Accepted Manuscript

Operations research models and methods in the screening, detection, and treatment of prostate cancer: A categorized, annotated review

Stuart Price, Bruce Golden, Edward Wasil, Brian Denton

PII: S2211-6923(15)00028-4
DOI: http://dx.doi.org/10.1016/j.orhc.2015.09.005
Reference: ORHC 88

To appear in: *Operations Research for Health Care*

Please cite this article as: S. Price, B. Golden, E. Wasil, B. Denton, Operations research models and methods in the screening, detection, and treatment of prostate cancer: A categorized, annotated review, *Operations Research for Health Care* (2015), http://dx.doi.org/10.1016/j.orhc.2015.09.005

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Operations Research Models and Methods in the Screening, Detection, and Treatment of Prostate Cancer: A Categorized, Annotated Review

Stuart Price
Robert H. Smith School of Business
University of Maryland
College Park, Maryland 20742

Bruce Golden
Robert H. Smith School of Business
University of Maryland
College Park, Maryland 20742

Edward Wasil
Kogod School of Business
American University
Washington, DC 20016

Brian Denton
Department of Industrial and Operations Engineering
Department of Urology
University of Michigan
Ann Arbor, Michigan 48109

Abstract

According to data from the Surveillance, Epidemiology, and End Results (SEER) program in the United States, approximately 15% of men will be diagnosed with prostate cancer during their lifetimes. Over the past 15 years, the battle against prostate cancer has been joined by researchers and practitioners who have used a wide variety of operations research (OR) models and methods to investigate decisions involving screening, detection, and treatment of prostate cancer. We provide a narrative review of articles falling into the following five categories: decision analysis, machine learning, optimization, simulation, and statistics. We identified a total of 523 archival journal articles describing the use of methods in these categories in the context of prostate cancer

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

since 2000. We categorize and annotate 49 of these articles in order to provide representative examples of the use of OR models and methods in each of these areas. We conclude with a summary of the trends in research using OR methods in the context of prostate cancer over the past 15 years, and a discussion about how these trends will influence future research.

June 26, 2015

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

1. Introduction

According to the Surveillance, Epidemiology, and End Results (SEER) program in the United States [61], based on 2008-2010 data, approximately 15% of men will be diagnosed with prostate cancer during their lifetime. In 2014, about 233,000 men in the United States were diagnosed with prostate cancer, making it the most common form of cancer among men. In 2011, more than 2.7 million men in the United States were estimated to be living with prostate cancer. Prostate cancer is the second leading cause of cancer deaths among men, with more than 29,000 deaths projected in the United States for 2014.

The high prevalence of prostate cancer, the large population at risk of prostate cancer, and recent advances in medical technology make prostate cancer care an important endeavor with potential to increase lifespan and quality of life for the large proportion of men who are affected. Prostate cancer is also expensive. At present, about $11.85 billion is spent annually in the United States on the screening, detection, and treatment of prostate cancer [42]. Thus, improving efficiency of decision making in the context of screening, detection, and treatment of prostate cancer has the potential to reduce cost to the health systems, allowing for more effective allocation of resources.

Over the years, the battle against prostate cancer has been joined by researchers and practitioners who have used a wide variety of operations research (OR) models and methods. A search on the Web of Knowledge using “prostate/prostatic and operations research method” revealed 523 published articles over the past 15 years. The citation count by year for these 523 articles is given in Figure 1. The increasing trend in citations suggest an increasing trend in the use and dissemination of knowledge about prostate cancer based on OR methods and models.

ACCEPTED MANUSCRIPT

In order to investigate the importance and impact of OR in prostate cancer research, we performed a literature search in which we categorized archival journal articles into five categories: decision analysis, machine learning, optimization, simulation, and statistics. These categories are not intended to provide an exhaustive enumeration of all OR approaches used for prostate cancer research. Rather, they were selected to provide examples of the use of a number of different types of OR models and methods including descriptive, predictive, and prescriptive approaches to decision making. We intentionally defined broad topic areas to include important approaches that are at the periphery of OR methods, including artificial intelligence, machine learning, and statistics. We annotate a representative set of 49 articles (selected from the 523 published articles) that illustrate the use of these approaches to deal with prostate specific antigen screening (quantifying the effects of screening, predicting prostate cancer, screening policy), detecting tumors (improving biopsies, using magnetic resonance imaging to detect prostate cancer), and treatment (staging, active surveillance, hormone therapy and tumor size, brachytherapy, patient choice, recurrence, costs). We take the 523 published articles and count how many articles fall into 13 OR modeling/method categories. Our count is given in Table 1A in the Appendix. Furthermore, we provide the OR model or method used by the 49 annotated papers in Table 1B in the Appendix.

In Sections 2, 3, and 4, we discuss OR models and methods for screening, detection, and treatment, respectively. Annotations are given in reverse chronological order, starting with the most recent article. We provide conclusions in Section 5.

ACCEPTED MANUSCRIPT

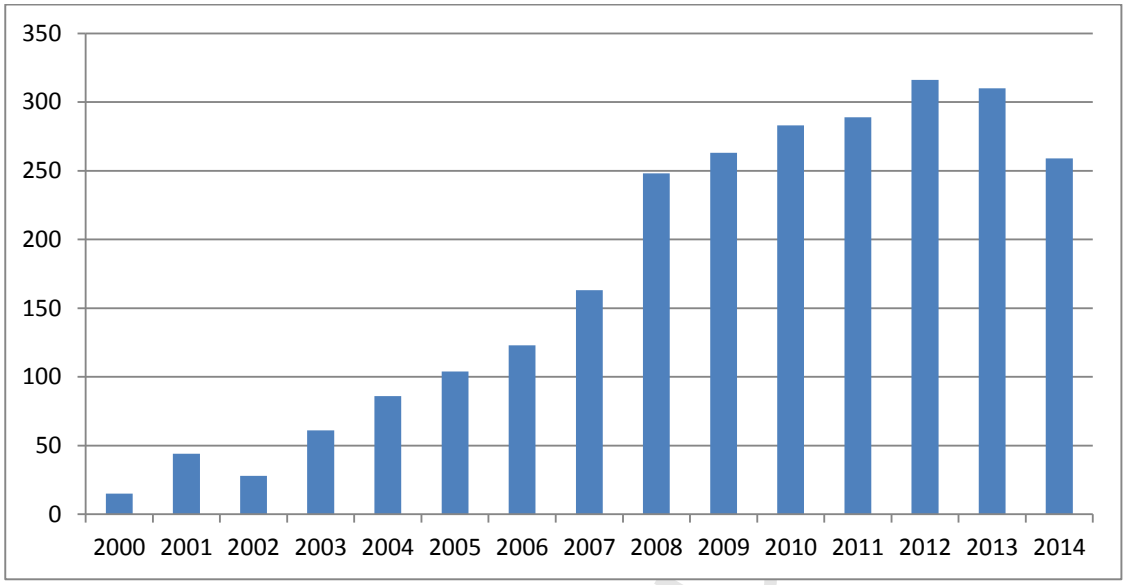

Figure 1. Number of citations by year for 523 articles using OR methods and models on prostate cancer research from the Web of Knowledge search (May 24, 2015).

2. Prostate Specific Antigen Screening

Since the FDA approved the prostate specific antigen (PSA) test for prostate cancer in 1994, millions of men have been tested. In the 2010 census, there were over 67 million American men over the age of 40 [31] whose screening decisions might be affected by screening recommendations. In 1980, prior to PSA screening, there were 106.0 new cases of prostate cancer per 100,000 men [61]. By 2000, after PSA screening was introduced, the rate jumped to 183.1, an increase of 72% [61]. Screening accounts for 89% of new prostate cancer diagnoses, with screen detected tumors being more clinically localized than clinically detected tumors (tumors detected due to the onset of other symptoms) [30]. In 2009, Medicare spent an estimated $447 million on PSA screening [41]. However, the benefits of screening are unclear. In 2012, the U.S. Preventive Service Task Force [45] recommended against PSA screening in men of average risk due to its limited benefit and potential harm from unnecessary treatment. There are three important

ACCEPTED MANUSCRIPT

questions that the field of operations research has helped address with respect to prostate cancer screening. What are the effects of prostate cancer screening? How can the use of the PSA test in detecting prostate cancer be improved? What should be the policy for PSA screening?

2.1 Quantifying the Effects of PSA Screening

The benefit of PSA screening is the early diagnosis of prostate cancer. The time between when the cancer is diagnosed by screening and when it would have presented clinically in the absence of screening is the lead time. The lead time is important because early treatment of the cancer can help improve outcomes.

Typically, prostate cancer is slow growing with a low mortality rate. Many men diagnosed with prostate cancer due to screening may never have been affected by prostate cancer within their lifetimes if the cancer had gone undetected. Men who are diagnosed due to screening and would not have presented the clinical symptoms within their lifetimes are referred to as having been *overdiagnosed*. Overdiagnosis can lead to overtreatment, i.e., treatment of a likely indolent cancer, resulting in some men unnecessarily living with the side effects of surgery or radiation therapy, with potential negative effects such as urinary and bowel dysfunction, erectile dysfunction, and loss of fertility. An important question is: Do the benefits from early diagnosis outweigh the risks of overdiagnosis?

There have been two major clinical trials focused on the effect of PSA screening on mortality rates from prostate cancer. The European Randomized Study of Screening for Prostate Cancer [53] found a 20% decline in mortality due to prostate cancer as a result of screening. The U.S. Prostate, Lung, Colorectal, and Ovarian Cancer Screening

ACCEPTED MANUSCRIPT

Trial [3] found that screening had no statistically significant effect on reducing mortality. These conflicting results have contributed to the controversy over the benefits and harms of prostate cancer screening. The contradictory results have been attributed to a number of causes of bias, such as members of the control groups (who should not be screened) being screened.

The failure of clinical trials to provide definitive answers about the merit of PSA screening provides motivation for the use of OR models to answer these questions. We summarize six articles that used statistics, Markov modeling, and simulation to estimate the effects of PSA screening on mortality, quality of life, and lead time.

Gulati et al. [27] used a simulated population to build a nomogram for estimating the probability that a screen-detected prostate cancer would not have been diagnosed within the patient's lifetime. A nomogram is a diagram that is used to calculate the probability of an event from a predictive model (such as the probability that a screen-detected prostate cancer would not have been diagnosed within a patient's lifetime). Overdiagnosis rates in the United States are estimated to be between 23% and 42% for screen-detected prostate cancer. This high rate of overdiagnosis leads to overtreatment which can be detrimental to a patient's health and happiness. Overdiagnosis is not directly observable. After an individual has screen-detected prostate cancer and has been treated, we do not know if the cancer would have been diagnosed within the individual's lifetime without screening. Using a microsimulation model of the progression of the disease, a population of 10,000 simulated prostate cancers was developed to train a logistic regression model indicating if a patient was overdiagnosed. Patient age, cancer Gleason score, and PSA level were used to predict if the patient had been overdiagnosed. The probability of overdiagnosis

ACCEPTED MANUSCRIPT

increased with age, decreased with PSA level, and decreased if the Gleason score was greater than or equal to seven. The nomogram had an area under the ROC curve (AUC) of 0.75.

