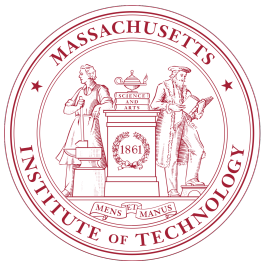

15.095

Machine Learning Under A Modern Optimization Lens

Deriving Treatment Policies for Prostate Cancer Patients using Optimal Policy Trees

Project Report

9^{th} December 2022

Moritz Barutsch

Rachit Jain

Contents

1 Introduction

Prostate cancer is the 2^{nd} most commonly occurring cancer in men and the 4^{th} most common cancer overall. The American Cancer Society [1] estimates $268,490$ new cases and about $34,500$ deaths in the US due to prostate cancer in 2022. Considering the widespread and severe nature of prostate cancer, especially in the US, there is a dire need of improving drug prescription policies. One approach to do so is by prescribing personalized treatment based on the characteristics for each patient.

2 Motivation

In their original paper, Byar & Green [2] postulated the idea of having an "optimal treatment for each patient based on the individual characteristics" in the application of clinical trials. Assigning the optimal treatment, if its exists, to patients based on their personal-level characteristics can go a long way in improving the healthcare environment for any country. With the onset of data-driven prescription strategies, there has been a big trend towards personalized medicine based not only on the diagnosis but also on the individual characteristics of the patient. This includes information such as demographics, genomics, and the patients disease history.

In addition, a preliminary exploratory data analysis on this dataset (in detail later) was explored. We found that the in the clinical trial that was conducted, the overall average treatment effect was close to zero as can be seen in Figure 1. Only the 1mg dose seemed to increase survival time, but that too slightly. Therefore, it seems that the treatment in general was rather ineffective. However, Bertsimas et. al. [3] showed that there exist subgroups of patients for which the different between best treatment and placebo is the largest. Motivated by this, we aim to derive an individual treatment assignment policy that assigns the treatment that is most effective for the respective patients (if there is one) to elongate their average survival time.

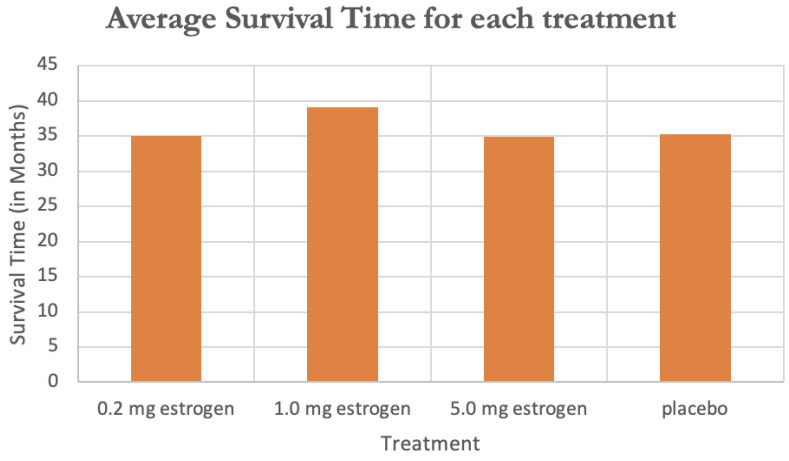

Figure 1: Months of survival by treatment

3 Problem Statement

Taking the idea of a potential optimal treatment further, this work puts an effort to making that original thought by Byar & Green into a data-driven reality for prostate cancer patients. The aim of this project is to create an optimal policy that personalizes the treatment assignment for patients.

4 Dataset

Diethylstilbestrol is a form of synthetic estrogen that has been used to treat late-stage prostate cancer. Byar and Green [2] explored the effect of a placebo and three different levels, 0.2mg, 1.0 mg and 5.0mg, of this drug involving estrogen, over 502 patients affected with stage 3 or 4 prostate cancer. It also contains 18 other patient-level features. These include, cancer stage, months of follow-up, age in years, blood pressure levels, weight index, history of cardiovascular disease, etc.

