0.1</td>
</tr>
</tbody>
</table>

For clearer analysis, we constructed a representative strategy for each risk group derived from the descriptive statistics in Table 5.2. For each representative strategy, we assumed a constant PSA threshold and screening frequency. Each strategy was then constructed directly from the sample-mean point estimates of the following quantities from Table 5.2: the age at initial screening; the age at final screening; the average PSA threshold; and the frequency of screening. These representative strategies are described in Table 5.3.

Table 5.3: Means-based strategies constructed on the basis of the sample means from Table 5.2 of the following quantities: the age to start screening; the age to stop screening; the average PSA threshold, and the frequency of screening.

<table>
<thead>
<tr>
<th>Risk group</th>
<th>Means-based screening strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whites</td>
<td>PSA-screen from age 51 to 57 every 2 years at threshold 0.2 ng/mL</td>
</tr>
<tr>
<td>African Americans</td>
<td>PSA-screen from age 51 to 56 annually at threshold 0.3 ng/mL</td>
</tr>
<tr>
<td>Family History</td>
<td>PSA-screen from age 51 to 57 every 2 years at threshold 0.2 ng/mL</td>
</tr>
<tr>
<td>CCI=1</td>
<td>Biopsy at age 52</td>
</tr>
<tr>
<td>CCI≥2</td>
<td>Biopsy at age 48</td>
</tr>
</tbody>
</table>
Having constructed the representative strategies from the basecase optimization runs in this chapter, we compared them with the strategies from Chapter 3 and the basecase optimized strategies from Chapter 4 on the basis of estimated E[QG]. The results from this comparison are shown in Table 5.4.
Table 5.4: Expected QG estimates across risk groups using basecase parameter levels comparing strategies #1–#14 from Chapter 3 to the best strategies found by optimization in Chapters 4 (the strategies in Figures 4.15a through 4.15e), and the representative strategies from Chapter 5 (the strategies in Table 5.3). Each estimate is based on a sample of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. For each risk group, the labels “opt4” and “opt5” denote, respectively, the corresponding optimized strategy from Chapter 4 and representative strategy from Chapter 5.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Whites</th>
<th>African Americans</th>
<th>Family History</th>
<th>CCI=1</th>
<th>CCI≥2</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
<td>0.0000 ± 0.0000</td>
</tr>
<tr>
<td>#2</td>
<td>-0.0067 ± 0.0008</td>
<td>0.0099 ± 0.0011</td>
<td>-0.0043 ± 0.0008</td>
<td>-0.0269 ± 0.0006</td>
<td>-0.0346 ± 0.0005</td>
</tr>
<tr>
<td>#3</td>
<td>0.0029 ± 0.0008</td>
<td>0.0251 ± 0.0012</td>
<td>0.0073 ± 0.0009</td>
<td>-0.0236 ± 0.0007</td>
<td>-0.0350 ± 0.0006</td>
</tr>
<tr>
<td>#4</td>
<td>0.0043 ± 0.0008</td>
<td>0.0278 ± 0.0012</td>
<td>0.0095 ± 0.0009</td>
<td>-0.0221 ± 0.0007</td>
<td>-0.0338 ± 0.0006</td>
</tr>
<tr>
<td>#5</td>
<td>0.0024 ± 0.0008</td>
<td>0.0282 ± 0.0012</td>
<td>0.0075 ± 0.0009</td>
<td>-0.0259 ± 0.0007</td>
<td>-0.0382 ± 0.0006</td>
</tr>
<tr>
<td>#6</td>
<td>0.0112 ± 0.0009</td>
<td>0.0477 ± 0.0013</td>
<td>0.0187 ± 0.0010</td>
<td>-0.0229 ± 0.0007</td>
<td>-0.0384 ± 0.0006</td>
</tr>
<tr>
<td>#7</td>
<td>0.0025 ± 0.0008</td>
<td>0.0246 ± 0.0012</td>
<td>0.0069 ± 0.0009</td>
<td>-0.0240 ± 0.0007</td>
<td>-0.0354 ± 0.0006</td>
</tr>
<tr>
<td>#8</td>
<td>0.0020 ± 0.0008</td>
<td>0.0277 ± 0.0012</td>
<td>0.0071 ± 0.0009</td>
<td>-0.0263 ± 0.0007</td>
<td>-0.0386 ± 0.0006</td>
</tr>
<tr>
<td>#9</td>
<td>0.0174 ± 0.0008</td>
<td>0.0431 ± 0.0012</td>
<td>0.0242 ± 0.0009</td>
<td>-0.0143 ± 0.0007</td>
<td>-0.0297 ± 0.0006</td>
</tr>
<tr>
<td>#10</td>
<td>0.0079 ± 0.0008</td>
<td>0.0367 ± 0.0012</td>
<td>0.0138 ± 0.0009</td>
<td>-0.0240 ± 0.0007</td>
<td>-0.0383 ± 0.0006</td>
</tr>
<tr>
<td>#11</td>
<td>0.0138 ± 0.0007</td>
<td>0.0314 ± 0.0011</td>
<td>0.0190 ± 0.0008</td>
<td>-0.0120 ± 0.0006</td>
<td>-0.0244 ± 0.0005</td>
</tr>
<tr>
<td>#12</td>
<td>0.0088 ± 0.0004</td>
<td>0.0200 ± 0.0007</td>
<td>0.0115 ± 0.0005</td>
<td>0.0036 ± 0.0004</td>
<td>0.0006 ± 0.0003</td>
</tr>
<tr>
<td>#13</td>
<td>0.0073 ± 0.0005</td>
<td>0.0128 ± 0.0008</td>
<td>0.0095 ± 0.0005</td>
<td>-0.0049 ± 0.0004</td>
<td>-0.0109 ± 0.0004</td>
</tr>
<tr>
<td>#14</td>
<td>0.0002 ± 0.0005</td>
<td>0.0001 ± 0.0007</td>
<td>0.0010 ± 0.0005</td>
<td>-0.0150 ± 0.0004</td>
<td>-0.0215 ± 0.0003</td>
</tr>
<tr>
<td>opt4</td>
<td>0.0881 ± 0.0010</td>
<td>0.1426 ± 0.0014</td>
<td>0.1052 ± 0.0011</td>
<td>0.0349 ± 0.0008</td>
<td>0.0148 ± 0.0005</td>
</tr>
<tr>
<td>opt5</td>
<td>0.0867 ± 0.0011</td>
<td>0.1733 ± 0.0015</td>
<td>0.1135 ± 0.0012</td>
<td>0.0332 ± 0.0007</td>
<td>0.0146 ± 0.0005</td>
</tr>
</tbody>
</table>
The comparison in Table 5.4 is illustrated in Figure 5.3.

