

Integrating Machine Learning and Optimization Methods for Imaging of Patients with Prostate Cancer

Selin Merdan

SMERDAN@UMICH.EDU

*Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, Michigan, USA*

Khurshid Ghani

KGHANI@MED.UMICH.EDU

*Department of Urology
University of Michigan
Ann Arbor, Michigan, USA*

Brian Denton

BTDENTON@UMICH.EDU

*Department of Industrial and Operations Engineering
Department of Urology
University of Michigan
Ann Arbor, Michigan, USA*

Abstract

We combine predictive modeling techniques from machine learning and optimization methods to design coordinated imaging protocols for detection of metastatic cancer. Our approach considers different combinations of imaging tests to reduce imaging while also ensuring that the average risk of missing a metastatic cancer in the population does not exceed a desirable threshold. To account for the imperfect calibration of probability estimates obtained from predictive models, we formulate the decision problem of determining the optimal assignment of patients to imaging protocols as a robust mixed-integer program. Furthermore, we propose fast, easy-to-understand and clinically motivated approximation algorithms that can mitigate the effects of statistical error in predictions. We illustrate the practical performance of the proposed approximation algorithms and optimization models based on medical data collected by a large state-wide prostate cancer collaborative. The work presented in this article will help lay the groundwork to improve medical decision making by integrating machine learning and optimization in other disease areas.

1. Introduction

Multiple diagnostic imaging tests are routinely used for detection of cancer. However, despite the tremendous advances in imaging in recent years, difficulty remains in selecting tests for patients. This difficulty stems from a tradeoff between the benefits of an accurate diagnosis of the anticipated disease and harms and costs associated with the imaging tests themselves. It is therefore challenging to determine how to use imaging tests optimally. We study this problem in the context of prostate cancer (PCa); however, the models and methods we describe could apply equally well to many other forms of cancer. To optimize the decision making for PCa imaging, we combine optimization and predictive analytics methods into *robust optimization models* to design optimal imaging protocols that can ac-

count for errors in predictions. Given its clinical significance, our work generates important insights and findings for clinicians, health systems and other stakeholders seeking a satisfactory tradeoff between the benefits and harms of using imaging tests for detection of metastatic PCa.

PCa is the most common cancer among men (Hricak et al. (2007)). It is estimated that in 2018, 164,690 new cases of PCa would be diagnosed and 29,430 men would die of disease in the United States. For each of these newly-diagnosed cancer cases, clinical *staging* will be performed to determine the extent of the disease. The most significant health outcome to consider when determining the stage of PCa is whether the cancer has metastasized (i.e., spread to other parts of the body). Although metastatic PCa is still considered incurable, there are treatment options that can increase survival. Therefore, accurate staging is crucial for the clinical management of PCa changing from possible cure to alleviating symptoms and improving quality of life.

Conventional imaging tests for PCa staging include bone scan (BS) and computed tomography (CT scan) for detection of bone and lymph node metastases, respectively. However, not all men with newly-diagnosed PCa are at the same risk of harboring metastatic cancer. This is an important consideration because there are harms associated with both under- and over-imaging. Under-imaging results in patients' metastatic PCa going undetected. In such cases, patients are subjected to treatments, such as radical prostatectomy (surgical removal of the prostate), that is unlikely to be beneficial, and can lead to serious side effects and negative health outcomes due to delays in chemotherapy. Over-imaging causes potentially harmful radiation exposure (Prasad et al. (2004); Lin (2010)), anxiety for the patient, and false positive findings that lead to risky and painful follow-up procedures (i.e., bone biopsy). Not only do these imaging tests expose the patient to excess radiation, but they also increase financial and time burdens both on the patient and healthcare system.

To facilitate the optimal imaging of newly-diagnosed PCa, professional societies such as the National Comprehensive Cancer Network, American Urological Association and European Urological Association have established international evidence-based guidelines indicating the need for BS and CT scan only in patients with certain unfavorable risk factors; however, the guidelines vary in their recommendations (Mohler et al. (2010); Heidenreich et al. (2008); Greene et al. (2009)). Thus, there exists persistent variation in the utilization of these imaging tests among urologists, including unnecessary imaging in patients at low risk for metastatic disease and potentially incomplete staging of patients at high risk. In 2012, the American Urological Association highlighted the need to reduce imaging for low-risk PCa in the *Choosing Wisely* campaign, a multidisciplinary effort to reduce unnecessary imaging, decrease overuse of healthcare resources, and improve quality of care (see <http://www.choosingwisely.org/>).

In this work, we are motivated by the fact that there is no evidence-based imaging guideline addressing the need for both BS and CT scan in a holistic approach. Therefore, clinicians often order both imaging tests simultaneously or no tests. However, given the correlation observed between BS and CT scan results, the result of one imaging test can be used to predict the result of another follow-on test, which in turn, motivates a sequential imaging paradigm in which some patients may benefit from having the imaging tests one at a time. In a more general context, applicable also to disease areas other than PCa staging, we are concerned with the problem of optimal assignment of diagnostic testing protocols

that may combine multiple tests to more accurately and efficiently detect the presence of disease. This is an important problem as diagnostic resources are often expensive and limited, and poor decisions can lead to serious health outcomes, resulting in high healthcare costs and a significant reduction in quality of life.

The appropriateness of testing is dependent on the likelihood that a patient has the suspected disease, which in turn may depend on a number of clinical and demographic factors. Hence, we study the problem from a perspective in which individualized patient probability estimates for the presence of metastatic disease are estimated using predictive models. Consistent with the incentive of published imaging guidelines to reduce the overuse of imaging, the objective of the optimization models we develop is to reduce the total number of imaging tests performed at the population-level, subject to a certain budget level. In this context, the budget represents the maximum acceptable rate of missed metastatic disease in the population. The significant impact of the preoperative detection of metastases on the selection of appropriate treatment, quality of life and survival underscores the importance of incorporating the missed disease rate as a constraint into our mathematical formulations.

To our knowledge, we are the first to integrate robust optimization models with predictive models to optimize diagnostic testing decisions. We formulate our models using medical data from a large state-wide prostate cancer collaborative. These models are used to address the lack of a standardized holistic approach for recommending imaging tests on the basis of individuals' risk of disease while accounting for errors in predictions. In addition to exact methods, we propose approximation algorithms that incorporate the perspectives of multiple stakeholders participating in the decision making process for imaging and that lead to more predictable decisions than solving an optimization model. Finally, we summarize the benefits of using our approach to optimize multi-modality imaging for PCa staging.