Estimates of lead time and overdiagnosis due to PSA screening varies widely in the literature. Draisma et al. [14] used three independently developed models to compare predicted results for lead time and overdiagnosis of prostate cancer. The lead time can be interpreted in three ways: lead time for non-overdiagnosed cancers only, censored lead times for both non-overdiagnosed and overdiagnosed cancers with lead time for overdiagnosed patients stopping at death, and uncensored lead times (for overdiagnosed patients, these continue until the patients would have been clinically diagnosed). The authors used two microsimulation models from the literature and a statistical mixed model [64] to compare the estimates of lead times and overdiagnosis. A microsimulation model simulates the health outcomes for each patient of the population. The overdiagnosis estimates varied from 23% to 42% of all screen-detected cancers, which is consistent with population-based trial estimates. The non-overdiagnosed lead times ranged from 5.4 to 6.9 years, which allows for treatment to begin at an earlier stage, thereby improving outcomes. The censored lead time ranged from 5.7 to 7.8 years (this included the time that patients who would die from other causes are diagnosed with prostate cancer). The uncensored lead time ranged from 7.2 to 10 years (this included the time that patients who die from other causes would gain before diagnosis without PSA levels).

Etzioni et al. [18] used the microsimulation model to measure the effect of PSA screening on advanced stage prostate cancer and prostate cancer mortality. In 1986, PSA

ACCEPTED MANUSCRIPT

screening started in the United States. From 1990 to 1999, there was a 21% decline in observed mortality from prostate cancer. Their model compared outcomes with and without PSA screening (with some percentage of men receiving biopsies based on PSA level). PSA screening accounted for 80% of the observed drop in the incidence of advanced prostate cancer. PSA screening only accounted for two-thirds of the 21% drop in mortality from prostate cancer.

Schröder and Kattan [52] analyzed the results of 36 nomograms and artificial neural networks (ANNs) from the literature to compare their efficacy against using PSA level alone. The authors reviewed 23 articles that used three or more variables for risk assessment in the form of a nomogram or ANN. Model accuracy was compared using the AUC when available; otherwise sensitivity and specificity were used. The studies drew from different populations, with some including referred patients and those participating in screening studies. The populations ranged in size from 151 to 8,851 patients. The models considered a variety of input variables, depending on available information. Age, family history, and PSA velocity often lacked significance and were not included in the final models. Model validation on external populations is an important step in arguing generalizability. Ten of the 36 models were externally validated on a total of 16 external populations. In 13 external validations, the AUC decreased, in two it increased, and in one it stayed the same. In general, the models improved the AUC over PSA alone by approximately 0.10. The authors raised important questions about the independence of the input variables, some of which, like age, prostate volume, and PSA level, have relationships that are well understood.

ACCEPTED MANUSCRIPT

Tsodikov, Szabo, and Wegelin [64] constructed a statistical model to estimate lead time, overdiagnosis, and other relevant characteristics of prostate cancer screening. The authors used a three-stage model of the natural history of the disease (disease-free stage, pre-clinical stage, clinical stage). Overdiagnosis was defined as the fraction of screening detected cancers that would not be detected in the absence of screening. Lead time was defined only for non-overdiagnosed patients. The study was limited to men over 50, since the probability of prostate cancer is small for men under 50. Data on 350,000 cases of prostate cancer were taken from the SEER database and population counts for the relevant areas of diagnosis. To simulate PSA testing schedules to be used by the statistical model, a simulator for PSA schedules from the National Cancer Institute's Statistical Research and Applications Branch based on data from the National Health Interview Survey and SEER was used. The statistical model estimated a 6-year mean lead time and 25% overdiagnosis among detected patients.

Etzioni et al. [19] applied a microsimulation of PSA screening and prostate cancer development to estimate the rate of overdiagnosis in a hypothetical cohort of men. A hypothetical cohort based on census data of two million men between 60 and 84 years old in 1988 was used. The simulation was run on the group with and without screening. Testing and detection rates for the model varied by year, age, and race. The model used three, five, or seven years mean lead time to estimate how long before the cancer would have been clinically detected without screening. Overdiagnosis was estimated at 28.8% for white men and 43.8% for black men using five-year and seven-year lead time estimates, respectively. The overdiagnosis rate was sensitive to the mean lead time provided by screening, but not to the relative frequency of screen-detection. When the

ACCEPTED MANUSCRIPT

lead time was reduced to three years, the overdiagnosis rate dropped to 17.7% for white men and 20.3% for black men.

2.2 Predicting Prostate Cancer from the PSA Level

When the Food and Drug Administration (FDA) in the United States approved the use of PSA level to test for prostate cancer in 1994, a threshold of 4 ng/mL was used as the upper limit for a normal PSA level [22]. Using a PSA threshold of >4ng/mL to indicate a need for a biopsy, about 20% of prostate cancers would be detected and 30-40% of patients without prostate cancer would be above this threshold [25]. In 2009, the American Urological Association (AUA) recommended against using a single threshold level, and advocated the use of additional information to improve detection [25]. For example, PSA level changes with age.

Improving the predictive value of PSA screening means fewer unnecessary biopsies which carry the potential for infection, temporary erectile dysfunction, and lingering urinary problems. In the eight articles that follow, researchers applied machine learning and statistical methods to improve the prediction of prostate cancer by increasing sensitivity (probability of correctly identifying prostate cancer) and, in some cases, identifying only clinically significant prostate cancer (having a Gleason score greater than six).

Prostate biopsies can have one of five outcomes: benign, atypical small acinar proliferation (ASAP), high-grade prostatic intraepithelial neoplasia (PIN), non-significant prostate cancer, or clinically significant prostate cancer. Benign means that the biopsy was non-cancerous. ASAP means the biopsy result was indeterminate and could not be classified as benign or malignant. PIN is thought to be a precursor of a malignant tumor.

ACCEPTED MANUSCRIPT

Non-significant prostate cancer is considered not life threatening, while clinically significant prostate cancer typically requires treatment. Lawrentschuk et al. [36] constructed a polychotomous logistic regression model and an artificial neural network model to predict biopsy results. Age, PSA level, digital rectal examination (DRE), presence of hypoechoic lesion (a visible abnormality) during transrectal ultrasound (TRUS), and TRUS prostate volume for 3,025 men with PSA level less than 10 ng/mL who underwent a TRUS-guided biopsy were used to construct the models. The models were trained on two-thirds of the data and tested on the remaining one-third. The regression and neural network models correctly identified benign tumors 86% and 88% of the time, respectively and clinically significant prostate cancer 65% and 66% of the time, respectively. Neither model correctly identified any ASAP/PIN outcome. Both models predicted only 2% of non-significant prostate cancer correctly. Prostate volume and a positive TRUS lesion were the most significant in correctly identifying patients with clinically significant prostate cancer from benign biopsy results with odds ratios of 0.19 and 5.2, respectively. The authors concluded that additional predictors would be necessary to correctly distinguish among the five outcomes.

Many predictive models for prostate cancer biopsy results are developed using a population of patients referred for early cancer detection or urinary tract symptoms. Therefore, these models may not generalize to all men being screened for prostate cancer. Sooriakumaran et al. [59] used a population of screened patients to construct two predictive models in order to reduce unnecessary biopsies. The authors began with 3,838 men from the Tyrol Prostate Cancer Screening Study. They removed patients missing total PSA (tPSA) level, DRE, prostate volume, or percent free PSA (fPSA), leaving 2,271

ACCEPTED MANUSCRIPT

patients. The authors used a urologically referred population of 599 patients, from Weill Cornell Medical College, for external validation. Multivariate logistic regression models were constructed where Model 1 used age, DRE, and tPSA, and Model 2 added percent fPSA. Nonlinear relationships between the features and the outcome were evaluated using a multiple fractional polynomial method. A bootstrap method with 200 bootstrap samples was used during internal validation. The AUCs for the two models were 0.691 and 0.710, respectively. The authors then constructed nomograms based on the two models. If patients with a 10% or greater risk of cancer were biopsied, then 30 biopsies (1.3%) would be avoided and two cancers (0.3%) would be missed by Model 1 applied to the Tyrol sample. Model 2 avoided 95 biopsies (4.2%) and missed six cancers (0.9%). When applied to the urologically referred population for patients with 20% or greater risk of cancer, Model 1 avoided 126 biopsies (21.0%) and missed 18 cancers (9.8%). Model 2 avoided 169 biopsies (28.2%) and missed 19 cancers (10.3%). The authors concluded that their models had sufficient predictive power to aid in clinical decision making regarding the need to biopsy.

Gülkesen et al. [28] developed a decision tree to classify the risk level of prostate cancer for 750 patients with serum fPSA and PSA levels less than or equal to 10 ng/mL who underwent prostate biopsy at the Urology Department of Akdeniz University Hospital. Age, PSA level, free PSA level, percent-free PSA, DRE 1 (with three possible outcomes for prostate cancer: negative, suspicious, or positive), and DRE 2 (with two possible outcomes for prostate cancer: negative or not negative) were used to construct the model. The QUEST algorithm (quick, unbiased, and efficient statistical tree) produced a model using a training set of 562 patients. The algorithm classified the

ACCEPTED MANUSCRIPT

patients into five groups with the risk of cancer ranging from 0% to 25%. The AUC was 0.62 compared to 0.68 for logistic regression using the same features. The lowest risk group, with no prostate cancer in either the training or testing sets, identified by the decision tree had patients with a PSA level less than or equal to 5.98, a negative DRE, and an fPSA level greater than .81.

Nam et al. [46] constructed a nomogram to assess the risk of a patient having prostate cancer with a Gleason score greater than or equal to seven. In a data set of 3,010 men with PSA levels greater than 4.0 ng/mL or abnormal DREs referred to the Prostate Centers of the University of Toronto, 2,700 had PSA levels less than 50 ng/mL, were able to provide complete information including family history of prostate cancer, and consented to participate in the study. An additional 408 men with a PSA level less than or equal to 4.0 ng/mL agreed to undergo a biopsy and were added to the study. Age, ethnicity, family history, symptom score, PSA level, ratio of free PSA to total PSA, and DRE were used as predictor variables. A model to predict the probability of no cancer, low-grade cancer (Gleason score less than or equal to six), and high-grade cancer (Gleason score greater than or equal to seven) was constructed using ordinal logistic regression. The logistic regression model used data from 2,108 patients and was tested on 1,000 patients. A nomogram was designed using the results of the logistic regression model. The AUC for predicting any prostate cancer was 0.74 based on the nomogram. The total AUC for predicting high-grade cancer was 0.77 based on the nomogram. The logistic regression model with the full set of predictor variables performed much better than the model based only on PSA levels and DRE results, which had an AUC of 0.59.

ACCEPTED MANUSCRIPT

Aggressive prostate cancer, defined as having a Gleason score greater than or equal to seven, benefits from early detection and treatment. Spurgeon et al. [60] developed a classification and regression tree (CART) model to predict the presence of aggressive prostate cancer in patients. Age, PSA level, PSAD, DRE, race, family history, vasectomy, transrectal ultrasound findings, and prostate volume for 1,563 consecutively referred men from the Portland Veterans Administration Hospital were used in the training and testing of the CART model. The authors used 1,067 patients for training and 492 patients for testing. Their model identified 43 of the 47 cancer patients with aggressive prostate cancer (91.5% sensitivity) and 149 of the 445 patients with Gleason score less than seven (31.9% specificity). When their model predicted a Gleason score greater than or equal to seven (i.e., aggressive prostate cancer), it was correct only 12.7% of the time. When their model predicted a Gleason score below seven, it was correct 97.4% of the time.

Thompson et al. [62] used a logistic regression model based on data from the Prostate Cancer Prevention Trial to create an online risk calculator based on data for 5,519 men in the control group of the trial. The statistically significant model risk factors were race, age, family history, PSA level, DRE result, and whether the patient had a prior biopsy or not. The AUC for out of sample data, based on 4-fold cross validation, was approximately 70%, indicating satisfactory discrimination. A more recent revision of the model in 2012, referred to as PCPTRC 2.0, adds the ability to estimate the likelihood of low- and high-grade cancers [4]. Statistics about the risk calculator website indicate hundreds of thousands of visits in the last 48 months, suggesting the calculator is commonly used by physicians and/or patients.