An important aspect of this project is the missing counterfactual values. For a patient, only one of the 4 treatments is given. Thus, the survival time for that patient for the other 3 treatments is missing and needs to be predicted before-hand. Although every confounding feature can never be taken into account, the dataset allows us to find some some high-level patterns.

Furthermore, dataset contains right-censored survival data. While some patients in the dataset died during the study period and therefore their exact time of survival is known, there are patients that were still alive at the end of the study period. For the these, only a lower bound for their survival time can be observed. To address this, the commonly used algorithms and models need to be adjusted.

5 Methodology

We take the following approach to derive a individualized treatment policy for prostate cancer based on our dataset. First, the data is split into training and testing data (see Section 5.1). After that, the best model to estimate the rewards under each treatment for each patient in the training set is determined (see Section 5.2). Based on the resulting reward matrix, an Optimal Policy can be trained (see Section 5.3). Similarly as for the training data, the best model to estimate the rewards matrix of the test data is determined using the test data only. Based on the resulting reward matrix, the fitted Optimal Policy Tree is used to predict the treatment for each of the patients in the testing data. Finally, the policy is evaluated by comparing it to the actually observed policy.

5.1 Train/Test split

As reward estimation is done separately for training and testing data to avoid information leakage, the commonly used data split of 70/30 is not appropriate for this problem. To ensure a high quality of reward estimates also on the testing set, a 50/50 train test split is chosen.

5.2 Reward Estimation

To prescribe the optimal treatment, i.e. the one maximizing the months of survival, we need to know what the months of survival would have been if we had prescribed each of the other treatments. As we are dealing with observational data we do not have these so called *counterfactuals* and, thus, have to rely on estimating them. To avoid any leakage of information between the data used for training and testing, this needs to be done separately for the training and testing data. It is important to note that this step is crucial as the quality of the prescription policy determined can only be as good as the reward estimation that preceded it. In general, there are three ways to estimate the rewards $\Gamma_i(t)$ for each patient i under each treatment t. We will shortly discuss each of these in the following.

5.2.1 Direct method

The most straightforward method to estimated the counterfactuals $\Gamma_i(t)$ is to train a separate reward prediction model $f_t(x_i)$ for each of the treatments t and then use these models to predict the rewards for all the patients i.

$$\Gamma_i(t) = f_t(x_i) \tag{1}$$

The reward estimation process can then be evaluated using Harrel's C-index [4], a commonly used metric in survival analysis. It can be calculated as follows:

$$\text{Harrel's C-index} = \frac{\text{number of } \textbf{concordant} \text{ pairs}}{\text{number of } \textbf{comparable} \text{ pairs}} \tag{2}$$

with a pair of patients being comparable if it is known that one patient died before the other. This can be the case if either the other patient died later or did not die and his/her survival time is higher. In addition, if the estimated reward is higher for the patient that died earlier, the pair is also said to be concordant, otherwise it is discordant.

The resulting index for the train and test set for two different types of Survival Learners implemented by Interpretable AI [5] can be seen in Table 1. We can observe that for both, the training as well as the testing data, the XGBoostSurvivalLearner performs best with a fairly good C-index between 70 to 80 % on the respective validation data from the train and test set. Note that the reward estimation procedures for the training and testing data is done separately, so that no data leakage from training to testing can occur.

Harrell's C-index	Train	Test
RandomForestSurvivalLearner	0.6440	0.5434
XGBoostSurvivalLearner	**0.7897**	**0.7128**

Table 1: Harrell's C-index for Random Forest and XGBoost Survival Learners

The rewards estimated using the direct method can be easily interpreted as the months of survival. However, the model used to predict the outcome for a certain treatment is trained exclusively on the patients that actually received that treatment and, thus, is biased by the initial treatment assignment.