The remainder of this article is organized as follows. In Section 2.1, we introduce mathematical notation and formulations as well as approximation algorithms. In Section 2.2, we describe the methodological approach for development and validation of a multinomial logistic regression model, and the analytical approach to quantify statistical variation in the probability estimates obtained from predictive models. In Section 3, we present numerical results using real medical data. Finally, in Section 4, we highlight our main conclusions.

2. Methods

In a clinical context, the focus of predictive models is often on risk prediction rather than classification because medical decisions are often influenced by individual risk preferences. Therefore, in order for a predictive model to be useful in decision making, it must provide validated and accurate estimates of probabilities of specific health conditions or outcomes. In practice, it is always the case that a predictive model will have imperfect calibration. Several factors such as the challenges in data collection and management (incomplete, heterogeneous, incorrect, or inconsistent data), small sample size, existence of large numbers of candidate predictors and the increased uncertainty surrounding rare events contribute to the imperfect nature of predictive models. In our proposed framework, the predictions are used to inform the assignment of imaging protocols to patients on the basis of their estimated risk of disease. Therefore, it is important to immunize imaging decisions against the *statistical error* in calibration. For this purpose, we utilize robust optimization.

2.1. Model Formulations and Analysis

Each branch of the decision tree in Figure 1 represents an imaging protocol, indexed by p , $p = 1, \dots, 4$, and the circles represent the application of the tests, with random outcomes denoted by branches. We consider the assignment of *patient types* into ideal imaging protocols and assume N types of patients differentiated on the basis of clinical risk factors, indexed by $j = 1, \dots, N$. The most straightforward approach to define patient types is to use the risk factors that are associated with the presence of disease. Similar to treating a continuous variable as a dichotomous variable in statistical modeling, some established criterion or cutoff point can be used to create certain categories of risk factors that are clinically relevant. These categories can then be used to define patient types (in Section 2.2 we describe the risk types we used in the context of PCa staging). We let w_j denote the proportion of patient type j in the population.

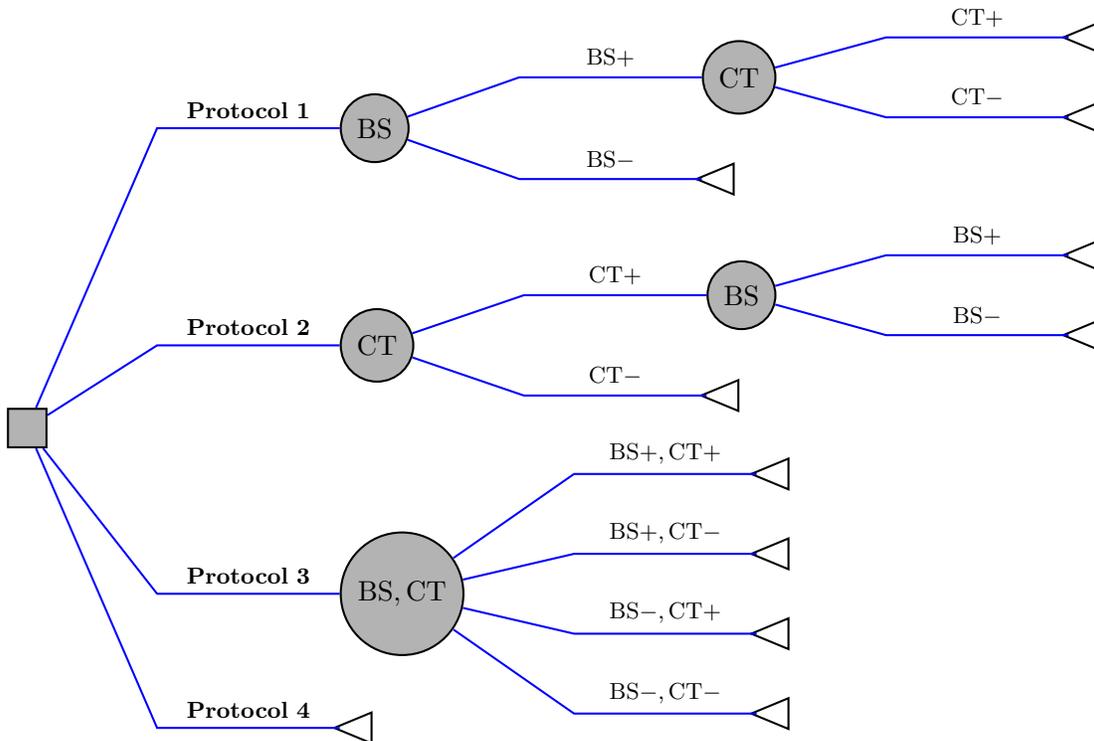


Figure 1: Decision tree for designing coordinated imaging protocols.

We let n_{jp} and m_{jp} denote the expected cost and the expected missed disease rate for patient type j under protocol p , respectively. We let α , $\alpha \in [0, 1]$, represent the maximum allowable rate of missed metastatic disease for the population determined by the decision maker, which we let refer to as the missed-rate budget. We introduce a binary variable x_{jp} defined as:

$$x_{jp} = \begin{cases} 1, & \text{if patient type } j \text{ is assigned to imaging protocol } p \\ 0, & \text{otherwise} \end{cases}$$

The decision problem of determining the optimal assignment of patient types to imaging protocols can be formulated as a *multiple-choice knapsack problem* (MCKP):

$$\max \left\{ \sum_{j=1}^N \sum_{p=1}^4 (\bar{n}_j - n_{jp}) x_{jp} \mid \sum_{j=1}^N \sum_{p=1}^4 m_{jp} x_{jp} \leq \alpha, \sum_{p=1}^4 x_{jp} = 1, \forall j, \text{ and } x_{jp} \in \{0, 1\}, \forall j, p \right\} \quad (\text{MIM})$$

which we refer to in our context as the *multiple imaging model* (MIM). This formulation is obtained by transforming the original minimization problem into the standard MCKP formulation by multiplying the objective by -1 , and adding the constant $\bar{n}_j = \max\{n_{jp} \mid p = 1, \dots, 4\}$ to all n_{jp} for patient type j . Note that we identify 4 alternative protocols for PCa staging, but the number and type of protocols would vary for other diseases.