ACCEPTED MANUSCRIPT

Garzotto et al. [24] developed a decision tree using CART to improve the accuracy of PSA tests. The authors trained their model using 1,433 consecutive referred patients who underwent an initial prostate biopsy procedure. They used 5-fold cross validation with the cost of misclassifying cancer as normal three times more than the cost of classifying a normal prostate as potentially cancerous. They mirrored the medical decision process used by physicians by first creating a decision tree using only PSA and DRE data. Then they added demographic data and family history to identify patients at high risk of prostate cancer. Their model identified 278 of the 291 cancer patients (95.5% sensitivity) and 334 of the 882 noncancerous patients (37.9% specificity). If their model predicted a cancerous biopsy, it was correct 33.7% of the time. If their model predicted a negative biopsy, it was correct 96.3% of the time. Furthermore, of the 15 cancer patients missed by the CART model, 14 had Gleason scores less than or equal to six.

2.3 PSA Screening Policy

Researchers have tried to improve screening policies for men in the United States by analyzing when to start screening, how often to screen, and at what PSA level to receive a biopsy. In 2009, the AUA recommended that screening start as early as age 40 [25]. In 2013, the AUA updated their guidelines and recommended against screening men at average risk of prostate cancer between the ages of 40 and 54 [7]. In the six articles summarized, researchers used optimization and simulation to compare the effects of different screening policies on mortality, quality-adjusted life years (QALYs), overtreatment, and lead time.

Gulati, Gore, and Etzioni [26] used a microsimulation model of prostate cancer to compare 35 screening policies based on lives saved, overdiagnoses, and mean time of life saved. A microsimulation model simulated at the unit of individual patients. This model

ACCEPTED MANUSCRIPT

simulated the PSA growth and disease progression of 100 million men. Thirty-two screening strategies that used starting age (40 or 50 years), stop age (69 or 74 years), screening intervals (annual or biennial), and biopsy referral conditions (PSA level of 4 ng/mL, PSA level of 2.4 ng/mL, PSA level of 4 ng/mL or PSA velocity of 0.35 ng/mL per year, or PSA level greater than 95th percentile for age) were compared. The authors evaluated screening policies recommended by the American Cancer Society (ACS), the National Comprehensive Cancer Network (NCCN), and the medical literature. They compared results to a base plan of annual screening from age 50 to 74 with biopsy threshold set at 4.0 ng/mL. The NCCN plan had the greatest improvement on reducing cancer deaths and increased the mean time of life saved from 0.86 month to 1 month. However, the NCCN plan doubled the number of false negatives and increased the probability of overdiagnosis from 3.3% to 6%. The ACS plan performed the same as the base strategy for lives saved and overdiagnoses, but reduced the number of tests conducted by nearly a quarter. Using age-specific PSA thresholds for biopsy referral reduced false positives by 25%, overdiagnoses by 30%, and lives saved by 17%.

Underwood et al. [65] constructed a simulation, based on a non-stationary, finite horizon Markov process, to model various PSA-based screening policies. Each screening policy was represented as a set of PSA thresholds for biopsy referral by age. They optimized these policies using a genetic algorithm and compared their results to policies presented in the literature. The simulation used 51,294 PSA levels from 11,872 men from Olmsted County, Minnesota and age-specific death and incidence data from a variety of other sources. The authors judged the policies in terms of quality-adjusted life years (QALYs), giving a one-time penalty for unnecessary biopsies. The genetic algorithm

ACCEPTED MANUSCRIPT

converged on a plan that had less than annual screening between the ages of 54 and 76, with the threshold for a biopsy based on a PSA level set lower than the current standard. The optimized policy slightly improved the results of the best policy given in the literature, and indicated the maximum number of QALYs that screening policies could provide. Underwood et al. suggested that screening over a shorter time period, with lower PSA thresholds, might do the most to improve QALYs.

Zhang et al. [67] constructed a partially observable Markov decision process to examine the benefits of PSA screening to patients and society. Patients progress through health states (no cancer and cancer are not directly observable; treated, metastasis, and death are considered known) and observable PSA intervals. Each year a decision is made to perform a biopsy or defer the decision until the next time period. The objective was to maximize QALYs for a patient. Detriments to the quality of life were occurrence of biopsy, treatment upon detection of cancer, long-term complications resulting from treatment, and symptoms from metastasis and its treatment. A secondary objective, from the societal perspective, is to maximize the expected monetary value of the QALYs minus the costs of screening, biopsies, and treatments. Prostate cancer probabilities conditional on PSA level were estimated using 11,872 patients from Olmsted County, Minnesota. A finite fixed-grid method was used to obtain the optimal policies for patients and society. The optimal policy from a patient's perspective (maximize QALYs) had screening stop at age 76 and improved QALYs by 0.131 over no screening and 0.165 over current policy (note that current policy underperforms no screening). The optimal policy from a societal viewpoint stops screening at age 71 and improved QALYs by 0.110 over no screening and 0.161 over traditional guidelines.

ACCEPTED MANUSCRIPT

Zhang et al. [68] used the partially observable Markov decision process model to examine prostate biopsy referral decisions. The optimal policy often took the form of a control-limit type policy, that is, a biopsy is performed only when the belief state for cancer exceeds a threshold value. There were three conditions for a control-limit type policy: annual probability of prostate cancer incidence is less than or equal to a multiple of the probability of treated patients developing metastasis, the annual probability of death from other causes is non-decreasing, and the annual probability of death from metastatic prostate cancer is non-increasing. There exist conditions when it is optimal to discontinue biopsy referral at a specific age. The optimal policy was estimated to improve the AUA's age-adjusted guidelines by 0.115 QALYs (0.306%). The authors noted that personalized utility assessment, consideration of comorbidity, and family history of prostate cancer might be added to the model in the future.

Prostate cancer screening strategies that use a PSA score vary with the age that a patient begins screening, the threshold result from a PSA test to recommend a biopsy, and the time intervals to receive screening. It is impractical to run randomized screening trials for all possible screening strategies. Ross et al. [51] constructed a Monte-Carlo simulation that tested a range of screening policies based on the number of prostate cancer deaths prevented. They tracked the number of PSA tests and biopsies per 1,000 men taken over the entire run of the simulation. The simulation was based on a Markov model that transitioned between no prostate cancer, three levels of cancer, and mortality due to prostate cancer and other causes. They simulated populations of one million men starting at 40 years of age for 40 years. Eight policies (including a baseline of no screening) with different PSA thresholds, screening frequency, and age recommendations

ACCEPTED MANUSCRIPT

were tested. The medical standard at the time called for annual PSA testing from age 50 to 75 with a PSA score above 4 ng/mL requiring a biopsy. The authors constructed a policy of screening at ages 40, 45, and 50, and then every two years until 75 with the PSA level at 4 ng/mL. Their new plan reduced the lifetime biopsies per 1,000 men from 600 to 450 and the average number of PSA tests in a lifetime from 10.5 to 7.5. Prevented deaths increased to 3.3 per 1,000 men from 3.2.

Etzioni, Cha, and Cowen [17] constructed a computer model to measure the effect of PSA screening strategies on years of life saved, PSA tests performed, false-positive PSA tests, and rate of overdiagnosis. The authors constructed a simulation model of prostate cancer and its progression in the population using real-world data from the SEER registry. A Markov model simulated the disease progression of five-year birth cohorts. The mortality rates for all causes of death other than prostate cancer and deaths due to prostate cancer were computed for each cohort. Clinical histories were developed for the patients that included the date of clinical diagnosis of prostate cancer (in the absence of screening) and the date of death not due to prostate cancer. Not all men develop prostate cancer. If the clinical diagnosis of prostate cancer was made before death from other causes, a history of disease progression was needed. Using the simulated progression of prostate cancer, natural histories were constructed for patients that included dates for disease onset and progression of the cancer. PSA levels that resulted from screening were simulated and grew at different rates based on the health state of the prostate. Five screening strategies were tested: annual PSA screening with age-specific biopsy thresholds, biannual PSA screening with age-specific biopsy thresholds, annual PSA screening with >4ng/mL threshold, biannual screening with >4ng/mL threshold, and

ACCEPTED MANUSCRIPT

screening every five years with threshold of >4ng/mL. Biannual screening results in about half the total number of screenings and number of false-positive screenings, and captures 95% of the cases that would have been caught with annual screenings. Using age specific PSA levels reduces false positives by 13.7%, but reduces the percentage of patients that are clinically detected from 63% to 54.8%.

2.4 Screening with New Biomarker

The prior literature on prostate cancer screening has focused on the principal methods of screening which include clinical exams (digital rectal examination) and PSA testing. No other biomarkers have been approved for prostate screening as of yet. However, numerous biomarkers have been discovered and are in late state clinical investigation. Future research will consider whether and when to use these biomarkers, and if and how they should be combined with standard PSA-based screening approaches.

3. Detecting Tumors

PSA screening is relatively inexpensive, costing between $17 and $62 [41]. However, it is not a definitive test, because its sensitivity is too low to justify treatment for a patient. A biopsy is necessary for definitive diagnosis of prostate cancer. Typically, a biopsy involves taking 12 samples of the prostate using hollow core needles. The results are reviewed by a pathologist to provide an assessment which includes the Gleason grade of the cancer and other clinical factors associated with patient health outcomes.

In some cases, additional diagnostic tests may be used to provide information about the size, grade, and/or location of a tumor. These may include magnetic resonance imaging (MRI), or the use of additional biomarkers. There are a number of new

ACCEPTED MANUSCRIPT

biomarkers that have been discovered, and some have recently received FDA approval, such as the urine based PROGENSA prostate cancer antigen 3 (PCA3) test. Similar to PSA, the PCA3 test provides a score that can be used to predict the presence and grade of cancer [21].

3.1 Improving Biopsies

Each year, approximately one million prostate biopsies are performed in the United States [66]. Studies have shown that the standard sextant transrectal ultrasound-guided needle biopsy has a false negative diagnosis rate over 20% [5,49]. Since the sensitivity of the standard biopsy is less than 80%, multiple biopsies may be needed to confirm prostate cancer. A biopsy is a painful, invasive procedure that has a risk of infection, with 4% of patients developing an infection. Within 30 days of a prostate biopsy, 6.9% of patients are admitted to the hospital [40]. To prevent the need for multiple biopsies, it is important to raise the accuracy of a biopsy. In the article summarized below, the authors tried to reduce the false negative rate by analyzing tumor locations and developing an optimal biopsy procedure.

Sofer, Zeng, and Mun [58] developed optimal biopsy procedures for fixed numbers of needles. The authors used 301 surgically removed prostates to generate a 3D statistical distribution of cancer occurrence in the prostate. The prostate was divided into 48 zones and 6,000 subzones. A biopsy protocol specified the number of samples to take and the zones where the samples are taken. There is variability as to the exact position and angle the physician aligns the needle for the biopsy. The authors optimized the biopsy protocols to maximize the probability of detecting cancer. Using their 3D model, they estimated the detection rate for the current sextant method at 67.3%. The optimized detection rate for a

ACCEPTED MANUSCRIPT

biopsy with six needles confined to the posterior of the prostate was 78.8%. By increasing the number of samples taken (to as high as 12), the authors showed that the detection rate could be as high as 85.5%. Biopsy procedures were developed for small and large prostates that had better detection rates than the current sextant procedure. Further work will examine the possibility of alternate protocols based on age, race, and PSA level.