![Figure 5.3: Estimated expected QG for each of the compared risk groups for all the strategies from Chapter 3, the basecase optimized strategies from Chapter 4, and the means-based strategies derived from the optimized strategy statistics in Table 5.2. For each risk group, the labels “opt4” and “opt5” denote, respectively, the corresponding optimized strategy from Chapter 4 and representative strategy from Chapter 5. As mentioned in Section 4.3.2, it is more natural from a clinical perspective to compare screening strategies based on NNSQ or NNS than based on expected QG. So, we compare the NNSQ and NNS for each risk group at the corresponding representative strategy. And we compare these NNSQ and NNS estimates with the corresponding estimates for the basecase optimized strategies from Chapter 4. The comparisons between Chapters 4 and 5 on the basis of NNSQ and NNS are depicted, respectively, in Figures 5.4a and 5.4b.](image-url)
Figure 5.4: Comparison of the estimated expected NNSQ and NNS for the basecase optimized strategies in Chapters 4 and 5. The strategies from Chapter 4 are those depicted in Figures 4.15a through 4.15e. The strategies from Chapter 5 are the representative strategies described in Table 5.3. All NNSQ and NNS estimates are based on samples of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. The 95% CIs reflect inter-patient stochastic variability; they do not reflect generalized uncertainty in any of the model parameters.
Comparing the basecase results for the White risk group with those for the Family History risk group, there are no statistically significant differences in the characteristics of the optimized screening strategies from Chapters 4 and 5. Although there are statistically significant differences between the characteristics of the optimized strategies for the African American risk group and the White and Family History risk groups, it is not obvious that these differences are practically significant. For each of these three risk groups, the optimized strategies suggest very aggressive screening, i.e., screening that is likely to result in multiple biopsies, over a similar period of time.

Both of the comorbidity risk groups consist of a biopsy for all patients in that risk group at one point in time. The biopsy for men in the CCI\(\geq 2\) risk group occurs at an earlier age than the biopsy for men in the CCI=1 risk group. A higher comorbidity rating corresponds to an increased other-cause mortality rate, especially in later years. So, the diagnosis and treatment of PCa later in life that would be beneficial for the noncomorbidity risk groups may not be beneficial for the comorbidity risk groups since comorbidity patients are more likely to die before these late-age tumors become life threatening. In other words, a patient with comorbidities is less threatened by late-age PCa than are noncomorbidity patients. And as comorbidity patients are less threatened by late-age PCa than are noncomorbidity patients, so severe comorbidity (CCI\(\geq 2\)) patients are less threatened by late-age PCa than are moderate comorbidity (CCI=1) patients. For these reasons, it is reasonable that screening in CCI=1 males may not be indicated to continue as late in age as in males from the noncomorbidity risk groups, and it is reasonable that screening in CCI\(\geq 2\) males may be indicated to take place earlier than in CCI=1 males.

It is evident that the binary encoding method produces optimized strategies that are significantly better (in terms of QALYs and PCa mortality) than those from Chapter 3. And the performance of the binary encoding method appears to be roughly equivalent to the performance of the three-stage method from Chapter 4.