The parameters n_{jp} and m_{jp} of MIM are determined based on the patient type-specific probabilities. To estimate the patient-type specific probabilities, we used the probability estimates for each patient in the study population obtained from the predictive models, and averaged them over patient types to obtain the mean predicted probabilities for each type (discussed in Section 2.2). We let $g_j(\cdot)$ denote the probability of an imaging outcome for patient type j . We let c_1 and c_2 denote the costs of BS and CT scan, respectively. We refer to c_1 and c_2 as costs; however, they can be generalized to represent asymmetrical penalties for imaging tests on the basis of factors that differentiate imaging tests such as cost, side effects, or patient or physician preferences in different concepts. The expected cost of imaging is defined for patient type j as $n_{j1} = w_j(c_1 + c_2 g_j(\text{BS+}))$ under Protocol 1, $n_{j2} = w_j(c_2 + c_1 g_j(\text{CT+}))$ under Protocol 2, $n_{j3} = 2w_j(c_1 + c_2)$ under Protocol 3 and $n_{j4} = 0$ under Protocol 4. The expected missed disease rates are defined for patient type j as $m_{j1} = w_j g_j(\text{BS-}, \text{CT+})$ for Protocol 1, $m_{j2} = w_j g_j(\text{BS+}, \text{CT-})$ for Protocol 2, and $m_{j4} = w_j g_j(\text{BS+ or CT+}) = w_j(1 - g_j(\text{BS-}, \text{CT-}))$ for Protocol 4. Note that $m_{j3} = 0$ since the protocol performing both tests simultaneously has a zero missed rate by assumption.

We adopt the *greedy algorithm* developed for the MCKP (Kellerer et al. (2004)). In the greedy algorithm for MCKP, the concept of *dominance* is important in the solution of MCKP because several variables that will never be chosen in an optimal solution can be deleted a priori.

Definition 1 Given two protocols s and t for patient type j , protocol s dominates t if it results in a lower cost and lower missed disease rate than protocol t . More formally:

$$m_{js} \leq m_{jt} \quad \text{and} \quad n_{js} \leq n_{jt} \quad (1)$$

Definition 2 If three protocols r, s and t for patient type j with $m_{jr} < m_{js} < m_{jt}$ and $n_{jr} < n_{js} < n_{jt}$ satisfy:

$$\frac{n_{jt} - n_{js}}{m_{jt} - m_{js}} \geq \frac{n_{js} - n_{jr}}{m_{js} - m_{jr}} \quad (2)$$

then Protocol s is LP-dominated by Protocols r and t .

Sinha and Zoltners (1979) showed that dominated and LP-dominated protocols will never be chosen in the optimal solution to a MCKP and the decision variables associated with these protocols are equal to 0. Dominated and LP-dominated protocols can therefore

be eliminated a-priori for each patient type. We let R_j denote the set of *LP-extreme* protocols that are nondominated for type j . The size of set R_j is denoted by r_j . We assume the ordering $m_{j1} < m_{j2} < \dots < m_{jr_j}$ in R_j .

Algorithm 1: Greedy.

For each patient type j , derive R_j . The following indices refer to protocols in R_j with respect to the increasing order of $m_{j1} < m_{j2} < \dots < m_{jr_j}$.

Construct an instance of the binary knapsack problem by setting $\tilde{n}_{jp} = n_{j,p-1} - n_{jp}$ and $\tilde{m}_{jp} = m_{jp} - m_{j,p-1}$ for each R_j and $p = 2, \dots, r_j$. Each combination of patient type and protocol in this problem can be seen as 2-tuples of (j, p) .

Calculate the incremental efficiencies $\tilde{e}_{jp} = \tilde{n}_{jp}/\tilde{m}_{jp}$ for each of the 2-tuples and sort them according to decreasing \tilde{e}_{jp} . With each value of \tilde{e}_{jp} , we associate the original indices j, p during the sorting.

Set $x_{j1} = 1$ and $x_{jp} = 0$ for $p = 2, \dots, r_j$ for all j .

$z^G = 2$ z^G is the total imaging tests performed based on the current allocation

$\bar{m} = \alpha$ \bar{m} is the residual missed-rate budget

```

for  $\forall(j, p) \in \{\tilde{e}_{jp}\}$  do
    while  $\tilde{m}_{jp} \leq \bar{m}$  do
        Assign type  $j$  to Protocol  $p$ 
         $\bar{m} = \bar{m} - \tilde{m}_{jp}$ 
         $z^G = z^G + \tilde{n}_{jp}$ 
         $x_{jp} = 1, x_{j,p-1} = 0$ 
    end
end

```

end

return The solution (x_1, \dots, x_N) with value z^G .

In addition to the greedy algorithm, we propose a *patient-centered greedy algorithm* that assigns patient types to protocols on the basis of their estimated probability of missed metastatic disease. Each protocol is considered sequentially until one is found for which the probability of missed disease for the patient type falls below the budget α . If the probability of missed disease for a patient type under Protocol 4 is above the budget, we check the probability of missed disease for Protocols 1 and 2. If both Protocols 1 and 2 result in a probability of missed disease below the budget, we assign the type to the protocol with the lowest missed-rate. It is guaranteed that the greedy solution is feasible to **MIM** given the properties of w_j , i.e., $w_j \in [0, 1]$ for all j and $\sum_{j=1}^N w_j = 1$.

Scalability. The proposed framework can be extended to the cases of more than two tests. Given t number of tests, a collection of sequential protocols can be generated based on the 2^t possible subsets of protocols and the alternative sequences of protocols for each subset. The challenge as t grows is that there is an exponentially increasing number of outcomes to be predicted.

The Robust Model. Robust optimization has emerged as a powerful modeling tool to handle erroneous or noisy data in decision-making over the last decade because of its computational tractability and practicability. For a detailed overview, see [Ben-Tal and Nemirovski \(2008\)](#); [Ben-Tal et al. \(2009\)](#) and [Bertsimas et al. \(2011\)](#), and the references therein. In contrast to stochastic optimization where the goal is to optimize an expectation,

the goal in the robust optimization framework is to find a solution that is feasible for any realization of the uncertainty in a given uncertainty set. In other words, it optimizes against the worst-case instances using a min-max objective. This is sometimes viewed as a competition between a decision maker who wishes to minimize the objective *v.s.* an *adversary* who can control the uncertain model parameters to cause harm by maximizing the objective subject to constraints on the variation in model parameters. Given the practicability and tractability of the robust counterpart optimization approach in combinatorial problems that are subject to uncertainty, we adopt the uncertainty set proposed by [Bertsimas and Sim \(2003\)](#) in our mathematical formulations for the robust optimal design of imaging protocols.