3.2 Using an MRI to Detect Prostate Cancer

Using an MRI to detect prostate cancer has been shown to reduce the number of men needing a biopsy by 51%, lower the false negative rate, and identify low-risk tumors [63]. Each year, unnecessary biopsies are estimated to cost $2 billion. An MRI could help prevent unnecessary biopsies, thereby reducing this cost [1]. An MRI does not carry the same risks as a biopsy such as infection, bleeding at the biopsy site, and painful urination. Therefore, it seems reasonable to have an MRI and analyze the results before deciding to have a biopsy. In the following four articles, the authors analyzed the MRI data of patients who had a tumor removed (the procedure is known as a prostatectomy) and used machine learning to automate the detection of tumors.

Proton magnetic resonance spectroscopic imaging (MRSI) allows for the detection and quantification of biochemical markers within the prostate and has been shown to be an improvement over a conventional MRI. A 3D grid divides the prostate into voxels (a voxel is a volumetric element that serves as the unit of analysis). Each voxel has its own spectra data from the MRSI. Matulewicz et al. [43] constructed an ANN model to automatically detect cancerous voxels in the prostate using MRSI data. MRSI data from 18 patients with positive biopsies for prostate cancer who later underwent radical

ACCEPTED MANUSCRIPT

prostatectomies were used to construct the model. In all, 5,308 voxels were used. One hundred forty eight voxels were labeled suspicious by an experienced spectroscopist, and 129 of the 148 voxels were confirmed to have cancer following the radical prostatectomy. Only voxels identified as cancerous by both histopathological maps and the spectroscopist were labeled as cancerous. The ANN model had input, hidden, and output layers, each with a single node. Two types of models were constructed. The first model had 256 variables based on the spectra. The second model added the percentage of the voxel that was in the peripheral zone, transition zone, periurethral region, and outside the prostate. Seventy percent of the data was used as a training set, 15% as a test set to prevent overtraining, and 15% as a validation set to measure performance. Since there is a random element in the generation of the ANN model, 100 models of each of the two types were applied to the same training and test sets. The model of each type that performed best on the test set was selected, and their performances on the validation set were compared. The AUC was 0.949 for model 1 and 0.968 for model 2. The sensitivity on the validation set for model 1 was 50% (identifying 8 of 16 cancerous voxels) and 62.5% for model 2 (identifying 10 of 16 cancerous voxels). The specificity on the validation set for model 1 was 98.7% (identifying 770 of 780 noncancerous voxels) and 99.0% for model 2 (identifying 772 of 780 noncancerous voxels).

Anderson et al. [2] constructed logistic regression, nearest neighbor, and hybrid classifiers to predict the risk of prostate cancer using MRI data. Multi-parametric images from dynamic contrast MRI, diffusion weighted MRI, and MRSI were used to generate features (apparent diffusion coefficient, volume transfer constant, conventional average of T2 values, and spectroscopy scores) for 223 slices of prostates from 28 patients who

ACCEPTED MANUSCRIPT

had a radical prostatectomy (all of the prostate gland is removed). Gleason scores were assigned to each slice by an experienced radiologist. Leave-one-out cross-validation was used to separate the data into training and test sets. The models were trained to classify a slice as having a Gleason score between 0 and 4 or between 5 and 8. Logistic regression had an accuracy of 64.6% and an AUC of 0.66, and KNN had an accuracy of 74%. A hybrid approach was developed using linear regression with the number of cancerous neighbors from KNN. The hybrid approach had an accuracy of 77% and an AUC of 0.85. The hybrid approach identified the most aggressive cancers (Gleason score of 7 or 8) with an accuracy of 82% and an AUC of 0.86.

Parfait et al. [47] constructed a support vector machine (SVM) and a multilayer perceptron to classify magnetic resonance spectra (MRS) of prostates as healthy or pathological. MRS data from 22 patients with positive biopsies for prostate cancer were used. There were a total of 2,464 spectra, with 1,062 spectra localized in the peripheral zone where the majority of prostate cancer lesions are typically found. The peripheral zone spectra were manually classified as undetermined (286), healthy (636), or pathological (140). A variety of preprocessing algorithms could be applied including phase correction, baseline correction, and normalization. Nine combinations of preprocessing algorithms and classification method were compared. The best average performance using 5-fold cross validation was SVM with preprocessing using phase correction and baseline correction. This combination correctly classified 82.2% of undetermined, 94.3% of healthy, and 75.0% of pathological spectra. When restricted to the healthy and pathological classifications, 4.5% are misclassified with 83.6% sensitivity and 98.1% specificity.

ACCEPTED MANUSCRIPT

Kelm et al. [34] compared the use of quantitative and subspace feature extraction methods in linear and nonlinear machine learning classifiers for identifying prostate tumors with MRSI. A subspace method projects higher dimensional data onto a lower dimensional subspace. Two quantitative feature extraction methods from the MRS literature and a new quantitative method were compared to four subspace feature extraction methods – principal components analysis (PCA), partial least squares (PLS), independent component analysis, and nonnegative matrix factorization. These features were then used in either a linear classifier (logistic regression, generalized PLS, P-spline signal regression) or a nonlinear classifier (a random forest method that uses multiple decision trees, SVM, Gaussian processes). MRSI data from 24 patients were used to construct the models. Using published guidelines, 4,188 voxels were identified as healthy, undecided, or having a tumor. Cross-validation was performed by training on 23 patients and testing on the one remaining patient. None of the classifiers using quantification features outperformed a conventional ratio feature used by clinicians. Subspace methods improved the performance of the classifiers, particularly the nonlinear classifiers. There was little difference among the performances of the four subspace methods.

4. Prostate Cancer Treatment

There have been great advances in the treatment of prostate cancer in the past 30 years. In 1985, the five-year survival rate was 75%; in 2006, it was 99.6% [61]. This increase is due to better treatments and earlier detection due to PSA screening. Surgery (prostatectomy) has long been the primary form of treatment. Surgery carries a risk of

ACCEPTED MANUSCRIPT

pain, bleeding, infection, and death. There is also a chance of incontinence and erectile dysfunction.

Other treatment options include external beam radiation therapy, brachytherapy, and hormone therapy. Often treatments are used in tandem. For example, hormone therapy is used to shrink a tumor before radiation treatment. Because prostate cancer can be slow growing, a patient with a low risk cancer (e.g., Gleason score ≤ 6, PSA < 10) may forgo immediate treatment and elect to monitor the progression of the cancer (known as active surveillance) through regular clinical exams, PSA testing, and surveillance biopsies.

Operations research models have been used to improve the delivery of treatment, compare treatment options, and help patients make treatment selections.

4.1 Staging

Staging describes the classification of a cancer. This is important information when making a treatment decision. Prostate cancer can be slow growing, posing little risk to the life of some patients who have low risk based on clinical factors. Unnecessary treatment can result in worse health outcomes. Identifying a low-risk tumor can prevent unnecessary treatment, while identifying a high-risk tumor can encourage a more aggressive treatment strategy. We summarize four articles that used data mining to identify the type or severity of the prostate cancer.

Chandana, Leung, and Trpkov [9] constructed models with different combinations of automatic feature selection, sampling, and classifier to predict the stage of prostate cancer. Age, primary Gleason grade, secondary Gleason grade, PSA level, PSAD, DRE, TRUS, gland volume, positive biopsy core, total percent of cores involvement, and total cancer length in mm for 1,054 patients were used in the models. The prostate cancer was organ-confined in 934 patients. There was an extra prostatic extension in 120 patients.

ACCEPTED MANUSCRIPT

The authors generated 18 models with combinations of automatic feature selection (rough set features, PCA, continuous feature selection using a genetic algorithm (GA)), sampling (under-sampling, synthetic minority over-sampling technique (SMOTE), combined under-sampling and SMOTE), and classifier (SVM, K-Nearest Neighbor (KNN)). SMOTE generated synthetic examples of extra prostatic extension patients, based on those in the data set, that were used in training. Dempster-Shafer fusion produced classification probabilities based on the classification probabilities from multiple classifiers and generated new classifiers from combinations of the 18 models. A GA identified the best set of classifiers for fusion. SVM with rough set for feature selection and under-sampling had the best performance of the 18 models with an AUC of 0.8376. A Dempster-Schafer fusion of four models had a total accuracy of 90.1% and AUC of 0.8626.

Kattan et al. [32] developed a statistical model to predict the probability that prostate cancer is indolent or in need of treatment. They used data from 1,022 patients treated with retropubic radical prostatectomy. Patients with a PSA level greater than 20, Gleason grade four or five, greater than 50% biopsy cores positive, total cancer in biopsy greater than 20 mm, or total benign tissue in all cores less than 40 mm were excluded as being unlikely to have indolent cancer. This left 409 patients. The authors constructed three statistical models. The base model used PSA level, clinical stage, and primary and secondary biopsy Gleason grades. The medium model added percent of cores that were positive, and ultrasound prostate volume to the features of the base model. The full model replaced percent of cores that were positive with millimeters of cancerous and noncancerous tissue. Bootstrapping was used to build the model and leave-one-out

ACCEPTED MANUSCRIPT

analysis was used to examine the predictive probabilities. The three models had AUC values of 0.64, 0.74, and 0.79, respectively. The full model predicted too many indolent cancers (17% too high). The authors then translated their statistical models into nomograms.

Bone scan and CT imaging are commonly used to detect metastatic cancer by identifying bone metastases and enlarged lymph nodes, respectively. This is important in determining whether treatment with curative intent, such as surgery, is appropriate, or whether patients should initiate chemotherapy. Merdan et al. [44] and Risko et al. [50] developed logistic regression models based on data from more than 80% of community urology practices in Michigan to estimate the probability of a positive bone scan and a positive CT scan. The models were found to provide very good discrimination, based on estimates of AUC greater than 80%. Furthermore, the models were used to estimate efficiency and effectiveness of proposed guidelines for determining which patients should receive imaging tests. Criteria of Gleason score > 7 or PSA > 20 (for bone scan guidelines) and Gleason score > 20, PSA > 20, Clinical state > T2a (for CT guidelines) were estimated to result in fewer than 1% of patients having missed positive results. Implementation of these guidelines was estimated to substantially reduce the number of negative imaging results and the total number of imaging studies overall. The latter results are particularly important given the high cost of imaging.

4.2 Active Surveillance

Active surveillance has become a viable alternative to treatment. Active surveillance involves close observation of a patient with frequent visits to a doctor for a DRE, a PSA test, and possibly a biopsy. If there is a change in the tumor, a patient may need to decide

ACCEPTED MANUSCRIPT

on a treatment. Active surveillance allows a patient to delay treatment, thereby avoiding any side effects. The two articles summarized compared active surveillance to different treatment options.

Eldefrawy et al. [16] compared the cost of active surveillance to common treatment options (radical retropubic prostatectomy, robotic assisted radical prostatectomy, external beam radiation therapy, brachytherapy) for low-risk prostate cancer. Screening for prostate cancer has led to an increase in diagnosis, particularly in early stage prostate cancer. Though 17% to 20% of men will be diagnosed with prostate cancer within their lifetime, only 3% will die from prostate cancer. Overtreatment can have a negative impact on patient quality of life due to treatment side effects such as erectile dysfunction and incontinence. Active surveillance delays treatment, while having the patient undergo surveillance involving DRE, PSA testing, and biopsies. Procedure cost, professional fees, and inpatient costs were estimated based on Medicare reimbursement levels in the Miami area and the costs for a Miami area hospital for 2010. The authors estimated costs for treating low-risk prostate cancer over a 10-year period using a Markov model. The model considered the probabilities of various complications and recurrence from treatment options and the resulting costs. For example, each TRUS guided biopsy had a 1% chance of causing sepsis. Robotic assisted radical prostatectomy, the most commonly performed procedure for prostate cancer in the United States, had the second highest costs with a one-year cost of $17,824 and a 10-year cost of $22,762. Active surveillance had the lowest costs with a one-year cost of $1,154 and a 10-year cost of $13,116.