### 5.3.2 Sensitivity Analysis on Model Parameters

We performed one-way sensitivity analysis over parameters \(e_t\), leadtime, and \(D_\ast\) for each risk group. At each experimental level, we ran the binary-encoding optimization method 5 times, using a different
random-number seed each time to produce the initial population and to govern the GA operators. Tables 5.5 through 5.9 show the estimated expected QG for the optimized strategy produced for each risk group at every experimental configuration and random number seed. Whenever a value of $0.000 \pm 0.000$ is shown for the estimated expected QG, the corresponding screening strategy is the no-screening strategy.

Table 5.5: Expected QG estimates from one-way sensitivity analysis over model parameters for the Whites risk group, based on samples of 5 independent runs of the GA for each risk-group–experimental-configuration combination. Each estimate is based on a sample of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. The symbol $\mathcal{D}\_*$ denotes that all QALY disutilities ($\mathcal{D}_{\text{Scr}}, \mathcal{D}_{\text{Dia}}, \mathcal{D}_{\text{Biop}}, \mathcal{D}_{\text{Tre}},$ and $\mathcal{D}_{\text{Met}}$) were simultaneously varied to their greatest or smallest plausible values. Seed means the random number seed used in the construction of the initial population for the GA and in the crossover and mutation operators of the GA.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Seed</th>
<th>Low level</th>
<th>Basecase</th>
<th>High level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_t$</td>
<td>1</td>
<td>0.0000 $\pm$ 0.0000</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.3115 $\pm$ 0.0016</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0000 $\pm$ 0.0000</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.3126 $\pm$ 0.0016</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0000 $\pm$ 0.0000</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.3159 $\pm$ 0.0016</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0000 $\pm$ 0.0000</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.3126 $\pm$ 0.0016</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0000 $\pm$ 0.0000</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.3126 $\pm$ 0.0016</td>
</tr>
<tr>
<td>leadtime</td>
<td>1</td>
<td>0.0782 $\pm$ 0.0011</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.0978 $\pm$ 0.0011</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0788 $\pm$ 0.0010</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.0979 $\pm$ 0.0011</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0755 $\pm$ 0.0010</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.0964 $\pm$ 0.0011</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0755 $\pm$ 0.0010</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.0972 $\pm$ 0.0011</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0772 $\pm$ 0.0010</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.0980 $\pm$ 0.0011</td>
</tr>
<tr>
<td>$\mathcal{D}_*$</td>
<td>1</td>
<td>0.1323 $\pm$ 0.0011</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.0358 $\pm$ 0.0010</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.1333 $\pm$ 0.0011</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.0358 $\pm$ 0.0010</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.1351 $\pm$ 0.0011</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.0358 $\pm$ 0.0010</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.1331 $\pm$ 0.0010</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.0328 $\pm$ 0.0010</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.1346 $\pm$ 0.0011</td>
<td>0.0882 $\pm$ 0.0011</td>
<td>0.0358 $\pm$ 0.0010</td>
</tr>
</tbody>
</table>
Table 5.6: Expected QG estimates from one-way sensitivity analysis over model parameters for the African Americans risk group, based on samples of 5 independent runs of the GA for each risk-group–experimental-configuration combination. Each estimate is based on a sample of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}$, $D_{Dia}$, $D_{Biop}$, $D_{Tre}$, and $D_{Met}$) were simultaneously varied to their greatest or smallest plausible values. Seed means the random number seed used in the construction of the initial population for the GA and in the crossover and mutation operators of the GA.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Seed</th>
<th>Low level</th>
<th>Basecase</th>
<th>High level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_t$</td>
<td>1</td>
<td>0.0000 ± 0.0000</td>
<td>0.1728 ± 0.0015</td>
<td>0.5931 ± 0.0023</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0000 ± 0.0000</td>
<td>0.1728 ± 0.0015</td>
<td>0.5968 ± 0.0023</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0000 ± 0.0000</td>
<td>0.1728 ± 0.0015</td>
<td>0.5914 ± 0.0022</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0000 ± 0.0000</td>
<td>0.1728 ± 0.0015</td>
<td>0.5955 ± 0.0023</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0000 ± 0.0000</td>
<td>0.1728 ± 0.0015</td>
<td>0.5981 ± 0.0023</td>
</tr>
<tr>
<td>leadtime</td>
<td>1</td>
<td>0.1489 ± 0.0015</td>
<td>0.1728 ± 0.0015</td>
<td>0.1927 ± 0.0015</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.1503 ± 0.0015</td>
<td>0.1728 ± 0.0015</td>
<td>0.1907 ± 0.0016</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.1504 ± 0.0015</td>
<td>0.1728 ± 0.0015</td>
<td>0.1927 ± 0.0015</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.1510 ± 0.0015</td>