The parameters n_{jp} and m_{jp} of [MIM](#) are determined based on the probability estimates obtained from predictive models, which in turn are both affected by statistical errors. For the uncertainty of the objective, we assume each uncertain \tilde{n}_{jp} of type j and protocol p takes values in $[\bar{n}_j - n_{jp}, \bar{n}_j - n_{jp} + \sigma_{jp}]$, where σ_{jp} ($\sigma_{jp} \geq 0$) represents the maximum deviation from the nominal value n_{jp} . For the uncertainty of the missed-rate budget constraint, we assume that the missed rates \tilde{m}_j are independently distributed and follow symmetric distributions in $[m_j - \delta_j, m_j + \delta_j]$, where δ_j ($\delta_j \geq 0$) represents the maximum deviation from the nominal value m_j . The robust counterpart formulation of [MIM](#) is as follows:

$$\begin{aligned}
 & \text{maximize } \sum_{j=1}^N \sum_{p=1}^4 (\bar{n}_j - n_{jp}) x_{jp} + \min_{\{S_0 \mid S_0 \subseteq J_0, |S_0| \leq \Gamma_0\}} \left\{ \sum_{(j,p) \in S_0} \sigma_{jp} x_{jp} \right\} \\
 & \text{subject to} \\
 & \sum_{j=1}^N \sum_{p=1}^4 m_{jp} x_{jp} + \max_{\substack{\{S_1 \cup \{k,l\} \mid S_1 \subseteq J_1, \\ |S_1| \leq \lfloor \Gamma_1 \rfloor, \{k,l\} \in J_1 \setminus S_1\}}} \left\{ \sum_{\{j,p\} \in S_1} \delta_{jp} x_{jp} + (\Gamma_1 - \lfloor \Gamma_1 \rfloor) \delta_{kl} x_{kl} \right\} \leq \alpha \quad (3) \\
 & \sum_{p=1}^4 x_{jp} = 1, \quad \forall j
 \end{aligned}$$

where J_0 and J_1 are the sets of coefficients of the objective and missed-rate budget constraint, respectively, that are subject to uncertainty, $J_0, J_1 \subseteq \{(j, p) \mid j \in \{1, \dots, N\} \text{ and } p \in \{1, \dots, 4\}\}$. The parameters Γ_0 and Γ_1 are used to control the level of robustness in the objective and the missed-rate budget constraint due to the uncertainty in the model parameters, respectively. For example, if $\Gamma_1 = 0$, the coefficients \tilde{m}_{jp} equal to their nominal values m_{jp} , which implies that there is no protection against uncertainty. If $\Gamma_1 = |J_1|$, the missed-rate budget constraint is fully protected against uncertainty yielding a very conservative solution. Thus, as Γ_1 increases, more protection is given and the solution is more robust to uncertainty. The robust counterpart formulation in (3) has the following equivalent mixed-integer program formulation:

$$\begin{aligned}
 & \text{maximize} && \sum_{j=1}^N \sum_{p=1}^4 (\bar{n}_j - n_{jp}) x_{jp} + t_0 \mathbf{\Gamma}_0 + \sum_{(j,p) \in J_0} u_{jp} \\
 & \text{subject to} && \sum_{j=1}^N \sum_{p=1}^4 m_{jp} x_{jp} + t_1 \mathbf{\Gamma}_1 + \sum_{(j,p) \in J_1} v_{jp} \leq \alpha \\
 & && \sum_{p=1}^4 x_{jp} = 1, \quad \forall j \\
 & && t_0 + u_{jp} \geq \sigma_{jp} x_{jp}, \quad \forall (j,p) \in J_0 \\
 & && t_1 + v_{jp} \geq \delta_{jp} x_{jp}, \quad \forall (j,p) \in J_1 \\
 & && t_0 \geq 0, \quad t_1 \geq 0 \\
 & && u_{jp} \geq 0, \quad \forall (j,p) \in J_0 \\
 & && v_{jp} \geq 0, \quad \forall (j,p) \in J_1 \\
 & && x_{jp} \in \{0, 1\}, \quad \forall (j,p)
 \end{aligned} \tag{R-MIM}$$

which we refer to as **R-MIM**. The variables u_{jp}, v_{jp} and t_0, t_1 of **R-MIM** correspond to the dual variables of the linearized constraints in (3) (Bertsimas and Sim (2003)).

2.2. Predictive Modeling

The robust optimization models described in the previous section are not limited to any one type of predictive model. Thus, we provide an example based on logistic regression (LR). LR is the most commonly used predictive modeling method in the biomedical literature. To predict the positive outcome of BS and CT scan, we utilize the binary LR models that were developed and validated in our previous work (Merdan et al. (2017)). In this section, we describe how we utilize the LR method to predict the probabilities of nominal imaging outcomes, and how we measure the uncertainty in predictions obtained from these models.

We develop a multinomial LR model to calculate the probabilities of joint outcomes of BS and CT scan. Suppose that n patients received both BS and CT scan, and we are given the empirical training data $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n) \in \mathbb{R}^d \times \{1, 2, 3, 4\}$ of those patients, where y is the categorical dependent variable of which the categories result from the concurrent application of BS and CT scan under Protocol 3. We assume that the categories of y are coded 1, 2, 3 or 4: category 1 corresponds to BS + and CT+, category 2 corresponds to BS + and CT-, category 3 corresponds to BS - and CT+, and category 4 corresponds to BS - and CT-. We fit three independent binary LR models, in which the last outcome is chosen to be the baseline outcome and the other three outcomes are separately regressed against the baseline outcome. We estimate binary LR models for $k < 4$ as follows:

$$\log \left(\frac{\mathbb{P}(y = k \mid \mathbf{x})}{\mathbb{P}(y = 4 \mid \mathbf{x})} \right) = \boldsymbol{\beta}_k^T \mathbf{x} \tag{4}$$

where $\boldsymbol{\beta}_k$ represents a set of regression coefficients for each category k with respect to the reference category 4. Exponentiating both sides of (4) and using the fact that all four of the probabilities must sum to one, we have:

$$\mathbb{P}(y = 4 \mid \mathbf{x}) = \frac{1}{1 + \sum_{l=1}^3 e^{\boldsymbol{\beta}_l^T \mathbf{x}}} \quad \text{and} \quad \mathbb{P}(y = k \mid \mathbf{x}) = \frac{e^{\boldsymbol{\beta}_k^T \mathbf{x}}}{1 + \sum_{l=1}^3 e^{\boldsymbol{\beta}_l^T \mathbf{x}}} \quad k < 4 \tag{5}$$

Because minimizing the conditional negative log-likelihood for an LR model is a convex optimization problem, we use the limited memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) method to find the maximum likelihood estimates of β_k for binary models in (5) for $k < 4$ (Byrd et al. (1995)).