Hayes et al. [29] used a simulation model to compare the QALYs for men with low-risk prostate cancer associated with active surveillance and initial treatment

ACCEPTED MANUSCRIPT

(brachytherapy, intensity-modulated radiation therapy (IMRT), radical prostatectomy). Active surveillance involves regular PSA testing, DRE, and biopsies. Active surveillance delays treatment until the prostate cancer progresses or the patient chooses to begin treatment. In the United States, 16% to 40% of newly diagnosed prostate cancer patients meet the criteria for active surveillance. The authors constructed a state transition model that they analyzed using Monte Carlo simulation. The model considered probabilities of side effects associated with brachytherapy, IMRT, and radical prostatectomy and with the disease in general. Probabilities were estimated based on a systematic literature review. Patient utilities were estimated by asking men not diagnosed with prostate cancer using the time-trade-off method. A hypothetical cohort of 65-year-old men with newly diagnosed, clinically localized, low-risk prostate cancer was used as the patient population. Patients undergoing active surveillance had the best performance with 11.07 QALYs, compared to 10.57 QALYs for brachytherapy, 10.51 QALYs for IMRT, and 10.23 QALYs for radical prostatectomy. Active surveillance had a higher risk of death from prostate cancer (11% compared to 9% for initial treatment). For those men undergoing active surveillance, 61% received treatment (brachytherapy, IMRT, radical prostatectomy) having been under active surveillance for a median of 8.5 years.

4.3 Hormone Therapy

Hormone therapy can be used to shrink a cancerous tumor in the prostate before it is treated with radiation. Hormone therapy shrinks a tumor to a minimum size, after which the therapy loses effectiveness and the tumor starts growing. Less radiation needs to be delivered when a tumor is at its smallest, thereby reducing the side effects from the

ACCEPTED MANUSCRIPT

treatment. The article summarized tried to identify when a tumor has reached its minimum size.

Lavieri et al. [35] used a dynamic Kalman filter model to predict when to begin radiation therapy. A Kalman filter is a recursive procedure that computes the optimal estimator of the state vector at each time period based on a series of noisy measurements. Since it is believed that the PSA level is lowest at the same time that the tumor's size is at a minimum, PSA level is used in lieu of tumor size. PSA level over time is modeled using a log quadratic curve, which has an average R^2 of 0.9 for the 163 patients in the study. The authors estimated the prior distribution of the curve parameters, then the distribution of the estimated time of the nadir. A Kalman filter is used to update the estimates of the curve parameters. Finally, clustering was used to identify subgroups of patients with similar responses to hormone therapy. Protocol in British Columbia required that radiation therapy begin if the PSA level started to rise, if after four months the PSA level was below 0.05 ng/mL, or eight months after beginning the therapy. Four new policies were proposed that used the predictive model to determine if the nadir had a threshold probability of being reached within a fixed time period of PSA tests. Two of these policies outperformed the current policy; beginning treatment if the nadir is predicted to occur before the next PSA test or beginning treatment if the nadir is likely to occur within one month of the next PSA test. The mean absolute difference between the time of the nadir to the beginning of radiation therapy is 45 days for the current policy, but only 29 or 36 days for the two new policies.

ACCEPTED MANUSCRIPT

4.4 Brachytherapy

Brachytherapy places radioactive seeds inside the prostate. The placement of the seeds affects the amount of radiation received by the tumor and other parts of the prostate. When developing a treatment plan, the objective is to deliver an appropriate amount of radiation to the tumor and to the tissue identified as at risk of cancer, while sparing healthy tissue. Brachytherapy carries risks including urinary problems and erectile dysfunction. We summarize five articles that used mathematical models to optimize the placement of the seeds and needles and analyzed the risks of underestimating the Gleason score when treating with brachytherapy.

Ferrari et al. [20] proposed a mathematical model that allowed for the simultaneous optimization of seeds and needles in brachytherapy planning. The model took into account the prostate, urethra, rectum, and bladder. The volume was divided into discrete points where the radioactive seeds could be placed. Several quality metrics were used including the percentage of each tissue type receiving given dose levels, dose non-uniformity ratio, and percent of treatment volume achieving the prescribed dose. The dose for each point was calculated based on the sum of the radiation received from the seeds (inversely related to the distance from the seed). Each seed is delivered by a needle and the needle can only deliver a finite number of seeds in a straight line that meet certain spacing criteria. The model minimized the number of needles and seeds, while maximizing the percentage of treatment volume above the prescribed dose and healthy tissue below a certain dose threshold. The authors developed a GA to find good solutions to the mathematical model. The GA used a two-dimensional chromosome of integers that corresponded to the seed placement and needles. The GA was used to develop plans for 11 test patients. The authors established criteria for acceptable plan performance based on

ACCEPTED MANUSCRIPT

how well the plan matched the prescribed dose. The GA took an average of 20 minutes and found acceptable solutions in 10 of 11 cases.

Lee and Zaider [38] devised optimization models and computational techniques for real-time intraoperative 3D treatment planning in brachytherapy. Traditionally, treatment plans had been constructed several days or weeks prior to the implantation of the radioactive seeds, following an ultrasound or computerized tomography scan of the prostate. The manual construction of a treatment plan is a lengthy process and can take several iterations, and the images used to construct the treatment plan are often different from the images obtained when it is time to insert the seeds. This is in part because the prostate volume measured at the first imaging is often different from the volume observed in the operating room. Discrepancies between the original imaging and the imaging during implantation can increase the chance of undesirable side effects. Lee and Zaider formulated a mixed integer program (MIP) to optimize the placement of the seeds throughout the prostate. Constraints included dosimetric constraints for the tumor and critical structures. Constraints desired by the physicians, such as limiting the number of seeds, can also be included. In order to solve the MIP in real time, various techniques, including matrix reduction and a branch-and-cut environment, were used. The resulting plans delivered the prescribed dose to 98% of the prostate, reduced the number of seeds implanted by 20-30%, and reduced urethra dose by 23-28% compared to other computerized techniques.

Cambio et al. [6] used the probability of each outcome and the cost of each decision to assess the cost-benefit and outcome of recommending brachytherapy to men with a prostate biopsy Gleason score of six. The Gleason score from a sextant biopsy agrees

ACCEPTED MANUSCRIPT

with the Gleason score from a prostatectomy 46% to 63% of the time. An extended core biopsy (10 or more cores) agrees 76% of the time. A Gleason score of six is considered low risk of biochemical recurrence of prostate cancer, while seven is considered intermediate risk. Brachytherapy is considered an appropriate standalone treatment for a patient with a Gleason score six or lower, though it is not recommended by itself for a patient with a Gleason score of seven or higher. The authors estimated the treatment cost and quality of life changes of brachytherapy from previous research. A group of 60-year-old patients with a prostate biopsy Gleason score of six were used for the analysis. Patients either had a true Gleason score of six or they had been undergraded and had a Gleason score of either seven or eight to 10. Depending on the true state of the cancer, there were different probabilities that brachytherapy would successfully prevent biochemical failure (a rise in PSA levels after treatment considered to signal the recurrence of prostate cancer) within five years. A variety of biopsy accuracies were evaluated. For example, raising the agreement of the biopsy and prostatectomy Gleason score from 60% to 80% reduced average costs of treatment and recurrence from $63,780 to $62,929 per patient.

Fu, Yu, and Liu [23] developed a GA for planning of prostate brachytherapy prior to the beginning of surgery, and replanning after some of the radioactive seeds have been placed. During brachytherapy surgery, needles can be deflected, changing the placement of radioactive seeds. Dynamic intraoperative treatment planning adjusts the placement of the remaining seeds based on the actual placement of the seeds already inserted into the prostate. The authors developed a GA that first optimized the needle pattern (assuming one of a few seed spacing rules) and then optimized the seed spacing. An original plan

ACCEPTED MANUSCRIPT

was constructed for a patient. Needles were randomly deflected, and seeds delivered as planned along the deflected needle. After one-quarter of the needles were delivered, the actual positions were known through imaging, and a new plan was constructed. After one-third of the needles in the second plan were delivered, a third plan was constructed. After one-half of the needles in the third plan were delivered, a fourth plan was constructed. After all the needles in the fourth plan were delivered, a fifth plan was constructed to ensure the tumor had received the necessary radiation. The quality of the reoptimized plans was compared to the original plan based on the number of seeds used, maximum urethral dose, and maximum rectal dose. The performance was based on three simulated runs for each of 10 patients using two different types of radioactive seeds, iodine-125 with a prescription dose of 145GY and palladium-103 with a prescription dose of 115Gy. Reoptimization led to a 10% to 20% increase in the number of seeds used, depending on the type of seed used. The deflections led to a 17% to 28% increase in maximum urethral dose and 16% to 42% increase in maximum rectal dose, depending on the type of seed.

Lee and Zaider [37] formulated a MIP to optimize the placement of the seeds throughout the prostate. A grid is placed over the potential seed locations in the prostate and binary variables were used to indicate whether a seed is placed at each location. The seed locations contribute to the radiation dosage at every point, and constraints include upper and lower bounds of radiation dosage at each point. Because it is not always possible to simultaneously meet all dose constraints, two models are constructed that either maximize the number of points meeting these constraints or minimize the weighted sum of the deviations from the dose bounds. Constraints included dosimetric constraints

ACCEPTED MANUSCRIPT

for the tumor and critical structures. Constraints desired by the physicians, such as limiting the number of seeds, can also be included. Computational strategies included matrix reduction, perturbation, and a penalty-based adaptive primal heuristic procedure. The MIP solver was able to solve sample problems within 15 CPU minutes and improve the results from existing computer-aided planning methods. The MIP solver used fewer seeds and needles, and provided better coverage and conformity.

4.5 Patient Choice

Men who receive an early prostate cancer diagnosis and are aggressively treated have a higher survival rate than those diagnosed with later stages of the disease. However, there is concern that screening can lead to unnecessary treatment, particularly for men with low-risk disease. When deciding on the best treatment for prostate cancer, a patient needs to consider two key factors: (1) the risk of prostate cancer progression and mortality and (2) the potential side effects of the various treatment options. We summarize three articles that developed models to help a patient select a course of treatment.

Simon [57] constructed a multiattribute utility model to help a patient consider five treatment alternatives: surgery (radical prostatectomy), external radiation, seed radiation, dual radiation, and no treatment. In the first part of the model, the life expectancy of a patient and the probability of death from prostate cancer were determined. The probabilities of three side effects (impotence, incontinence, toxicity) that depend on the type of treatment and factors specific to a patient were developed. In the second part of the model, a patient's preferences were analyzed. For the three side effects, a patient needed to determine what percentage of remaining life he would be willing to forego to

ACCEPTED MANUSCRIPT

avoid a particular side effect (this is known as the emotional weight). Simon expressed the possibility of each side effect in terms of a reduction in life span. For a patient, a life score was developed that took into account the length and quality of life. It provided a patient with a weighted average utility for a specific treatment. For example, a life score of 95 for seed radiation would be preferred to a life score of 85 for surgery. Simon ran the model on different patient profiles and found some interesting results. Younger men who did not have an aversion to side effects had the highest life score for surgery. Older men had high life scores for treatments that were less aggressive than surgery. One result was controversial: external radiation was used too often. Simon implemented a web-based version of this model that had an average of 400 hits per week in 2007.