<td>0.1728 ± 0.0015</td>
<td>0.1907 ± 0.0016</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.1506 ± 0.0015</td>
<td>0.1728 ± 0.0015</td>
<td>0.1927 ± 0.0015</td>
</tr>
<tr>
<td>$D_*$</td>
<td>1</td>
<td>0.2578 ± 0.0015</td>
<td>0.1728 ± 0.0015</td>
<td>0.0974 ± 0.0014</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.2575 ± 0.0015</td>
<td>0.1728 ± 0.0015</td>
<td>0.1035 ± 0.0015</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.2575 ± 0.0015</td>
<td>0.1728 ± 0.0015</td>
<td>0.0974 ± 0.0014</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.2544 ± 0.0015</td>
<td>0.1728 ± 0.0015</td>
<td>0.1086 ± 0.0015</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.2578 ± 0.0015</td>
<td>0.1728 ± 0.0015</td>
<td>0.1070 ± 0.0015</td>
</tr>
</tbody>
</table>
Table 5.7: Expected QG estimates from one-way sensitivity analysis over model parameters for the Family History risk group, based on samples of 5 independent runs of the GA for each risk-group–experimental-configuration combination. Each estimate is based on a sample of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. The symbol $D_*$ denotes that all QALY disutilities ($D_{\text{Scr}}, D_{\text{Dia}}, D_{\text{Biop}}, D_{\text{Tre}},$ and $D_{\text{Met}}$) were simultaneously varied to their greatest or smallest plausible values. Seed means the random number seed used in the construction of the initial population for the GA and in the crossover and mutation operators of the GA.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Seed</th>
<th>Low level</th>
<th>Basecase</th>
<th>High level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_t$</td>
<td>1</td>
<td>0.0000 ± 0.0000</td>
<td>0.1154 ± 0.0012</td>
<td>0.3978 ± 0.0018</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0000 ± 0.0000</td>
<td>0.1154 ± 0.0012</td>
<td>0.3910 ± 0.0018</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0000 ± 0.0000</td>
<td>0.1154 ± 0.0012</td>
<td>0.3970 ± 0.0018</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0000 ± 0.0000</td>
<td>0.1154 ± 0.0012</td>
<td>0.4012 ± 0.0018</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0000 ± 0.0000</td>
<td>0.1154 ± 0.0012</td>
<td>0.3971 ± 0.0018</td>
</tr>
<tr>
<td>Leadtime</td>
<td>1</td>
<td>0.1018 ± 0.0012</td>
<td>0.1154 ± 0.0012</td>
<td>0.1284 ± 0.0012</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.1024 ± 0.0011</td>
<td>0.1154 ± 0.0012</td>
<td>0.1286 ± 0.0012</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.1014 ± 0.0011</td>
<td>0.1154 ± 0.0012</td>
<td>0.1261 ± 0.0012</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.1014 ± 0.0011</td>
<td>0.1154 ± 0.0012</td>
<td>0.1261 ± 0.0012</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.1021 ± 0.0011</td>
<td>0.1154 ± 0.0012</td>
<td>0.1261 ± 0.0012</td>
</tr>
<tr>
<td>$D_*$</td>
<td>1</td>
<td>0.1694 ± 0.0012</td>
<td>0.1154 ± 0.0012</td>
<td>0.0577 ± 0.0011</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.1694 ± 0.0012</td>
<td>0.1154 ± 0.0012</td>
<td>0.0568 ± 0.0011</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.1707 ± 0.0012</td>
<td>0.1154 ± 0.0012</td>
<td>0.0577 ± 0.0011</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.1690 ± 0.0012</td>
<td>0.1154 ± 0.0012</td>
<td>0.0560 ± 0.0011</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.1694 ± 0.0012</td>
<td>0.1154 ± 0.0012</td>
<td>0.0594 ± 0.0011</td>
</tr>
</tbody>
</table>
Table 5.8: Expected QG estimates from one-way sensitivity analysis over model parameters for the CCI=1 risk group, based on samples of 5 independent runs of the GA for each risk-group–experimental-configuration combination. Each estimate is based on a sample of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. The symbol $D_*$ denotes that all QALY disutilities ($D_{\text{Scr}}, D_{\text{Dia}}, D_{\text{Biop}}, D_{\text{Tre}},$ and $D_{\text{Met}}$) were simultaneously varied to their greatest or smallest plausible values. Seed means the random number seed used in the construction of the initial population for the GA and in the crossover and mutation operators of the GA.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Seed</th>
<th>Low level</th>
<th>Basecase</th>
<th>High level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_t$</td>
<td>1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0332</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0332</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0332</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0332</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0332</td>
</tr>
<tr>
<td>leadtime</td>
<td>1</td>
<td>0.0284</td>
<td>0.0008</td>
<td>0.0332</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0287</td>
<td>0.0007</td>
<td>0.0332</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0287</td>
<td>0.0007</td>
<td>0.0332</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0296</td>
<td>0.0008</td>
<td>0.0332</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0287</td>
<td>0.0007</td>
<td>0.0332</td>
</tr>
<tr>
<td>$D_*$</td>
<td>1</td>
<td>0.0748</td>
<td>0.0008</td>
<td>0.0332</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0710</td>
<td>0.0008</td>
<td>0.0332</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0754</td>
<td>0.0008</td>
<td>0.0332</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0752</td>
<td>0.0008</td>
<td>0.0332</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0748</td>
<td>0.0008</td>
<td>0.0332</td>
</tr>
</tbody>
</table>