Uncertainty in Predictions. To measure the uncertainty of parameters in MIM, we need to measure the uncertainty in the patient type-specific probabilities. Figure 2 displays the distribution of individual probability estimates obtained from an LR model predicting the positive outcome of BS for patients in a (a) low-risk and (b) high-risk type. The probability estimates for the low-risk type do not diverge significantly from zero (negative). For the high-risk type, the probability estimates exhibit significant variation.

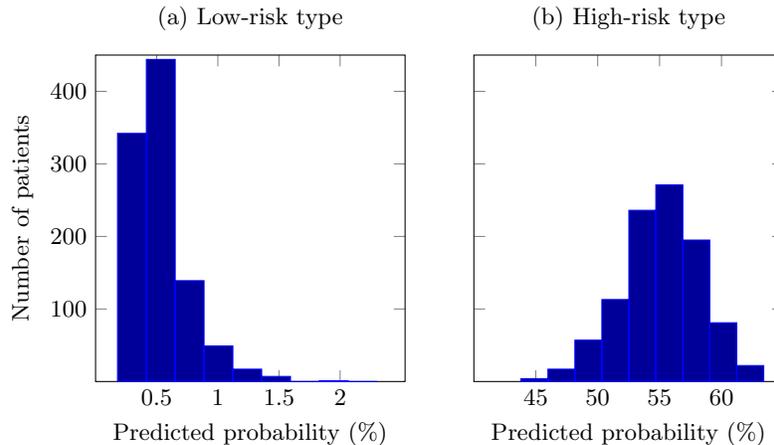


Figure 2: Distributions of individual probability estimates obtained from a LR model predicting the positive outcome of BS.

To measure the uncertainty in predictions, we employ random sampling of the coefficient vectors of the LR models based on the large-sample normal distributions of maximum likelihood estimates using the variances and covariances of the estimated coefficients of LR models. To illustrate the estimation of the variances and covariances, we consider a binary LR model for predicting the positive outcome of a BS: $\text{logit } \pi_i = \beta^T \mathbf{x}_i$ where the outcome is either $y_i = 1$ or $y_i = 0$ (1 corresponds to a positive test and 0 to a negative test) and $\pi_i = \mathbb{P}(y_i = 1 \mid \mathbf{x}_i)$ for patient i . The variances and covariances of the maximum likelihood estimates of β are obtained from the inverse of the so-called *observed information matrix*, denoted as $\mathbf{I}(\beta)$, i.e., $\text{Var}(\beta) = \mathbf{I}^{-1}(\beta)$. The estimators of the variances and covariances, denoted by $\widehat{\text{Var}}(\hat{\beta})$, are obtained by evaluating $\text{Var}(\beta)$ at the maximum likelihood estimate $\hat{\beta}$. The information matrix can be estimated as $\hat{\mathbf{I}}(\hat{\beta}) = \mathbf{X}^T \hat{\mathbf{V}} \mathbf{X}$, where \mathbf{X} is the data matrix and \mathbf{V} is a diagonal matrix defined as $\mathbf{V} = \text{diag}(\hat{\pi}_1(1 - \hat{\pi}_1), \dots, \hat{\pi}_n(1 - \hat{\pi}_n))$ (Hosmer et al. (2013)). In Appendix A, we describe the determination of R-MIM parameters based on the random sampling of coefficient vectors of the LR models.

3. Results

In this section, we present results for (1) the predictive models and (2) the robust optimization models. Model parameters were fit using data from a large PCa collaborative in the state of Michigan. Established in 2011 with funding from Blue Cross Blue Shield of Michigan, Michigan Urological Surgery Improvement Collaborative (MUSIC) is a consortium of 43 practices from throughout Michigan that aims to improve the quality and cost-efficiency of care provided to men with prostate cancer. Each practice involved in MUSIC obtained an exemption or approval for participation from a local institutional review board.

3.1. Predictive Model Results

Based on the univariate analyses conducted in our previous work, the following variables were found to have a statistically significant association with the presence of metastatic cancer: age at diagnosis, serum prostate-specific antigen (PSA) level, biopsy Gleason score, clinical T stage and the percentage of positive biopsy cores (Merdan et al. (2014); Risko et al. (2014)). PCa is diagnosed by biopsy, which involves extraction of tissue (normally 12 samples) from the prostate. These samples produce useful predictors of metastasis such as a pathology grading called Gleason score (GS) and percentage of positive samples (also called cores) that show cancer. A PSA test is a simple blood test that indicates the amount of PSA, a protein produced by cells of the prostate gland, that escapes into the blood from the prostate. Patients with higher than normal PSA values have a greater risk of metastatic PCa. Clinical T stage is part of the TNM staging system for PCa that defines the extent of the primary tumor based on clinical examination.

The study population included 938 newly-diagnosed PCa patients with complete preoperative data who received both BS and CT scan at diagnosis, of which 67 (7.1%) had both tests positive, 36 (3.8%) had BS positive but CT scan negative, 40 (4.3%) had BS negative and CT scan positive, and finally, 795 (84.8%) had both tests negative (see Appendix B for the clinical characteristics of the patients). We included the following covariates in the multinomial LR model: natural logarithm of PSA, biopsy GS ($\leq 3 + 4$, $4 + 3$, or $8 - 10$), clinical T stage (T1, T2, or T3/4) and the percentage of positive biopsy cores. Due to the high dispersion in PSA values, we used the natural logarithm transformation. We used a random sample of half of the data for training and other half for validation. Depictions of the mean predicted risk versus the true fraction of cases with $y = 1, 2$ and 3 along with the pairwise ROC curves for the binary models are shown in Appendix B. The results show good calibration in the validation samples. As expected, the binary model predicting $y = 1$ against $y = 4$ is good at discriminating patients who had both tests positive from patients who had both negative.

3.2. Optimization Model Results

To be consistent with the existing literature on knapsack problems, we proposed optimization models in maximization form in Section 2.1. For ease of interpretation, however, we present results in this section in the context of minimization of average number of imaging tests to be performed rather than the true objective value. MIM refers to the nominal model of which the model parameters are set to their average values based on randomly sampled

coefficient vectors of the LR models. Recall that **R-MIM** refers to the robust counterpart model. To define patient types, we used the clinical parameters that were highly associated with the positive outcome of BS and CT scan. We chose the most commonly used categories for each of these parameters. For PSA: ≤ 4 , $4 - 10$, $10 - 20$ and > 20 ; for GS: < 7 , $= 7$ and > 7 ; for clinical stage: T1, T2 and T3/4 were considered. Overall, we had 36 patient types. We assumed symmetrical cost penalties for BS and CT scan in our case study.