Liberatore et al. [39] developed a decision counseling method, based on the Analytic Hierarchy Process (AHP), to help men make informed decisions about whether to screen for prostate cancer. AHP is a technique for analyzing complex decisions that takes into account the factors affecting the final decision. As part of a randomized controlled trial, the authors scheduled sessions for 129 men with a trained health educator to discuss prostate cancer screening. They used AHP to process a participant's three top decision factors as identified by the participant during the session and to generate a preference score that reflected a participant's preference to screen. The score was computed during the session using a calculator and was then given to the participant prior to the scheduling decision. The preference score was a statistically significant predictor of a participant's decision to schedule a screening exam. An analysis of the preference score showed certain decision factors (emotional factors), such as fear of getting cancer, were the most influential in a patient's decision to be screened. Liberatore et al. concluded that AHP

ACCEPTED MANUSCRIPT

was successful in eliciting decision factors. They supported its implementation as part of decision counseling in the future.

Chapman et al. [10] constructed a multi-attribute utility (MAU) model to examine the trade-offs between different treatment options for prostate cancer patients, with the goal of aiding patients in making treatment decisions. The authors considered five health attributes that can be affected by patient treatment: pain, mood, sexual function, bladder and bowel function, and fatigue and energy. Their study had 57 patients from two Chicago Veterans Administration health clinics with either localized or metastasized prostate cancer. The patients' preferences were measured using time trade-off (TTO) judgments. The participants were asked how many years of full health were equivalent to 10 years in one of three levels for each health attribute. The patients then divided 100 points between the five health attributes to indicate their relative importance. Patients tended to place the most weight on pain followed by bladder and bowel function. The MAU model was compared to a global TTO preference assessment. The result of the comparison indicated that the MAU model may be useful in aiding patient decision making.

4.6 Recurrence

There are many forms of treatment available for prostate cancer. Younger patients, who are healthy, are often encouraged to receive a radical prostatectomy. Older patients who are not good candidates for surgery are more often steered away from a radical prostatectomy towards other forms of treatment such as external beam radiation therapy or brachytherapy. Regardless of the treatment selected, there is the potential for recurrence following treatment.

ACCEPTED MANUSCRIPT

If a patient's PSA level rises above 0.1 ng per mL following a radical prostatectomy, the patient is considered to be at risk that his cancer persists. According to Catalona and Smith [8], the five-year probability of nonprogression after a radical retropubic prostatectomy for prostate cancer is 78%. We summarize seven articles that used data mining to identify patients at risk of prostate cancer recurrence.

Dancea et al. [12] developed an SVM model to divide patients into risk classes prior to a radical prostatectomy to aid in surgical decisions. They used a data set with 14 medical attributes for 399 patients: classification of malignant tumors, Gleason score, presence of median endo-vesical lobe, prostate volume, pre-operative PSA level, international index of erectile function, quality of life, abort operation, surgery technique, nerve sparing, surgery time, postoperatory hospital time, complications, and risk class. They trained their SVM model on 369 patients and tested on 30. The SVM model correctly classified 93% of the test set; it correctly classified all low-risk and high-risk patients.

Shariat et al. [56] reviewed the literature to compare nomograms, risk groupings, artificial neural networks, probability tables, and regression trees to determine their effectiveness in predicting the risk of cancer recurrence. The published studies evaluated models on predictive accuracy, performance characteristics according to risk level, generalizability, and level of complexity. The studies that were examined by the authors applied nomograms and another model to a common external data set with known patient outcomes. Nomograms had superior predictive accuracy to risk groupings and probability tables. Regression trees constructed with CART analyses were easy to use and offered greater model-fitting flexibility than nomograms. In a head-to-head study, predictive

ACCEPTED MANUSCRIPT

accuracy was 70% for CART models and 84% for nomograms. ANNs have a high level of complexity and are typically used as a black box method, but they can fit very complex patterns. In a study that applied nomograms and ANNs to an external data set, predictive accuracy was 70.6% and 67%, respectively.

Androgen deprivation therapy is a common treatment for recurrent prostate cancer, but androgen-independent prostate cancer will eventually develop. NOXA and PUMA are two proteins whose presence may affect prostate cancer recurrence. Diallo et al. [13] generated recursive partitioning and regression tree (RPART) models with NOXA and PUMA to predict biochemical recurrence in prostate cancer patients. RPART models use a two-stage procedure to produce binary decision trees. The study had 43 patients with healthy prostate tissues, 62 patients presenting primary prostate cancer tissues, and 30 patients with hormone-refractory prostate cancer. The authors constructed RPART models using NOXA and/or PUMA with combinations of preoperative PSA level, Gleason score, pathologic stage, and resection margin status (the presence of cancer in the surgical margin that surrounded the designated tumor removed during the radical prostatectomy). The RPART model with the best performance was based on NOXA, PUMA, and the resection margin status. Nine of the top 10 models included NOXA to predict biochemical recurrence, while only three of the top 10 included PUMA.

Biochemical failure-free survival (bFFS) refers to the probability that a patient's PSA level does not rise for three consecutive tests following external beam radiotherapy for prostate cancer. If a patient's PSA level rises for three consecutive tests following treatment, this is a strong indicator of recurrent cancer. In order to determine a patient's risk of relapse, Churilov et al. [11] applied clustering techniques to a set of post-treatment

ACCEPTED MANUSCRIPT

patients who had received external beam radiation therapy. The data set had Gleason score, tumor stage, PSA level at diagnosis, and age for 258 patients treated at the William Buckland Radiotherapy Center (WBRC) in Melbourne, Australia. The PSA level of each patient was taken at three-month intervals for the first year following the start of treatment, six-month intervals the next year, and then annually. Churilov et al. applied an optimization algorithm for clustering based on earlier work. The clustering algorithm grouped the patients into 10 clusters. WBRC used a rule-based method with seven clusters. In each cluster, a patient was identified as low, intermediate, or high risk of having a biochemical failure in the five years following treatment. The WBRC method assigned 51.2% of all patients to an intermediate-risk category and 8.2% to a low-risk category. The clustering algorithm developed by Churilov et al. assigned 26% to low risk and 34% to intermediate risk, while maintaining a similar probability of bFFS for the categories as the WBRC method. Those defined as low-risk by the clustering method had a 71.6% of five-year bFFS, compared to 76.2% for those classified by the WBRC method. Patient's classified as intermediate-risk had a 53.1% and a 53.4% five-year bFFS for the clustering method and WBRC method, respectively. Churilov et al. improved the classification of patient's risk, allowing doctors to make better decisions regarding post-treatment care.

Seker et al. [54] constructed a Fuzzy K-Nearest Neighbors (FK-NN) classifier to predict the outcome of prostate cancer treatments for patients. Data from 41 men with histologically proven prostate cancer were used to construct FK-NN, linear regression, and ANN models. Four conventional indicators of prostate cancer (tumor stage, skeletal metastasis, Gleason score, serum PSA level) and two experimental indicators (p53

ACCEPTED MANUSCRIPT

immunostaining, bcl-2 immunostaining) were used to predict a patient's response to hormonal treatment, radical surgery, or observation. Response to treatment was classified as either having no response to any type of treatment, complete response to treatment (no tumor progression for patients undergoing observation), or relapse following successful start (tumor progression for patients undergoing observation). Each model was trained using a leave-one-out method. Each model achieved a best predictive accuracy using a different subset of the indicators. Serum PSA and treatment type were the best indicators of patient outcome. The logistic regression model had a predictive accuracy of 41.46%. The ANN model had a predictive accuracy of 53.66%. The FK-NN model had a predictive accuracy of 60.98% if the number of nearest neighbors was set to 1 and 63.42% if it was set to 2 or 3.

Seker et al. [55] extended their work by constructing a hybrid neuro-fuzzy rule-based system to predict the treatment outcome of a prostate cancer patient. A set of rules was generated to determine the classification of a patient. Each rule had a premise (a set of variables and conditions on which the rule was based) and a consequence (the classification based on the premise). A neural network computed the parameters of the premise for each rule. Singular value decomposition selected the most and least important rules generated by using a k-means clustering method. The hybrid approach achieved a predictive accuracy of 63.42%.

Kattan, Wheeler, and Scardino [33] developed a nomogram to predict the progression of prostate cancer in men treated with radical prostatectomy. Pretreatment PSA level, Gleason sum from the surgical specimen, prostatic capsular invasion (the presence of cancer cells in the capsule around the tumor), surgical margin status (the presence of

ACCEPTED MANUSCRIPT

cancer cells in the expected healthy tissue removed with the tumor), seminal vesicle invasion (the presence of cancer cells in muscular wall of the seminal vesicle), and lymph node status (the presence of cancer cells in the lymph nodes) for 996 men who underwent radical prostatectomy were used in the development and testing of the nomogram. Progression of prostate cancer was defined as PSA levels rising to 0.4 ng/ml or higher, a second PSA test that was higher than the first by any amount, or treatment with radiation or hormone therapy within seven years of surgery. The nomogram had an AUC of 0.88, implying that for 88% of pairs of patients, the patient with the larger score will relapse first. If variables whose values were known only after the radical prostatectomy were removed, the AUC would be reduced to 0.74.

4.7 Costs

With an aging population and a current expenditure of $11.8 billion for prostate cancer treatment in the United States [61], it is important to understand where money is currently being spent and what the costs are likely to be in the future. Since many of those affected by prostate cancer are above the age of 65, Medicare pays for much of the screening and treatment. Accurately predicting future Medicare expenditures is important in long-term planning. The two articles that are summarized explored the financial costs of prostate cancer with one article examining the costs prior to treatment and the other predicting future Medicare costs.

Ekwueme, Stroud, and Chen [15] conducted a systematic review of published cost data for prostate cancer treatment in the United States and other industrialized countries, and then performed a statistical analysis of the data. The authors identified 262 articles published between 1980 and 2003 whose title or abstract implied cost information on

ACCEPTED MANUSCRIPT

prostate cancer treatment. From these, 28 (15 from the United States and 13 from other industrialized countries) were available in full text, written in English, contained original resource cost data, and included screening, diagnosing, and staging costs. All costs were converted to 2003 U.S. dollars using the consumer price index (for the country of origin) and exchange rates. A weighted mean for the cost per man screened was taken, weighted by the number of men screened in each study. A Monte Carlo simulation method was used to determine uncertainty in the pooled resource costs. For the United States, the cost for PSA screening was $37.23 ($30.92 internationally) and DRE was $31.77 ($33.54 internationally). In the United States, costs have decreased for biopsies and PSA screening, while they have increased for clinical staging, pathologic or histologic analysis, TRUS, urology consultation, and DRE. Internationally, resource costs associated with biopsies, PSA screening, pathologic or histologic analysis, TRUS, urology consultation, and DRE have decreased over time (over the course of one or more of the published studies).

Penberthy et al. [48] analyzed the factors that predict Medicare expenditures in patients with breast, colorectal, lung, or prostate cancer. The authors combined data from the Virginia Cancer Registry, Medicare Provider Analysis and Review files, Medicare Automated Data Retrieval System, Medicare Health Insurance Master file, Medicare Annual Demographic files, Area Resource File, and 1990 Census Data for Zip Code Level information. Expenditures were analyzed for one year following diagnosis, with cost being defined as the amount the Health Care Financing Administration reimbursed the health care provider. The data set had 1,952 breast, 2,563 colorectal, 3,331 lung, and 3,179 prostate cancer patients. Treatments for prostate cancer were categorized as

ACCEPTED MANUSCRIPT

definitive surgery (287 patients), nonsurgical treatment (1,827 patients), surgery plus nonsurgical therapy (70 patients), and no treatment (995 patients). A two-stage least squares analysis was performed to predict costs. The R^2 value for prostate cancer was 0.38. The mean cost for prostate cancers was $14,361 for Virginia (1985-1988). Treatments that involved surgery tended to cost more as did treatment for patients with any comorbidity. Predicted cost increased with income of the patient. Patients who did not survive one year following diagnosis were found to cost 80% more than those who survived their first year.