95% CI estimates for expected QG
Table 5.9: Expected QG estimates from one-way sensitivity analysis over model parameters for the CCI≥2 risk group, based on samples of 5 independent runs of the GA for each risk-group–experimental-configuration combination. Each estimate is based on a sample of $10^7$ simulated patients from the corresponding risk group under the corresponding screening strategy. The symbol $D_*$ denotes that all QALY disutilities ($D_{Scr}, D_{Dia}, D_{Biop}, D_{Tre},$ and $D_{Met}$) were simultaneously varied to their greatest or smallest plausible values. Seed means the random number seed used in the construction of the initial population for the GA and in the crossover and mutation operators of the GA.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Seed</th>
<th>Low level</th>
<th>Basecase</th>
<th>High level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_t$</td>
<td>1</td>
<td>0.0000 ± 0.0000</td>
<td>0.0148 ± 0.0005</td>
<td>0.1029 ± 0.0010</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0000 ± 0.0000</td>
<td>0.0148 ± 0.0005</td>
<td>0.1022 ± 0.0009</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0000 ± 0.0000</td>
<td>0.0148 ± 0.0005</td>
<td>0.1022 ± 0.0009</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0000 ± 0.0000</td>
<td>0.0148 ± 0.0005</td>
<td>0.1022 ± 0.0009</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0000 ± 0.0000</td>
<td>0.0148 ± 0.0005</td>
<td>0.1037 ± 0.0010</td>
</tr>
<tr>
<td>leadtime</td>
<td>1</td>
<td>0.0128 ± 0.0005</td>
<td>0.0148 ± 0.0005</td>
<td>0.1016 ± 0.0005</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0128 ± 0.0005</td>
<td>0.0148 ± 0.0005</td>
<td>0.1016 ± 0.0005</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0128 ± 0.0005</td>
<td>0.0148 ± 0.0005</td>
<td>0.1016 ± 0.0005</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0121 ± 0.0005</td>
<td>0.0148 ± 0.0005</td>
<td>0.1016 ± 0.0005</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0128 ± 0.0005</td>
<td>0.0148 ± 0.0005</td>
<td>0.1016 ± 0.0005</td>
</tr>
<tr>
<td>$D_*$</td>
<td>1</td>
<td>0.0477 ± 0.0006</td>
<td>0.0148 ± 0.0005</td>
<td>0.0024 ± 0.0003</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0469 ± 0.0006</td>
<td>0.0148 ± 0.0005</td>
<td>0.0024 ± 0.0003</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0473 ± 0.0006</td>
<td>0.0148 ± 0.0005</td>
<td>0.0024 ± 0.0003</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0444 ± 0.0007</td>
<td>0.0148 ± 0.0005</td>
<td>0.0024 ± 0.0003</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0469 ± 0.0006</td>
<td>0.0148 ± 0.0005</td>
<td>0.0024 ± 0.0003</td>
</tr>
</tbody>
</table>
From the sensitivity analysis, we note that for each risk group, at the low level of $e_t$ (the probability of metastasis of an untreated PCa) the best decision is simply not to screen. Also, considering the magnitude of the changes in estimated expected QG when $e_t$ is raised from the basecase level to the high level in each risk group, we see that among the considered parameters $e_t$ has the greatest effect on expected quality-of-life outcome. Since parameter $e_t$ governs the rate of PCa growth, it is not surprising that this parameter is highly impactful. Low values of $e_t$ make PCa more slow-growing and thus less urgent to treat, whereas high values of $e_t$ make PCa more fast-growing and thus more urgent to treat.

As in the SA from Chapter 4, we found there to be more screen events in the optimized strategy when all QALY disutilities were set to their most favorable (i.e., lowest) levels, which is intuitive. The optimized strategies produced at the lower and higher levels of leadtime did not appreciably differ, which supports the inference that leadtime does not greatly affect the characteristics of near-optimal screening, though it may greatly affect the magnitude of the expected benefit of screening.