Figure 3 depicts the diminishing returns with respect to the increasing budget on missed disease rate for the optimal **MIM** and **R-MIM** solutions, and the greedy solutions. In this figure, **R-MIM** has the full protection level against statistical variation, i.e., $\Gamma_0 = |J_0| = 72$ and $\Gamma_1 = |J_1| = 108$. Both **MIM** and the greedy algorithm reduce the average number of imaging tests per patient significantly compared to the patient-centered greedy algorithm. For example, at a missed-rate budget of 1%, both the optimal **MIM** solution and the greedy solution reduce the average number of imaging tests per patient by 55.7% compared to the patient-centered greedy solution. Although the patient-centered greedy algorithm performs very poorly on the basis of mean imaging per patient for a given missed-rate budget, it guarantees all patient types have an actual missed-rate that falls below the missed-rate budget α . Thus, the difference between the optimal **MIM** solution and the patient-centered greedy solution can be viewed as the population benefit from “central planning”.

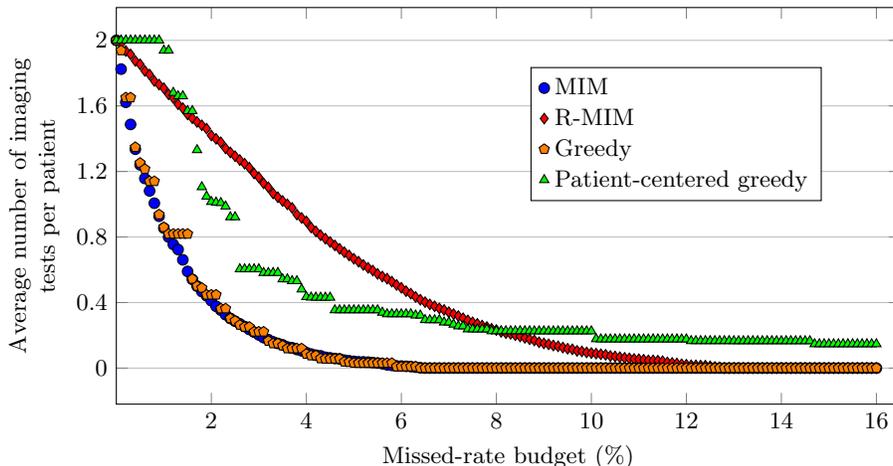


Figure 3: The average number of imaging tests per patient based on the optimal **MIM** and **R-MIM** solutions, and the greedy solutions as a function of the increasing missed-rate budget.

To evaluate the robustness of optimal **R-MIM** solutions, we considered the probability of missed-rate budget violation by randomly sampled coefficient vectors of the LR models. At a certain missed-rate budget, we constructed instances of **MIM** using 1,000 randomly selected coefficient vectors of the LR models. We estimated the probability of missed-rate budget violation as the fraction of **MIM** instances that were infeasible (resulting in a missed-rate higher than the budget α) by the original optimal solutions of **R-MIM** with various protection levels. Based on our sensitivity analyses, we found that the tradeoff between

the optimal average number of imaging tests per patient and the probability of missed-rate budget constraint violation was not affected by the changes in Γ_0 values, which controls the level of conservatism in the objective function of R-MIM (data not shown). As illustrated in Figure 4a, because the patient types at high risk of disease constitute a small portion of the population but are associated with high deviations in the probability estimates for positive imaging tests, it neutralizes the impact of Γ_0 on the robustness of the optimal R-MIM solution. Figure 4b demonstrates that the tradeoff between the robustness and optimality of coordinated imaging largely depends on the protection level Γ_1 . Moreover, there are choices of Γ_1 for which the optimal R-MIM solution provides substantial protection against missed-rate budget violation without greatly sacrificing on the reduction in imaging.

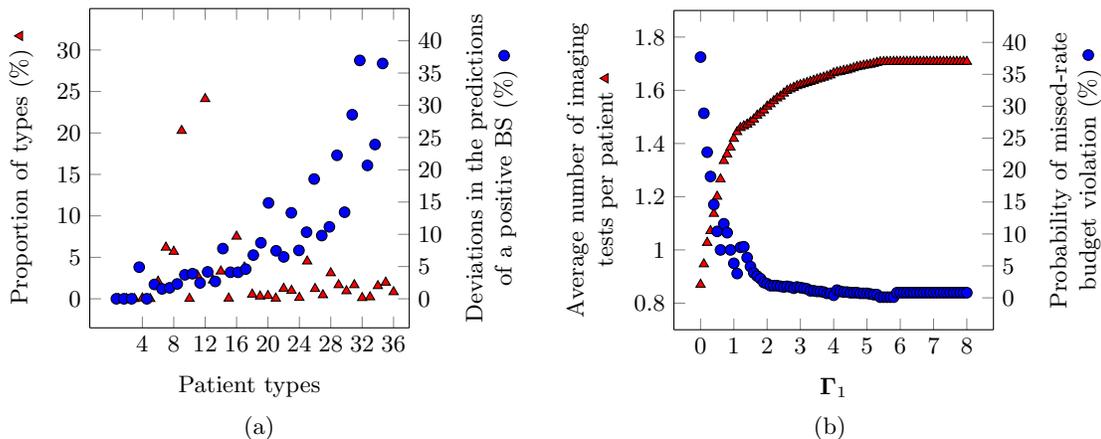


Figure 4: (a) The relation between the proportion of patient types and the variations in the estimated probability of positive BS. The patient types are sorted in the order of increasing risk of disease. (b) The tradeoff between the average number of imaging tests per patient and the probability of missed-rate budget violation of R-MIM as a function of Γ_1 at a missed-rate budget of 1%.

Table 1 presents a sample of the optimal average number of imaging tests per patient and the probability of missed-rate budget violation for R-MIM. The budget levels of 1% and 2% were chosen based on the clinical goal of having a very low miss disease rate for the population. At a missed-rate budget of 1%, the optimal MIM solution and the greedy solution exhibit similar performance in terms of the probability of missed-rate budget violation and the average number of imaging tests performed per patient. At a 2% missed-rate budget, the greedy solution, however, reduces the probability of missed-rate budget violation by 26.7% while increasing the average number of imaging tests per patient by 12.7% compared to the optimal MIM solution.