5.2.2 Inverse propensity weighting

To address the treatment assignment bias observed when using the direct method, the rewards can be estimated based on the probability of a patient receiving a each treatment. Therefore, a classification model $p(x_i, t)$ is trained to predict the probability for patient i with features x to receive treatment t. This so called propensity score is defined by the conditional probability $P(T_i = t|X = x)$.

To estimate the rewards while accounting for assignment treatment bias, the actual months of survival y_i are divided by the predicted propensity score $p(x_i, t)$. This is called inverse propensity weighting and can be denoted as follows:

$$\Gamma_i(t) = \mathbb{1}(T_i = t) * \frac{y_i}{p(x_i, t)} \tag{3}$$

The reward estimation process using propensity scores can be evaluated looking at the misclassification error of $p(x_i, t)$ on the respective validation data from train and test set (see Table 2). We can observe that a RandomForestClassifier outperforms the XGBoostClassifier on the training and testing data. For both, Random Forest and XGBoost the misclassification error is lower in the testing data than in the training data.

Misclassification error	Train	Test
RandomForestClassifier	**0.2734**	**0.1982**
XGBoostClassifier	0.2869	0.2183

Table 2: Misclassification error for Random Forest and XGBoost Classifiers

The resulting rewards, however, can no longer be interpreted as months of survival as they have been adjusted with the estimated propensity score. In addition, having small changes in propensity scores can lead to big changes in rewards as we divide by $p(x_i, t)$.

5.2.3 Doubly robust

Finally, the aforementioned methods can be combined to into one leading to a reduction in treatment assignment and sensitivity to changes in propensity scores at the same time. The so called doubly robust reward estimation employs the direct method while adjusting for treatment assignment bias present in the data. The reward estimates are given by:

$$\Gamma_i(t) = f_t(x_i) + \mathbb{1}\{T_i = t\} * \frac{y_i - f_t(x_i)}{p(x_i, t)} \tag{4}$$

The accuracy of the doubly robust reward estimation is evaluated by looking at both: the Harrel's C-index of the outcome prediction model $f_t(x_i)$ and misclassification of the propensity model $p(x_i, t)$. The fairly good performance of both models (see Table 1 and Table 2) allows us to use the rewards resulting from the doubly robust estimation method to train and test an Optimal Policy Tree. Still, we have to keep in mind that the conclusions drawn in the results section are only as good as this reward estimation procedure (see Table 1 and 2).

5.3 Optimal policy tree

After the rewards have been estimated, a policy tree can be trained that maximizes the resulting outcomes under the assigned treatments over all patients. This can be written as:

$$\max_{\text{Tree}(.)} \sum_i \sum_t \mathbb{1}(\text{Tree}(x_i) = t) * \Gamma_{it} \tag{5}$$

To train the tree, we perform a grid search with different values for the max_depth and minbucket hyperparameters. Specifically we searched over a maximum tree depth of up to 7 and minbucket as 1%, 5% and 10% of the data.

5.4 Policy Evaluation

Evaluating policies derived from estimated rewards is challenging, as the estimated rewards can only be interpreted as the months of survival in certain cases. Using the most state-of-the-art methods, the rewards can only be interpreted in aggregation across all patients.

To evaluate the performance of policies, the average reward over all observations of the policy can be compared to the actual treatments that were given. In addition, as the rewards resulting from the direct method can be interpreted as survival months for each patient, the policy determined based on the direct method can also be evaluated by comparing it to an oracle that always prescribes the best treatment. Thus, the number of times the policy would have prescribed the best treatment is compared to how many times the best treatment was actually given.

5.5 Comparison with Optimal Subset Strategy

Taking it further, we combine a personalized treatment assignment policy with a subset selection strategy. In essence, we first find the optimal subset for a particular treatment. Afterwards, we compare the survival time of giving that specific treatment to all the patients in that subset with giving the 'optimal' treatment to each individual.