We also conducted sensitivity analysis on the PSA sampling method. In particular, we varied the distribution mean $\mu_{\beta_2}$ of the normal random variable $\beta_2 \sim \mathcal{N}(\mu_{\beta_2}, \sigma_{\beta_2}^2)$ in Equation (3.10). The random variable $\beta_2$ is the postonset slope increment for the linear changepoint PSA sampling model. Once PCa onset occurs, then the randomly-sampled value $\beta_{2,j}$ determines for the $j$th patient the rate at which PSA growth continues. So, a larger value of $\beta_{2,j}$ means that the $j$th patient’s PSA levels will increase more rapidly over time after PCa onset. That is, larger values of $\beta_2$ correspond to PSA trajectories that more noticeably increase following PCa onset, and thus larger values of $\beta_2$ should tend to make PSA testing potentially more discriminating (and, therefore, potentially more useful). Conversely, smaller values of $\beta_2$ correspond to PSA trajectories that are less noticeably affected by PCa onset, which makes PSA testing less likely to be beneficial. Gulati et al. [40] provide 95% CI estimates of $\mu_{\beta_2}$: the lower and upper CI limits of $\mu_{\beta_2}$ are, respectively, 0.0919 and 0.1269.

At each of the values $\mu_{\beta_2} = 0.0919$ and $\mu_{\beta_2} = 0.1269$, we ran 5 independent runs of the binary-encoding method with all model parameters (except $\mu_{\beta_2}$) set to their basecase levels. We ran these experiments for each of the compared risk groups. Then, we compared by risk group the sample estimates of the following quantities obtained when $\mu_{\beta_2}$ is set to the lower and upper values: (i) the average age
to begin screening; (ii) the average age to stop screening; (iii) the average PSA threshold; and (iv) the average screening frequency. At the $\alpha = 0.05$ level, for no risk group was there a statistically significant difference in any of these four quantities at the lower and upper limits of $\mu_{\beta_2}$.

As in Chapter 4, we conducted additional SA experiments over the following: (i) the preclinical dwelling time; (ii) the QALY disutility of biopsy; (iii) the per-patient maximum number of biopsies; and (iv) the PSA sampling method. We set the disutility of biopsy to 0.05 QALYs, we set the per-patient maximum number of biopsies to 1, and we replaced the PSA sampling method of Gulati et al. [40] with the method in [107]. With these settings, we ran 5 independent runs of the binary encoding method at each of the following levels of the preclinical dwelling time: 0.0, 0.4, 0.8, and 12.0 years. The optimized strategies produced by these experiments are described in Table 5.10. We find that as the preclinical dwelling time decreased, the length of time between the age at initial screen and the age at final screen increased. Taking the sample mean of the difference between the age at final screen and the age at initial screen across random-number seeds, we have the estimates 8.4, 10.8, 7.6, and 1.0 years for the screening-interval length for the preclinical dwelling times of 0.0, 0.4, 0.8, and 12.0 years, respectively.
Table 5.10: Samples of optimized strategies produced by the binary-encoding method for the reference risk group at preclinical dwelling time ($t_{PD}$) levels of 0, 4, 8, and 12 years. PSA thresholds have units ng/mL, and ages and screening frequency have units years. The screening frequency means the time between screen events.

<table>
<thead>
<tr>
<th>Preclinical Dwelling Time</th>
<th>Seed</th>
<th>Age at initial screening</th>
<th>Age at final screening</th>
<th>Avg PSA threshold</th>
<th>Screening frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 years</td>
<td>1</td>
<td>52</td>
<td>60</td>
<td>0.7</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>51</td>
<td>63</td>
<td>1.5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>55</td>
<td>55</td>
<td>0.0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>55</td>
<td>65</td>
<td>0.7</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>52</td>
<td>64</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>4 years</td>
<td>1</td>
<td>53</td>
<td>65</td>
<td>0.8</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>53</td>
<td>61</td>
<td>1.1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>55</td>
<td>65</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>52</td>
<td>63</td>
<td>1.2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>50</td>
<td>63</td>
<td>1.2</td>
<td>1</td>
</tr>
<tr>
<td>8 years</td>
<td>1</td>
<td>56</td>
<td>56</td>
<td>0.0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>56</td>
<td>62</td>
<td>0.6</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>55</td>
<td>65</td>
<td>0.7</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>56</td>
<td>66</td>
<td>0.4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>50</td>
<td>62</td>
<td>0.8</td>
<td>3</td>
</tr>
<tr>
<td>12 years</td>
<td>1</td>
<td>56</td>
<td>56</td>
<td>0.0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>56</td>
<td>56</td>
<td>0.0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>54</td>
<td>54</td>
<td>0.0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>54</td>
<td>56</td>
<td>0.7</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>54</td>
<td>57</td>
<td>0.5</td>
<td>3</td>
</tr>
</tbody>
</table>
5.4 Discussion

The most important findings for screening in this chapter are confirmations of findings already discussed in Section 4.4. First, all evidence produced in the basecase analysis in this chapter suggests that men in the White, African American, and Family History risk groups should be screened very aggressively, and that men in the CCI=1 and CCI≥2 risk groups should be biopsied once. Given how low the PSA thresholds are in the optimized strategies for the White, African American, and Family History risk groups, these strategies have the effect of causing virtually all patients (who live long enough) to undergo a biopsy. Therefore, the inference to biopsying virtually all men is supported by the results in this chapter. Note that many men are likely to have 3 biopsies according to the noncomorbidity risk-group strategies in Table 5.2. This is not all that different from the results in Chapter 4. Not every patient is guaranteed to have 3 biopsies. And the loss of an additional 0.00577 QALYs may not be a high price to pay for the increased effective sensitivity achieved by having an additional biopsy.