We also investigated benefits of coordinated imaging over optimizing imaging decisions independently for BS and CT scan. For each imaging test, in the context of independent imaging, we introduce a binary variable x_j : 1 if the patient type j is assigned to imaging protocol or 0 otherwise. Considering the case for BS, the expected missed-rate is defined as $m_j = w_j g_j(\text{BS}+)$ and the expected number of BSs performed is defined as $n_j = w_j$ for

Table 1: Comparison of imaging solutions at missed-rate budgets of 1% and 2%.

	Probability of missed-rate budget violation (%)	Average no. of imaging tests per patient	Probability of missed-rate budget violation (%)	Average no. of imaging tests per patient
	$\alpha = 1\%$		$\alpha = 2\%$	
MIM-Greedy	37.09 (36.60 – 37.57)	0.854 (0.851 – 0.857)	30.76 (29.87 – 31.64)	0.461 (0.456 – 0.466)
Patient-centered greedy	0.0	1.938	1.85 (1.70 – 2.00)	1.017 (1.014 – 1.020)
R-MIM				
$\Gamma_1 = 0.0$	38.62 (38.22 – 39.03)	0.861 (0.857 – 0.864)	41.95 (41.54 – 42.36)	0.409 (0.406 – 0.413)
$\Gamma_1 = 0.2$	24.52 (23.16 – 25.88)	0.982 (0.967 – 0.997)	31.28 (30.28 – 32.27)	0.467 (0.461 – 0.472)
$\Gamma_1 = 0.5$	14.52 (13.23 – 15.80)	1.142 (1.115 – 1.169)	18.39 (16.69 – 20.09)	0.571 (0.547 – 0.595)
$\Gamma_1 = 0.7$	9.98 (8.75 – 11.20)	1.237 (1.208 – 1.266)	13.67 (12.17 – 15.16)	0.662 (0.630 – 0.694)
$\Gamma_1 = 2.0$	5.54 (4.77 – 6.31)	1.489 (1.474 – 1.505)	3.12 (2.80 – 3.44)	1.038 (1.008 – 1.068)
$\Gamma_1 = 6.0$	0.51 (0.41 – 0.61)	1.655 (1.640 – 1.670)	0.58 (0.47 – 0.69)	1.288 (1.256 – 1.319)

The numbers in the parentheses represent the 95% confidence intervals calculated based on the 30 independent samples of 1000 coefficient vectors of the LR models. The protection level Γ_0 is set to its maximum (i.e., $|J_0| = 72$) in **R-MIM**.

patient type j . The optimal assignment of patient types for imaging can be formulated as:

$$\min \left\{ \sum_{j=1}^N w_j x_j \mid \sum_{j=1}^N m_j (1 - x_j) \leq \alpha, x_j \in \{0, 1\}, \forall j \right\} \quad (\text{SIM})$$

which we refer to in our context as the *single imaging model* (**SIM**).

We solved **SIM** for both BS and CT scan, and used the optimal solutions from these models to create a solution to **MIM**. We applied the following rule in generating a solution to **MIM**: if type j is assigned to both BS and CT scan, then $x_{j3} = 1$; if type j is assigned to BS but not to CT scan, then $x_{j1} = 1$; if type j is assigned to CT scan but not to BS, then $x_{j2} = 1$; else $x_{j4} = 1$ for each patient type j . Next, we determined the average number of imaging tests performed per patient and the resulting missed-rate in the population using the solution we created to **MIM**. Figure 5a shows that the optimal **SIM** solutions for BS and CT scan, when evaluated in **MIM**, result in a lower average number of imaging tests per patient than the optimal **MIM** solution at varying missed-rate budgets. However, Figure 5b shows that the optimal **SIM** solutions, when evaluated in **MIM**, result in a higher missed disease rate in the population than the missed-rate budget, therefore, yielding infeasible solutions to **MIM**.

4. Discussion and Related Work

Motivated by the lack of a holistic clinical perspective that integrates imaging decisions for PCa staging, we combined predictive models from machine learning and optimization models into a robust optimization framework to design imaging guidelines that can account for imperfect calibration of predictions. We incorporated the perspectives of patients and physicians at the population level, and proposed models for sequential testing where the outcome of one imaging test informs the decision about the follow-up test. In addition to the optimization models, we proposed clinically motivated approximation algorithms and presented the results of the approximation methods that we showed can be used to solve large-scale models.

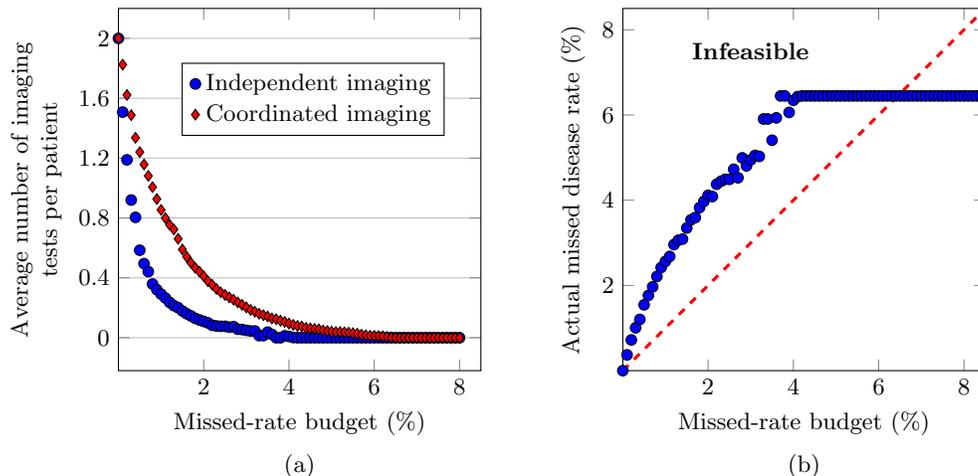


Figure 5: (a) The optimal independent imaging for BS and CT scan results in a lower imaging but (b) higher missed-rate than the budget, when evaluated in MIM.

The optimal selection of diagnostic tests for disease screening has been studied in the context of blood screening where the goal is to reduce the risk of transfusion-transmitted infectious diseases (TTIs), including the human immunodeficiency virus, hepatitis viruses B and C, and syphilis (Bish et al. (2011, 2014); Xie et al. (2012)). Earlier work addressing different aspects of this problem showed that optimized screening strategies result in a more effective and efficient screening for donated blood compared to the current screening strategies, without increasing resource requirement. Recently, El-Amine et al. (2017) expanded previous research on the optimization for blood screening by accounting for the uncertainty in the prevalence rates of TTIs and the limited information that the decision maker has.