6 Results

6.1 Doubly Robust

At first, we trained a policy tree based on the rewards estimated using the doubly robust method. The resulting tree can be found in Figure 2. We can observe that for patients with lower hemoglobin levels the 1 mg treatment should be prescribed. Moreover, older patients seem to be harmed by a higher dose of estrogen while highest dose works best for younger patients. Evaluating this tree on the testing data resulted in a 14 month average increase in patient survival time. However, as described in section 5.2.3 it is hard to further interpret these results on a patient level as the doubly robust method does not allow to individually interpret the estimated rewards.

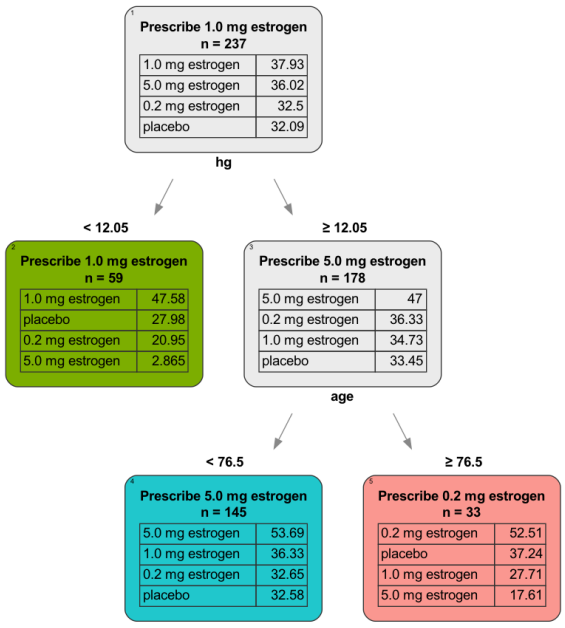

Figure 2: Optimal Policy Tree for All Patients (Doubly robust)

6.2 Direct Method

To be able to further explore these results on a patient level, we used the rewards estimated by the direct method to train Optimal Policy Trees. After training a global model for all patients in the dataset, we evaluate our performance on subsets of the data and again train Optimal Policy Trees for these.

6.2.1 Global model

At first, we trained a policy tree prescribing treatments to all patients. The resulting tree can be seen in Figure 3.

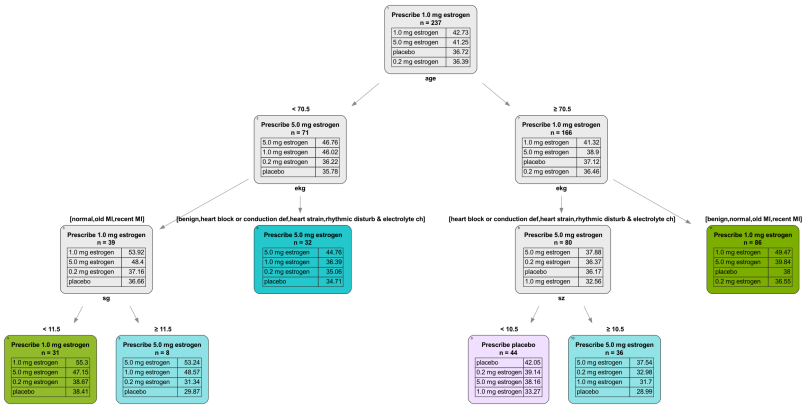

Figure 3: Optimal Policy Tree for All Patients (Direct method)

Evaluating its performance on the test data resulted in a mean reward of 40.85, while the actual treatment resulted in a mean reward of 39.71. However, the policy tree assigned the best treatment only for 37.39% of the patients in the test data while the best treatment according to the estimated reward was actually given to 21.85% of the treatments.