Second, the results in this chapter strongly suggest that men with comorbidities should be biopsied earlier in life than men without comorbidities, and that the more severe the comorbidities, the earlier the biopsy should take place.

Finally, since the optimized strategies for each risk group after each independent run of the binary-encoding optimization method at the low level of parameter $e_t$ was no screening, it is critical to have accurate estimates of this parameter. All of these screening-related findings agree with the findings from Chapter 4.

By the head-to-head comparison shown in Table 5.4, it is evident that the constraint imposed upon the PSA-threshold time series by virtue of the binary encoding scheme does not result in a statistically significant loss in estimated expected QG for the optimized strategy. Since the approach in this chapter does not require a separate regularization stage to produce regular strategies, this approach may be preferable.
Conclusions and Recommendations

6.1 Main Conclusions of the Research

In this dissertation, we developed and validated a new PSA-based simulation model of screening for PCa that consists of the following major components: (i) a PCa natural-history model, which is based on a discrete-event stochastic process representing a patient’s progression through key health states; and (ii) a method for simulating patient PSA levels over time. The model incorporates a method for using PCa leadtime to account for clinical incidence of PCa, which is important for estimating accurately the benefit of screening. This screening simulation model was extended to represent screening in specific well-known PCa risk groups by careful modification of relevant model parameters. After studying risk-based screening on the basis of PCa-specific mortality, through the NNS statistic, we developed a QALYs parameterization of the screening model, which enabled us to develop the NNSQ statistic and use a simple transformation of NNSQ, namely QG, as the performance measure in an optimization problem.

After developing the optimization problem, we developed and motivated a regularization approach to mitigate against the tendency of heuristic simulation-optimization methods to produce clinically unrealistic screening strategies. We developed a three-stage simulation-optimization approach based on a GA with a clean-up procedure and an implementation of the COMPASS algorithm, where the COMPASS procedure in the third stage attempts to make regularizing improvements to the best strategy produced by the GA in the first stage. Using this approach, we present numerical results on near-optimal screening in
each of the compared risk groups, including sensitivity analysis over key model parameters.

As an alternative way of avoiding irregular screening strategies, we also presented a binary-encoding
metaheuristic that guarantees by construction perfectly regular screening strategies. Using this alternative
method, we present numerical results and sensitivity analysis for each of the compared risk groups, and
we compared these results with results produced by the three-stage method. The following is a summary
of some of the most important findings from Chapters 3, 4, and 5.

In Chapter 3, we developed the risk-based PCa-screening simulation model and used synchronized
patient histories (an implementation of the common random numbers variance-reduction technique) to
compute the NNS statistic. We compared a set of screening strategies from the literature on the basis
of estimated expected NNS, and we presented a range of clinical statistics on each strategy by risk
group that may prove insightful to clinicians and that would be very costly to estimate by RCTs. We
found that, on the basis of expected NNS, the best strategy from the set considered for each risk group
was the strategy of screening annually from age 50 to 75 using the PSA threshold 2.5 ng/mL. Thus,
strictly considering PCa-specific mortality, a single “one-size-fits-all” screening strategy may be optimal.
However, we also found that this strategy yielded significantly different benefit for different risk groups.
The African American risk group appears to benefit most from screening, and the comorbidity risk groups
appear to benefit least.

In Chapter 4, we developed and used a regularization approach that could be adapted to other
simulation optimization problems where there is benefit from regularity in the decision-variable time-
series values that is not captured by the original objective function. We developed and derived a CI
estimator for NNSQ, which is a statistic similar to NNS that could be used in a great variety of disease-
screening problems. And while NNS, which is widely-known, only accounts for mortality reduction,
NNSQ accounts for mortality reduction and a range of desirable and undesirable aspects of the disease
in question and the process of screening for and treating it. We found using estimated expected QG,
which is a simple transformation of NNSQ, that for every risk group the best screening strategy called
for biopsying virtually all men. We found that patients with moderate to severe comorbidities should be
screened earlier than patients without comorbidities. Similar to the observed distribution of benefit from
near-optimal screening in Chapter 3; and we found that considering QALYs, patients with comorbidities stand to gain significantly less benefit from screening than do patients from the other risk groups. In addition to the three-stage method, we also developed an analytical formulation of a simplified single-period PCa screening problem. With this analytical approach, we found independent support for the conclusion derived from the heuristic approach that biopsying virtually all men maximizes expected QALYs.