Our work differs from the above literature on diagnostic testing decisions for blood screening in two main ways. First, we presented a new MCKP formulation to determine the optimal assignments of patient types into coordinated imaging protocols with the goal of minimizing imaging in the population while ensuring that the percentage of the population with missed disease is below a certain missed-rate budget predefined by the decision maker. In contrast to our objective, the previous work on blood screening aims to find the optimal test selection that achieves a low TTI risk. Second, our model involves the use of predictive models for estimating the probability of imaging outcomes based on patients' risk factors, and the individual probability estimates obtained from predictive models are used to define uncertainty sets for the model parameters that depend on predictions. The closest work that considers parameter uncertainty in this framework is that of El-Amine et al. (2017). They assume that the only information available to the decision maker is the support of the random prevalence rate vector of infectious diseases, and this information relies on the estimates reported in the literature in other studies.

Our case study on medical data from a state-wide prostate cancer collaborative demonstrated that the cost of robustness was high because of high statistical estimation error in predictions, rendering the incorporation of robust optimization models into clinical de-

cision making useful to trade off the protection against missed-rate budget violation with the number of imaging tests performed in the population. Furthermore, we showed that the coordinated imaging in PCa staging is more beneficial than the optimized single imaging. The coordinated imaging offers the potential to achieve better health outcomes in the population while reducing imaging tests performed. Hence, these models are particularly relevant for clinical decision making with implications for patients and physicians. To provide physicians with decision rules that are easy to interpret and implement, we considered the assignment of patient types into the imaging protocols (see Appendix C for an illustration of optimal decision rules). Our results show that optimizing at the population level (i.e., central planning) may differ significantly from optimizing for each patient type independently. This raises important questions about how to trade off between the different perspectives of patients and physicians.

The new robust optimization framework we present for diagnostic testing decisions provides a means to account for and mitigate the negative effects of errors in risk predictions. This model will help lay the foundation for future opportunities to combine machine learning and optimization for medical decision making.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. CMMI-1536444. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors would also like to acknowledge the significant contribution of Blue Cross Blue Shield of Michigan, which supports Michigan Urological Surgery Improvement Collaborative.

References

- Aharon Ben-Tal and Arkadi Nemirovski. Selected topics in robust convex optimization. *Mathematical Programming*, 112(1):125–158, 2008.
- Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust optimization*. Princeton University Press, 2009.
- Dimitris Bertsimas and Melvyn Sim. Robust discrete optimization and network flows. *Mathematical Programming*, 98(1-3):49–71, 2003.
- Dimitris Bertsimas, David B. Brown, and Constantine Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.
- Douglas R Bish, Ebru K Bish, Shiguang R Xie, and Anthony D Slonim. Optimal selection of screening assays for infectious agents in donated blood. *IIE Transactions on Healthcare Systems Engineering*, 1(2):67–90, 2011.
- Douglas R Bish, Ebru K Bish, Ryan S Xie, and Susan L Stramer. Going beyond “same-for-all” testing of infectious agents in donated blood. *IIE Transactions*, 46(11):1147–1168, 2014.

- Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.
- Hadi El-Amine, Ebru K Bish, and Douglas R Bish. Robust postdonation blood screening under prevalence rate uncertainty. *Operations Research*, 66(1):1–17, 2017.
- Kirsten L Greene, Peter C Albertsen, Richard J Babaian, H Ballentine Carter, Peter H Gann, Misop Han, Deborah Ann Kuban, A Oliver Sartor, Janet L Stanford, Anthony Zietman, and Peter Carroll. Prostate specific antigen best practice statement: 2009 update. *The Journal of Urology*, 182(5):2232 – 2241, 2009.
- Axel Heidenreich, Gunnar Aus, Michel Bolla, Steven Joniau, Vsevolod B Matveev, Hans Peter Schmid, and Filliberto Zattoni. EAU guidelines on prostate cancer. *European Urology*, 53(1):68 – 80, 2008.
- David W Hosmer, Stanley Lemeshow, and Rodney X Sturdivant. *Applied logistic regression*, volume 398. John Wiley & Sons, 2013.
- Hedvig Hricak, Peter L Choyke, Steven C Eberhardt, Steven A Leibel, and Peter T Scardino. Imaging prostate cancer: a multidisciplinary perspective. *Radiology*, 243(1):28–53, 2007.
- Hans Kellerer, Ulrich Pferschy, and David Pisinger. *Knapsack Problems*. Springer, Berlin, Germany, 2004.
- Eugene C Lin. Radiation risk from medical imaging. *Mayo Clinic Proceedings*, 85(12):1142–1146, 2010.
- Selin Merdan, Paul R Womble, David C Miller, Christine Barnett, Zhu Ye, Susan M Linsell, James E Montie, and Brian T Denton. Toward better use of bone scans among men with early-stage prostate cancer. *Urology*, 84(4):793–798, 2014.
- Selin Merdan, Christine Barnett, Brian Denton, James E Montie, and David C Miller. Data analytics for optimal detection of metastatic prostate cancer. *Working paper (Under review)*, 2017. URL <http://www-personal.umich.edu/~smerdan/#/pubs>.
- James Mohler, Robert R Bahnson, Barry Boston, J Erik Busby, Anthony D’Amico, James A Eastham, Charles A Enke, Daniel George, Eric Mark Horwitz, Robert P Huben, Philip Kantoff, Mark Kawachi, Michael Kuettel, Paul H Lange, Gary MacVicar, Elizabeth R Plimack, Julio M Pow-Sang, Mack Roach, Eric Rohren, Bruce J Roth, Dennis C Shrieve, Matthew R Smith, Sandy Srinivas, Przemyslaw Twardowski, and Patrick C Walsh. NCCN clinical practice guidelines in oncology: prostate cancer. *Journal of the National Comprehensive Cancer Network*, 8(2):162–200, 2010.
- Kedar N Prasad, William C Cole, and Gerald M Hasse. Health risks of low dose ionizing radiation in humans: a review. *Experimental Biology and Medicine*, 229(5):378–382, 2004.
- Rachel Risko, Selin Merdan, Paul R Womble, Christine Barnett, Zhu Ye, Susan M Linsell, James E Montie, David C Miller, and Brian T Denton. Clinical predictors and recommendations for staging computed tomography scan among men with prostate cancer. *Urology*, 84(6):1329–1334, 2014.

Prabhakant Sinha and Andris A Zoltners. The multiple-choice knapsack problem. *Operations Research*, 27(3):503–515, 1979.

Shiguang R Xie, Douglas R Bish, Ebru K Bish, Anthony D Slonim, and Susan L Stramer. Safety and waste considerations in donated blood screening. *European Journal of Operational Research*, 217(3):619–632, 2012.