5. Conclusions

Over the past 15 years, OR models and methods have been applied to diverse problems in the screening, detection, and treatment of prostate cancer. From Figure 1, we see that the number of publications on the use of OR models and methods for prostate cancer has increased substantially. Based on this figure alone, we would expect that the use and impact of OR will grow significantly in the next decade. But there is more.

A Google Scholar search for articles based on the keywords “prostate cancer” and “analytics” (a term often used synonymously for OR) turns up 202 articles in 2010, 255 articles in 2011, 353 articles in 2012, 461 articles in 2013, and 528 articles in 2014. This is a growth rate of about 27% per year. Furthermore, if we perform the search replacing “prostate cancer” with “colorectal cancer,” we obtain a similar annual growth rate of 25% from 2010 through 2014. We anticipate comparable trends for other types of cancer. We hope that our review and the attendant annotations will serve as a useful guidepost to help those conducting research on new problems in cancer care in the future.

ACCEPTED MANUSCRIPT

Appendix

Table 1A. Count of OR articles in prostate cancer research on the Web of Knowledge (2000-2015).

OR Model/Method	Number of Articles
Machine learning	186
Markov model	129
Markov chain	69
Integer programming	35
Linear programming	31
Microsimulation	23
Optimal control	21
Markov decision	19
Discrete event simulation	8
Dynamic programming	7
Analytic hierarchy process	5
Multi-attribute utility model	3
Agent-based model	2
Total	538
Total (unique articles*)	523

*15 articles use more than one model or method

Table 1B. OR models and methods used in prostate cancer research.

Area	Article Number in References	Model/Method
Decision Analysis	6	DT
	10	MAU
	16	MM
	39	AHP
	57	MAU
Machine Learning	2	KNN
	9	GA, KNN, PCA, SVM
	11	CL
	12	SVM
	13	RPART, RT
	24	CART, RT
	28	QUEST, RT
	34	ICA, NMF, PCA, PLS, SVM
	36	ANN
	43	ANN

ACCEPTED MANUSCRIPT

Table 1B. (continued)

	47	ANN, SVM
	52	ANN, NOM
	54	ANN, FKNN
	55	ANN, FKNN
	56	ANN, NOM, CART, RT
	60	CART, RT
Optimization	20	GA
	23	GA
	37	MIP
	38	MIP
	67	MDP
	68	MDP
Simulation	14	MIC, MM
	15	MC
	17	MIC, MM
	18	MIC, MM
	19	MIC, MM
	26	MIC, MM
	27	MIC
	29	MC
	51	MC, MM
	58	TDS
	64	MX
	65	MIC, MM
Statistics	4	LR
	32	NOM
	33	NOM
	35	KF
	44	LR
	46	NOM
	48	TSLS
	50	LR
	59	NOM
	62	LR

AHP	Analytic Hierarchy Process	MIP	Mixed Integer Program
ANN	Artificial Neural Network	MM	Markov Model
CART	Classification and Regression Tree	MX	Mixture Model
CL	Clustering	NMF	Nonnegative Matrix Factorization
DT	Decision Tree	NOM	Nomogram
FKNN	Fuzzy K-Nearest Neighbor	PCA	Principal Components Analysis

ACCEPTED MANUSCRIPT

GA	Genetic Algorithm
ICA	Independent Component Analysis
KF	Kalman Filter
KNN	K-Nearest Neighbor
LR	Logistic Regression
MAU	Multi-Attribute Utility
MC	Monte Carlo Simulation
MDP	Markov Decision Process
MIC	Microsimulation
PLS	Partial Least Squares
QUEST	Quick, Unbiased, and Efficient Statistical Tree
RPART	Recursive Partitioning and Regression Tree
RT	Regression Tree
SVM	Support Vector Machine
TDS	Three Dimensional Simulation
TSLS	Two Stage Least Squares

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

1. Ahmed, H. U., Kirkham, A., Arya, M., Illing, R., Freeman, A., Allen, C., and Emberton, M. "Is it Time to Consider a Role for MRI Before Prostate Biopsy?" *Nature Reviews Clinical Oncology,* 6(4), 197-206 (2009)

2. Anderson, D., Golden, B., Wasil, E., Zhang, H., "Predicting Prostate Cancer Risk Using Magnetic Resonance Imaging Data," forthcoming in *Information Systems and e-Business Management* (2014)

3. Andriole, Gerald L., et al., "Mortality Results from a Randomized Prostate-Cancer Screening Trial," *New England Journal of Medicine,* 360(13), 1310-1319 (2009)

4. Ankerst, D. P., Hoefler, J., Bock, S., Goodman, P. J., Vickers, A., Hernandez, J., Sokoll, L. J., Sanda, M. G., Wei, J. T., Leach, R. J., Thompson, I. M., "Prostate Cancer Prevention Trial Risk Calculator 2.0 for the Prediction of Low- vs High-grade Prostate Cancer," *Urology,* 83(6), 1362-1368 (2014)

5. Bankhead, C. "Sextant Biopsy Helps in Prognosis of PCa, but It's Not Foolproof," *Urology Times,* 25(8), (1997)

6. Cambio, A. J., Ellison, L. M., Chamie, K., deVere White, R. W., and Evans, C. P., "Cost-Benefit and Outcome Analysis: Effect of Prostate Biopsy Undergrading," *Urology,* 69(6), 1152-1156 (2007)

7. Carter, H. B., Albertsen, P. C., Barry, M. J., Etzioni, R., Freedland, S. J., Greene, K. L., Holmberg, L., Kantoff, P., Konety, B., Murad, M. H., Penson, D. F., and Zietman, A. L., "Early Detection of Prostate Cancer: AUA Guideline," *The Journal of Urology,* 190(2), 419-426 (2013)

8. Catalona, W. J., and Smith, D. S. "5-year Tumor Recurrence Rates after Anatomical Radical Retropubic Prostatectomy for Prostate Cancer," *The Journal of Urology,* 152(5), 1837-1842 (1994)

9. Chandana, S., Leung, H., and Trpkov, K., "Staging of Prostate Cancer Using Automatic Feature Selection, Sampling and Dempster-Shafer Fusion," *Cancer Informatics,* 7, 57-73 (2009)

10. Chapman, G. B., Elstein, A. S., Kuzel, T. M., Nadler, R. B., Sharifi, R., and Bennett, C. L., "A Multi-Attribute Model of Prostate Cancer Patients' Preferences for Health States," *Quality of Life Research,* 8(3) 171-180 (1999)

11. Churilov, L., Bagirov, A. M., Schwartz, D., Smith, K., and Dally, M., "Improving Risk Grouping Rules for Prostate Cancer Patients with Optimization," In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004.,* IEEE, 1-9 (2004)

ACCEPTED MANUSCRIPT

12. Dancea, O., Gordan, M., Dragan, M., Stoian, I., and Nedevschi, S., "Postoperatory Risk Classification of Prostate Cancer Patients Using Support Vector Machines," *IEEE International Conference on Automation, Quality and Testing, Robotics*, Vol. 3, 53-56 (2008)

13. Diallo, J. S., Aldejmah, A., Mouhim, A. F., Péant, B., Fahmy, M. A., Koumakpayi, I. H., Sircar, K., Bégin, L. R., Mes-Masson, A. M, and Saad, F., "NOXA and PUMA Expression Add to Clinical Markers in Predicting Biochemical Recurrence of Prostate Cancer Patients in a Survival Tree Model," *Clinical Cancer Research*, 13(23), 7044-7052 (2007)

14. Draisma, G., Etzioni, R., Tsodikov, A., Mariotto, A., Wever, E., Gulati, R., Feuer, E., and de Koning, H., "Lead Time and Overdiagnosis in Prostate-Specific Antigen Screening: Importance of Methods and Context," *Journal of the National Cancer Institute*, 101(6) 374-383 (2009)

15. Ekwueme, D. U., Stroud, L. A., and Chen, Y., "Cost Analysis of Screening for, Diagnosing, and Staging Prostate Cancer Based on a Systematic Review of Published Studies," *Preventing Chronic Disease*, 4(4), (2007)

16. Eldefrawy, A., Katkoori, D., Abramowitz, M., Soloway, M. S., and Manoharan, M., "Active Surveillance vs. Treatment for Low-Risk Prostate Cancer: A Cost Comparison," *Urologic Oncology: Seminars and Original Investigations*, 31(5), 576-580 (2011)

17. Etzioni, R., Cha, R., and Cowen, M. E., "Serial Prostate Specific Antigen Screening for Prostate Cancer: A Computer Model Evaluates Competing Strategies," *The Journal of Urology*, 162(3), 741-748 (1999)

18. Etzioni, R., Gulati, R., Falcon, S., and Penson, D. F., "Impact of PSA Screening on the Incidence of Advanced Stage Prostate Cancer in the United States: A Surveillance Modeling Approach," *Medical Decision Making*, 28(3), 323-331 (2008)

19. Etzioni, R., Penson, D. F., Legler, J. M., di Tommaso, D., Boer, R., Gann, P. H., and Feuer, E. J. "Overdiagnosis due to Prostate-Specific Antigen Screening: Lessons from US Prostate Cancer Incidence Trends," *Journal of the National Cancer Institute*, 94(13), 981-990 (2002)

20. Ferrari, G., Kazareski, Y., Laca, F., and Testuri, C. E. "A Model for Prostate Brachytherapy Planning with Sources and Needles Position Optimization," *Operations Research for Health Care*, 3(1), 31-39 (2014)

21. Filella, X., Foj, L., Mila, M. et al., "PCA3 in the Detection and Management of Early Prostate Cancer," *Tumor Biology*, 34(3), 1337-1347 (2013)

ACCEPTED MANUSCRIPT

22. Food and Drug Administration, HHS., "Hybritech Inc.; Premarket Approval of Tandem-R, E, and ERA PSA Assays," Federal Register, 59(198), Friday, October 14, 1994, FR Doc No: 94-25516, From the Federal Register Online via the Government Printing Office, http://www.gpo.gov/fdsys/pkg/FR-1994-10-14/html/94-25516.htm (1994)

23. Fu, L., Yu, Y., and Liu, H., "Genetic Algorithm-Based Dynamic Intraoperative Treatment Planning for Prostate Brachytherapy," *Artificial Intelligence Applications and Innovations*, Eds. Li, D. and B. Wang, Proceedings of Second IFIP Conference on Artificial Intelligence Applications and Innovations, Springer US, 425-434 (2005)

24. Garzotto, M., Beer, T. M., Hudson, R. G., Peters, L., Hsieh, Y. C., Barrera, E., Klein, T., and Mori, M., "Improved Detection of Prostate Cancer Using Classification and Regression Tree Analysis," *Journal of Clinical Oncology*, 23(19), 4322-4329 (2005)

25. Greene, K. L., Albertsen, P. C., Babaian, R. J., Carter, H. B., Gann, P. H., Han, M., Kuban, D. A., Sartor, A. O., Stanford, J. L., Zietman, A., and Carroll, P., "Prostate Specific Antigen Best Practice Statement: 2009 Update," *The Journal of Urology*, 182(5), 2232-2241 (2009)

26. Gulati, R., Gore, J. L., and Etzioni, R., "Comparative Effectiveness of Alternative Prostate-Specific Antigen–Based Prostate Cancer Screening Strategies: Model Estimates of Potential Benefits and Harms," *Annals of Internal Medicine*, 158(3) 145-153 (2013)

27. Gulati, R., Inoue, L. Y., Gore, J. L., Katcher, J., and Etzioni, R., "Individualized Estimates of Overdiagnosis in Screen-Detected Prostate Cancer," *Journal of the National Cancer Institute*, 106(2), (2014)