In Chapter 5, we developed a binary-encoding method that guarantees perfectly regular screening strategies and we used it with a GA to solve essentially the same problem that was posed in Chapter 4. We found that the near-optimal screening strategies produced by this alternative method were, as in Chapter 4, very aggressive strategies that result in biopsying virtually all patients. And we found that patients with moderate to severe comorbidities should be biopsied substantially earlier in life than patients without comorbidities or with comorbidities amounting to a score of CCI=1 (less severe).

In this research we concluded that PSA was not a reliable tool either for diagnosing the presence of PCa or for indicating the need to perform a follow-up biopsy. This finding is undoubtedly the most potentially controversial aspect of the research. It is, however, noteworthy that currently the USPSTF, the AUA, the EAU, the ACS, and the ACP all recommend against routine PSA-based screening for PCa (see Section 2.1.1). Note that the 2015 NCCN guidelines utilize a 1 ng/mL PSA threshold during early detection evaluation to determine the frequency of subsequent screening; thus, there is some precedent for using very low PSA thresholds to drive the PCa-screening decision-making process. Insofar as these recommendations are based on warranted distrust in PSA as a discriminator for PCa, they are consistent with our findings. The real departure from the prevailing medical opinion comes from our research’s finding that biopsying virtually all men maximizes expected quality-adjusted survival. But here it is important to note that there have not been large clinical studies designed to investigate the effect of such an aggressive strategy. That is, to our knowledge there is insufficient clinical data available to evaluate expected patient outcomes for the strategy of biopsying all men independent of PSA. The question of whether to routinely PSA-screen may be satisfactorily answered in the minds of many by the current set of trial studies on PCa screening, but the question of whether to mass-biopsy really has not been addressed.
6.2 Limitations and Future Research

In comparing our results with existing guidelines, it is important to note that there may be risk aversion associated with the decision to biopsy that is not reflected in the objective of maximizing expected QALYs. With respect to such potential unrepresented risk factors, our model assumes a risk neutral decision maker. Thus, from a societal perspective, biopsying all men may be ideal, but from the perspective of some individual patients it may not be. Our focus on maximizing expected QALYs is, therefore, an important limitation of this research.

Another important limitation of our model is that we do not consider alternative definitive-treatment modalities or active-surveillance protocols. Our model also assumes perfect adherence by every patient to the prevailing screening strategy. In these ways, our PCa screening simulation does not represent the entire range of clinical tools and patient behaviors. Another important limitation is the uncertainties inherent in the estimation of QALYs. Estimates of QALYs are subjective, and different patients have different degrees of preferences and dislikes.

There are also limitations associated with the risk group modeling in this research. Results for the nonreference risk groups should be considered with appropriate caution because most of the model parameter estimates for the nonreference risk groups were obtained by multiplicative adjustments of the corresponding estimates for the reference risk group. Also, because of the relative paucity of trial data on key PCa metrics for the nonreference risk groups, the nonreference risk-group models could not be validated as fully as the reference risk-group model.

Finally, because the major optimization methods used in this research are heuristic random-search based methods, we cannot assert that the optimized strategies we presented were optimal. We can, however, demonstrate that the methods produce significant improvement over the starting (initial) strategies and that the methods arrive at near-optimal solutions when solving difficult test problems with known optima.

There are several avenues for future extensions of this research that promise interesting findings. It would be interesting to develop simulations of other more narrowly-defined risk groups, such as African-American men with and without a family history of PCa or males with both a family history of PCa and comorbidities. On the one hand, the more fine-grained the risk groups, the more potentially
useful information could be produced for clinical application. On the other hand, the more fine-grained the risk groups, the more challenging it becomes to find accurate model-parameter estimates to distinguish the groups. As individualized risk-based medical research continues, it may become more feasible to construct fine-grained risk group representations. In the modern era in which nearly all PCa screening guidelines recommend shared decision-making that considers individual patient risk factors, there may be substantial benefit to this direction for future research.

It would also be helpful to examine the extent to which the sensitivity and specificity of PSA hinders clinicians from realizing improved outcomes in patients with PCa. This could be considered by a comparative analysis using new alternative PCa biomarkers or possibly even by some manipulation of the parameters of the PSA sampling method used in this research. Using empirical data, the clinical detection component of the model could be simulated with higher fidelity by randomly sampling patient PCa leadtimes—for example, using empirical distribution estimates from Savage et al. [90]—instead of treating every patient as having the same fixed value of leadtime.

In conclusion, we have produced valuable insights into the problem of risk-based PCa screening in particular and multiple generalizable approaches for dealing with irregularity in optimized strategies for disease-screening problems in general. The insights from this research may assist medical researchers and healthcare policy-makers in studying and dealing with the PCa problem in an individualized way in the future. The methods we have developed may help other researchers using simulation optimization to solve disease screening and treatment problems in a way that translates into actionable medical intelligence.
REFERENCES


[34] R. Etzioni, N. Howlader, P. A. Shaw, D. P. Ankerst, D. F. Penson, P. J Goodman, and I. M.