To interpret this result, an Optimal Classification Tree (OCT) was trained to interpret for which patients the policy would have prescribed the best treatment while a worse treatment was actually given. Therefore, class 0 is assigned to all patients for which the policy's prescription was "better" and class 1 to all other patients. The resulting tree can be seen in Figure 4.

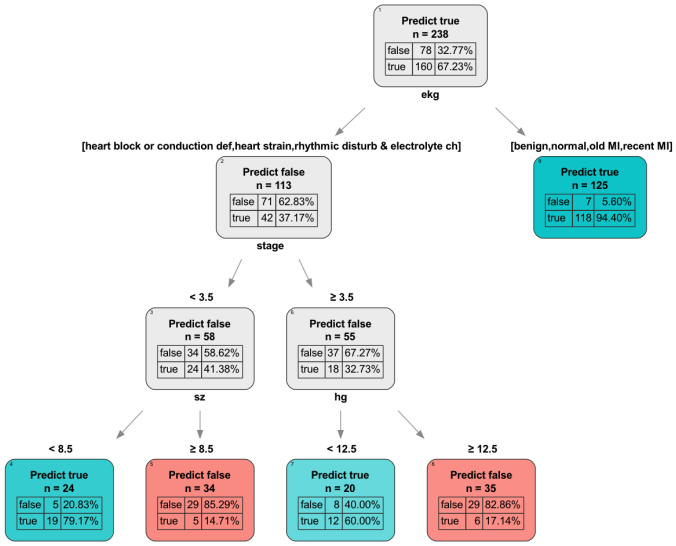

Figure 4: Optimal Classification Tree for comparison between policy and actual outcomes (= True, if policy tree assigned best treatment while it actually was not given

Although the overall results from the policy show an improvement of one months of survival over all the patients collectively, the OCT in Figure 4 shows that there are certain sub-sections of the dataset where the policy performs better than its counterpart. For instance, apart from an EKG separation, we can see that when the size of the rumor ('sz') is less than 8.5, the behaviour of our policy is very different from when it is more than 8.5.

6.2.2 Subset models

With motivation from the OCT shown, a separate policy tree was trained for patients with a higher tumor size than 8.5 and for those with lower. We cluster the patients into these two categories and find the resulting trees. These can be seen in Figure 10 and Figure 11.

For patients with larger tumor size, our intuition matches with the results from tree in Figure 10. These patients are better-off with a stronger treatment i.e. "5.0 mg estrogen" and "1.0 mg estrogen", thus the model assigns these drugs mostly. Infact, the placebo comes out to be the least effective in these cases which perfectly aligns with intuition. In contrast, patients with smaller tumor size are primarily given a moderate dosage of the drug or "placebo" as can be seen from Figure 11.

Quantifying with some metrics, for severe cases i.e. patient with higher size of tumor, the policy tree assigned the best treatment only for 36.55% patients while the best treatment according to the estimated reward was actually given to 22.76% of the treatments. This shows that the policy is giving a better treatment to significantly more patients that the actual treatments given. Similar results are obtained for less severe cases. Summarize result table is presented in Figure 5.

Alongside the tumor size criterion, we did similar analysis on different features like 'EKG', 'History of Cardiovascular Diseases', 'Status of Cancer', etc. The OPT on 'EKG' is shown in Figure **??**. This analysis was motivated from the different features present in the OCT in Figure 4.

		Survival (Months)			Percentage (%)	
		Policy	Actual	Oracle	Policy	Actual
Overall		40.85	39.71	48.11	37.39	21.85
SZ <= 8.5	SZ High (Individual OPT)	39.01	39.15	47.33	36.55	22.76
	SZ Low (Individual OPT)	43.88	40.60	49.32	39.78	20.43
	SZ High (Overall OPT)	39.01	39.15	47.33	36.03	22.06
	SZ Low (Overall OPT)	42.51	40.60	49.32	34.41	20.43

Figure 5: OPT Results for patients with low and high tumor size

6.3 Relating to Exceptional Responder Identification

Using the work [3] done for finding an optimal subset for '5.0 mg estrogen' treatment on this dataset, we find that patients that have no history of cardiovascular diseases, have stage 4 cancer and have diastolic blood pressure of greater than 70mmHg define the 'optimal subset', i.e. the subset with the highest average treatment effect.