28. Gülkesen, K. H., Köksal, İ. T., Özdem, S., and Saka, O., "Prediction of Prostate Cancer using Decision Tree Algorithm," *Turkish Journal of Medical Science*, 40(5), 681-686 (2010)

29. Hayes, J. H., Ollendorf, D. A., Pearson, S. D., Barry, M. J., Kantoff, P. W., Stewart, S. T., Bhatnagar, V., Sweeney, C. J., Stahl, J. E., and McMahon, P. M., "Active Surveillance Compared with Initial Treatment for Men with Low-Risk Prostate Cancer: A Decision Analysis," *Journal of the American Medical Association*, 304(21), 2373-2380 (2010)

30. Hoffman, R. M., Stone, S. N., Espey, D., and Potosky, A. L. "Differences Between Men with Screening-Detected versus Clinically Diagnosed Prostate Cancers in the USA," *BMC Cancer*, 5(1), 27 (2005)

31. Howden, L. M. and Meyer, J. A. "Age and Sex Composition: 2010," *United States Census Bureau*, May (2011)

ACCEPTED MANUSCRIPT

32. Kattan, M. W., Eastham, J. A., Wheeler, T. M., Maru, N., Scardino, P. T., Erbersdobler, A., Graefen, M., Huland, H., Koh, H., Shariat, S., Slawin, K., and Ohori, M., "Counseling Men with Prostate Cancer: A Nomogram for Predicting the Presence of Small, Moderately Differentiated, Confined Tumors," *The Journal of Urology*, 170(5), 1792-1797 (2003)

33. Kattan, M. W., Wheeler, T. M., and Scardino, P. T., "Postoperative Nomogram for Disease Recurrence after Radical Prostatectomy for Prostate Cancer," *Journal of Clinical Oncology*, 17(5), 1499-1499 (1999)

34. Kelm, B. M., Menze, B. H., Zechmann, C. M., Baudendistel, K. T., and Hamprecht, F. A., "Automated Estimation of Tumor Probability in Prostate Magnetic Resonance Spectroscopic Imaging: Pattern Recognition vs Quantification," *Magnetic Resonance in Medicine*, 57(1),150-159 (2007)

35. Lavieri, M. S., Puterman, M. L., Tyldesley, S., and Morris, W. J., "When to Treat Prostate Cancer Patients Based on Their PSA Dynamics," *IIE Transactions on Healthcare Systems Engineering*, 2(1), 62-77 (2012)

36. Lawrentschuk, N., Lockwood, G., Davies, P., Evans, A., Sweet, J., Toi, A., and Fleshner, N. E., "Predicting Prostate Biopsy Outcome: Artificial Neural Networks and Polychotomous Regression are Equivalent Models," *International Urology and Nephrology*, 43(1), 23-30 (2011)

37. Lee, E. K., and Zaider, M., "Mixed Integer Programming Approaches to Treatment Planning for Brachytherapy–Application to Permanent Prostate Implants," *Annals of Operations Research*, 119(1-4), 147-163 (2003)

38. Lee, E. K., and Zaider, M., "Operations Research Advances Cancer Therapeutics," *Interfaces*, 38(1), 5-25 (2008)

39. Liberatore, M., Nydick, R., Daskalakis, C., Kunkel, E., Cocroft, J., and Myers, R., "Helping Men Decide About Scheduling a Prostate Cancer Screening Exam," *Interfaces*, 39(3), 209-217 (2009)

40. Loeb, S., Carter, H. B., Berndt, S. I., Ricker, W., and Schaeffer, E. M. "Complications after Prostate Biopsy: Data from SEER-Medicare," *The Journal of Urology*, 186(5), 1830-1834 (2011)

41. Ma, X., Wang, R., Long, J. B., Ross, J. S., Soulos, P. R., Yu, J. B., Makarov, D. V., Gold, H. T., and Gross, C. P. "The Cost Implications of Prostate Cancer Screening in the Medicare Population," *Cancer*, 120(1), 96-102 (2014)

42. Mariotto, A. B., Yabroff, K. R., Shao, Y., Feuer, E. J., and Brown, M. L., "Projections of the Cost of Cancer Care in the United States: 2010-2020," *Journal of the National Cancer Institute*, 103(2), 1-12 (2011)

ACCEPTED MANUSCRIPT

43. Matulewicz, L., Jansen, J. F., Bokacheva, L., Vargas, H. A., Akin, O., Fine, S. W., Shukla-Dave, A., Eastham, J. A., Hricak, H., Koutcher, J. A., and Zakian, K. L. "Anatomic Segmentation Improves Prostate Cancer Detection with Artificial Neural Networks Analysis of 1H Magnetic Resonance Spectroscopic Imaging," *Journal of Magnetic Resonance Imaging*, Available at: http://dx.doi.org/10.1002/jmri.24487 Accessed January 24, 2014

44. Merdan, S., Womble, P. R., Miller, D. C., Barnett, C., Ye, Z., Linsell, S.M., Montie, J. E., Denton, B. T., "Toward Better Use of Bone Scans Among Men with Early-stage Prostate Cancer," *Urology* (in press)

45. Moyer, V. A. "Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement," *Annals of Internal Medicine*, 157(2), 120-134 (2012)

46. Nam, R. K., Toi, A., Klotz, L. H., Trachtenberg, J., Jewett, M. A., Appu, S., Loblaw, D. A., Sugar, L., Narod, S. A., and Kattan, M. W., "Assessing Individual Risk for Prostate Cancer," *Journal of Clinical Oncology*, 25(24) 3582-3588 (2007)

47. Parfait, S., Walker, P. M., Créhange, G., Tizon, X., and Mitéran, J., "Classification of Prostate Magnetic Resonance Spectra using Support Vector Machine," *Biomedical Signal Processing and Control*, 7(5), 499-508 (2012)

48. Penberthy, L., Retchin, S. M., McDonald, M. K., McClish, D. K., Desch, C. E., Riley, G. F., Smith, T. J., Hillner, B. E., and Newschaffer, C. J., "Predictors of Medicare Costs in Elderly Beneficiaries with Breast, Colorectal, Lung, or Prostate Cancer," *Health Care Management Science*, 2(3) 149-160 (1999)

49. Rabbani, F., Stroumbakis, N., Kava, B. R., Cookson, M. S., and Fair, W. R. "Incidence and Clinical Significance of False-Negative Sextant Prostate Biopsies," *The Journal of Urology*, 159(4), 1247-1250 (1998)

50. Risko, R., Selin, M., Womble, P. R., Barnett, C., Ye, . Z, Linsell, S. M., Montie, J. E., Miller, D. C., Denton, B. T., "Clinical Predictors and Recommendations for Staging CT Scan Among Men with Prostate Cancer," *Urology* (in press)

51. Ross, K.S., Carter, H.B., Pearson, J.D., and Guess, H.A., "Comparative Efficiency of Prostate-Specific Antigen Screening Strategies for Prostate Cancer Detection," *JAMA: The Journal of the American Medical Association*, 284(11), 1399-1405 (2000)

52. Schröder, F., and Kattan, M. W., "The Comparability of Models for Predicting the Risk of a Positive Prostate Biopsy with Prostate-Specific Antigen Alone: A Systematic Review," *European Urology*, 54(2), 274-290 (2008)

53. Schröder, F. H. et al.,"Screening and Prostate-Cancer Mortality in a Randomized European Study," *New England Journal of Medicine*, 360(13), 1320-1328 (2009)

ACCEPTED MANUSCRIPT

54. Seker, H., Odetayo, M. O., Petrovic, D., and Naguib, R. N. G., "A Fuzzy Logic Based-Method for Prognostic Decision Making in Breast and Prostate Cancers," *IEEE Transactions on Information Technology in Biomedicine*, 7(2), 114-122 (2003)

55. Seker, H., Odetayo, M. O., Petrovic, D., Naguib, R. N. G., and Hamdy, F. C., "An Intelligent Hybrid Neuro-Fuzzy Rule-Based System for Prognostic Decision Making in Prostate Cancer Patients," in *4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, 2003.* IEEE, 173-176 (2003)

56. Shariat, S. F., Karakiewicz, P. I., Suardi, N., and Kattan, M. W. "Comparison of Nomograms with Other Methods for Predicting Outcomes in Prostate Cancer: A Critical Analysis of the Literature," *Clinical Cancer Research*, 14(14), 4400-4407 (2008)

57. Simon, J., "Decision Making with Prostate Cancer: A Multiple-objective Model with Uncertainty," *Interfaces*, 39(3), 218-227 (2009)

58. Sofer, A., Zeng, J., and Mun, S. K., "Optimal Biopsy Protocols for Prostate Cancer," *Annals of Operations Research*, 119(1-4), 63-74 (2003)

59. Sooriakumaran, P., John, M., Christos, P., Bektic, J., Bartsch, G., Leung, R., Herman, M., Scherr, D., and Tewari, A. "Models to Predict Positive Prostate Biopsies using the Tyrol Screening Study," *Urology*, 78(4), 924-929 (2011)

60. Spurgeon, S. E., Hsieh, Y. C., Rivadinera, A., Beer, T. M., Mori, M., and Garzotto, M., "Classification and Regression Tree Analysis for the Prediction of Aggressive Prostate Cancer on Biopsy," *The Journal of Urology*, 175(3), 918-922 (2006)

61. Surveillance, Epidemiology, and End Results (SEER) Program Populations (1969-2012) (www.seer.cancer.gov/popdata), National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released March 2014

62. Thompson, I. M., Ankerst, D. P., Chi, C., Goodman, P. J., Tangen, C. M., Lucia, M. S., Feng, Z., Parnes, H. L., Coltman, C. A. Jr., "Assessing Prostate Cancer Risk: Results from the Prostate Cancer Prevention Trial," *Journal of the National Cancer Institute*, 98 (8), 529-534 (2006)

63. Thompson, J. E., Moses, D., Shnier, R., Brenner, P., Delprado, W., Ponsky, L., Pulbrook, M., Bohm, M., Haynes, A. M., Hayen, A., and Stricker, P. D. "Multiparametric Magnetic Resonance Imaging Guided Diagnostic Biopsy Detects Significant Prostate Cancer and Could Reduce Unnecessary Biopsies and Over Detection: A Prospective Study," forthcoming in *The Journal of Urology* (2014)

64. Tsodikov, A., Szabo, A., and Wegelin, J., "A Population Model of Prostate Cancer Incidence," *Statistics in Medicine*, 25(16), 2846-2866 (2006)

ACCEPTED MANUSCRIPT

65. Underwood, D.J., Zhang, J., Denton, B.T., Shah, N.D., and Inman, B.A., "Simulation Optimization of PSA-Threshold Based Prostate Cancer Screening Policies," *Health Care Management Science,* 15(4), 293-309 (2012)

66. Welch, H. G., Fisher, E. S., Gottlieb, D. J., and Barry, M. J. "Detection of Prostate Cancer via Biopsy in the Medicare-SEER Population during the PSA Era," *Journal of National Cancer Institute*, 99(18), 1395–1400 (2007)

67. Zhang, J., Denton, B. T., Balasubramanian, H., Shah, N. D., and Inman, B. A., "Optimization of PSA Screening Policies: A Comparison of the Patient and Societal Perspectives," *Medical Decision Making*, 32(2), 337-349 (2012)

68. Zhang, J., Denton, B. T., Balasubramanian, H., Shah, N. D., and Inman, B. A., "Optimization of Prostate Biopsy Referral Decisions," *Manufacturing & Service Operations Management,* 14(4), 529-547 (2012)

ACCEPTED MANUSCRIPT