Using the rewards achieved from the above techniques, we find the average treatment effect for all patients in the box if they were given the same treatment. For instance, if all those patients were given '5.0 mg estrogen', then their survival time (in months) would have been 4 months lesser than what our optimal policy on the entire dataset would prescribe. These results are presented in Figure 6. Training an OPT for that specific subset, could potentially further improve on this.

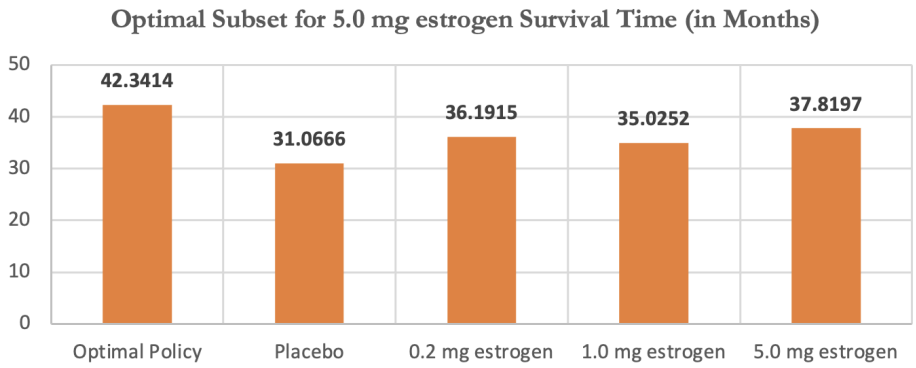

Figure 6: Survival Time when all patients in the optimal subset are given the same treatment vs given the policy treatment

7 Impact & Conclusion

Using the trained policy tree would have extended the survival time of the patients in the test set by 14 months on average. This policy can be used as a starting point for further research on the treatments used, as it offers insights into a potential treatment strategy that is individualized to both the patient and cancer.

Furthermore, Optimal Policy Trees can not be used to derive an individualized treatment policy if the treatment is ineffective for all patients. However, if the treatment is effective for all or a clearly defined subset of the patients, an optimal policy tree can be trained to interpretably prescribe treatments. Thus, in certain cases, it might be sensual to use exceptional responder identification (as done by Bertsimas et al. [3]) to identify a subset of patients responding to the treatment before applying optimal policy trees.

8 Contribution

Most of the work was done together, especially in the initial phases. We then tried to distribute the work amongst ourselves so that we could parallely work, run models, make evaluations, and then come back together every week to get the next steps aligned. Most contribution is seen in Figure 7.

CONTRIBUTION	
Rachit	**Moritz**
Brainstorming Problem Formulation	
Modeling Discussion	
Modeling Discussion	
Model Selection & Evaluation	
Report Structure	
Policy Tree Analysis	
Data EDA	Survival Analysis
Optimal Subset Comparison	Reward Estimation
Results [Report]	Methodology [Report]

Figure 7: Contribution

A Appendix

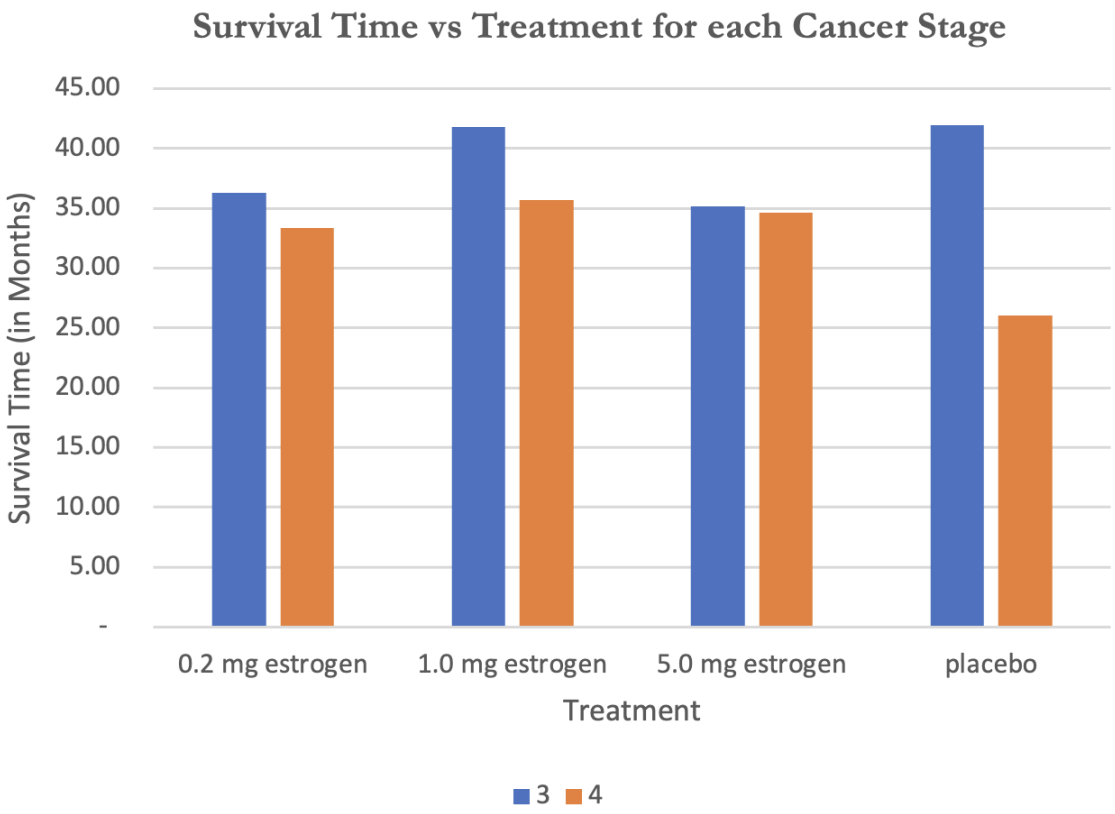

Figure 8: Survival Time for different treatments

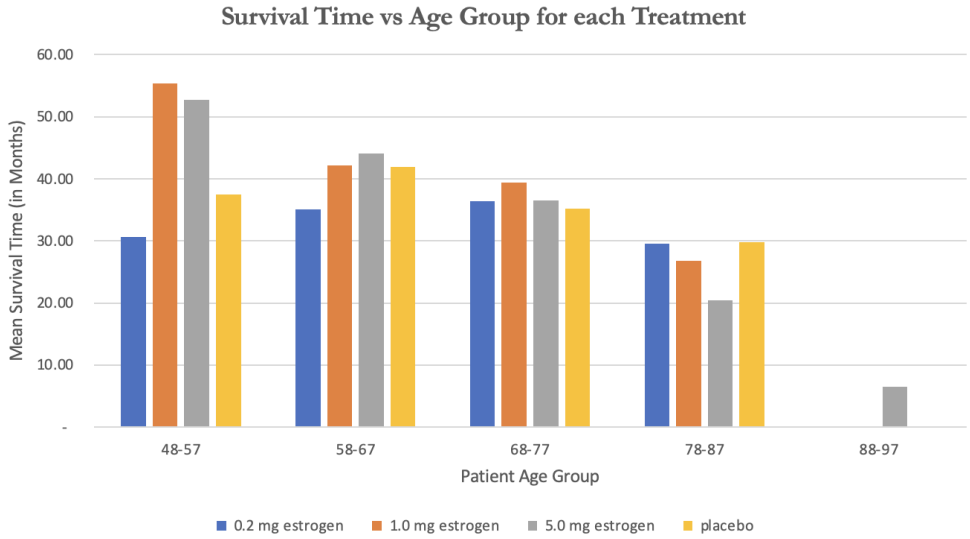

Figure 9: Treatmeant effect with respect to age group

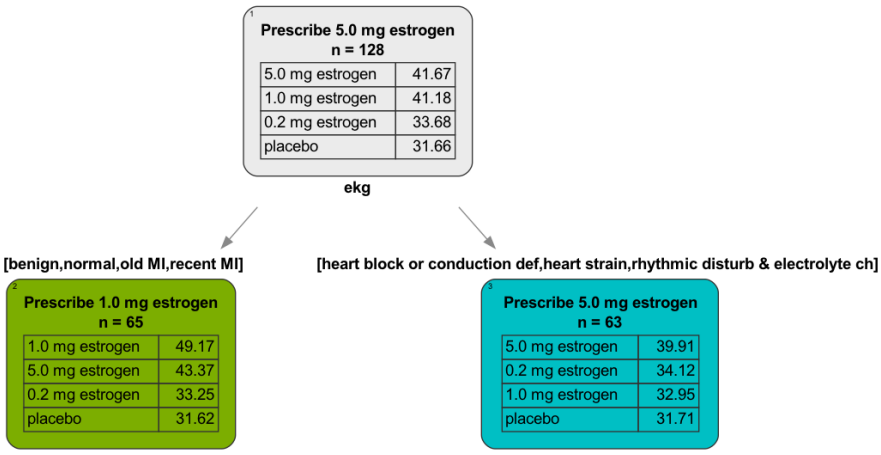

Figure 10: OPT on Patients with Tumor Size more than 8.5

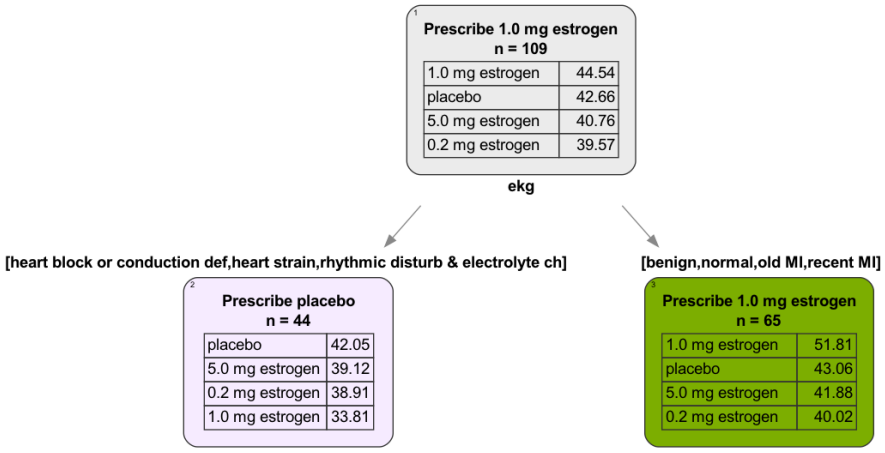

Figure 11: OPT on Patients with Tumor Size ≤ 8.5

References

[1] "Statistics for Prostate Cancer Patients."

[2] G. S. Byar DP, "The choice of treatment for cancer patients based on covariate information," *Bull Cancer*, vol. 67, no. 4, pp. 187–191, 1980.

[3] D. Bertsimas, N. Korolko, and A. M. Weinstein, "Identifying exceptional responders in randomized trials: An optimization approach," *INFORMS Journal on Optimization*, 2019.

[4] J. Harrell, Frank E., R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati, "Evaluating the Yield of Medical Tests," *JAMA*, vol. 247, pp. 2543–2546, 05 1982.

[5] "Interpretable AI."