Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Yuhan Ji	July 27, 2020

Risk Stratification for Overall Survival among Metastatic Prostate Cancer Patients treated by ADT as the first line treatment

By

Yuhan Ji
MSPH

Department of Biostatistics and Bioinformatics

Yuan Liu
Thesis Advisor

Robert Lyles
Reader

Risk Stratification for Overall Survival among Metastatic Prostate Cancer Patients treated by ADT as the first line treatment

By

Yuhan Ji

B.S., University of Minnesota – Twin Cities, 2018

Thesis Committee Chair: Yuan Liu, Doctor

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in
Biostatistics and Bioinformatics Department
2020

Abstract

Risk Stratification for Overall Survival among Metastatic Prostate Cancer Patients treated by ADT as the first line treatment

By Yuhan Ji

Background: Risk stratification is an important clinical process to divide patients into different groups based on their health signs and to carry right level of care, and there have already existed numerous standards to do risk stratification towards prostate cancer patients. In this study, we examine if nomogram, decision tree and regression coefficient-based scoring systems are useful as risk stratification tools towards prognostic assessment in metastatic prostate cancer patients.

Method: In this study, we implemented 3 alternative methods for risk stratification: nomogram, decision tree, three kinds of regression coefficient-based scoring system (Beta/Schneeweiss score, Beta/Sullivan score and HR/Charlson score). These 3 methods were conducted for an application on metastatic prostate cancer from NCDB dataset, where overall survival among patients were examined through Cox proportional hazard model. Prediction ability of 3 risk stratification methods were examined.

Conclusion: This study reveals that race, age, Charlson-Deyo Score, clinical stage, PSA, bone metastasis involvement, positive biopsy cores percentage are independent prognostic indicators for overall survival for metastatic prostate cancer patients. For risk stratification, nomogram was constructed with 1 and 2-year survival rate and had a C-index of 0.614, decision tree was not ideal with only age, Charlson-Deyo Score and positive biopsy cores percentage left in the model. Beta/Sullivan scoring system outperform other two regression coefficient-based scoring algorithms and had an average C-index of 0.595.

KEYWORD: Risk stratification, Nomogram, Decision tree, Regression coefficient-based scoring system

Risk Stratification for Overall Survival among Metastatic Prostate Cancer Patients treated by ADT as the first line treatment

By

Yuhan Ji

B.S., University of Minnesota – Twin Cities, 2018

Thesis Committee Chair: Yuan Liu, Doctor

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in
Biostatistics and Bioinformatics Department
2020

1. INTRODUCTION

Prostate cancer is known to be the most common cancer and ranked sixth in the cancer-related death amongst men in the world (Jemal et al., 2011). During the cancer progression, metastasis could happen when the cancer cells spread locally or regionally to other body parts like lymph nodes, organs, bones, etc. (NIH,2017). Most of the prostate cancer cases are diagnosed at early stage through screening, while there are still less than 5% patients diagnosed with metastasis condition (Sumi et al., 2017). There are various types of treatment for prostate cancer including chemotherapy, radiation, surgery and so on. Androgen-deprivation therapy (ADT) is one of the major treatments for prostate cancer since 1940s and this treatment helps patients to relieve their medical symptoms, decelerate tumor growth and increase overall survival (Christopher et al., 2015).

The purpose of this study was to investigate the association between several variables and survival outcome amongst 6274 patients with metastatic prostate cancer, and to do a risk stratification. Here the study population of interest were the male patients greater than 40 years old who used ADT as their initial treatment and started ADT within 90 days of diagnosis. Besides, patients who received radiation therapy or surgery after ADT treatment were excluded. The reason we wanted to explore this specific cohort was that there were clinical trials conducted to compare the treatment effects of “radiation + ADT ” vs. “ADT alone” (Lei et al., 2015) and “ADT + docetaxel” vs. “ADT alone” (Gravis et al., 2016), while they didn’t seem to have the precondition or restriction of using ADT as the initial and the only

treatment, they also allowed for some other treatments before ADT treatment whereas we requested ADT as the first-line treatment. For papers dealing with cancer survival prediction, univariate and multivariate survival analysis with Cox regression model were the common methods to establish the multivariable association between variants and survival outcome (Yang et al., 2019; Grivas et al., 2013). Another goal of the current study after the establishment of survival model was risk stratification for metastatic prostate cancer patients. Risk stratification was a very important clinical process to divide patients into different groups based on their health signs and to carry right level of care (NACHC, 2019). risk stratification was applied with various machine learning methods. Kruppa have written extensively about more than 20 machine learning methods to make risk prediction and the author also implied that if machine-learning algorithms were appropriately utilized and interpreted, it would be great in medical decision making (Kruppa et al., 2012), Thus, our further investigation goal of the current study including the stratification of patients according to their risk for disease, or to make risk predictions. To accomplish this, nomogram, decision tree and regression coefficient-based scoring system were applied as three algorithms of my own interest and compared prediction accuracies derived from these methods. Hopefully the results of this paper would be helpful in clinical decision making towards the prognosis and prediction of metastatic prostate cancer.

The structure of the paper was as follows: In the method and result section, basic demographic information of the dataset from NCDB as well as the exclusion criteria were summarized. Cox proportional hazard model was used to determine

the multivariate association of predictors and overall survival and variable selection was performed. Then, three machine learning methods were applied to do the risk stratification and results were compared. In the last section, we summarized our findings and discussed deficiencies and further research directions.

2. METHOD

2.1 Define study population

Data in this paper was derived from the National Cancer Database (NCDB). NCDB was a national oncology outcomes database where about 70% information of newly diagnosed cancer in the United States could be found (Bilimoria et al, 2007). In this study, the dataset of interest was NCDB prostate PUF data which contained prostate cancer cases diagnosed in 2004 – 2015. The selection and inclusion criteria were shown in Table 1. Since the biopsy Gleason score were included into the patients' profile after 2010, and this is an important variable related to prostate cancer, so only cases from 2010-2015 were included. By implementing these criteria, we reduced the original sample size from 1490799 to 6274.

Table 1. *Inclusion criteria for eligible patients*

Selection and Exclusion Criteria	Sample Size	Excluded
NCDB Prostate PUF Cancer Cases	1490799	-
Male with Age ≥ 40	1489806	993
Invasive tumor behavior	1489530	276

Selection and Exclusion Criteria	Sample Size	Excluded
Has only one or the first cancer diagnosis	1367820	121710
Include Metastasis cases	64208	1303612
Include cases treated by ADT as the first line treatment and started within 90 days after diagnosis	40972	23236
Exclude Palliative Care	34181	6791
Exclude patients got radiation or surgery	26349	7832
Delete missing values in T stage, PSA, and GS	15440	10909
Delete missing values in OS	12580	2860
Only include cases from 2010-2015	6274	6306

Outcome of interest was overall survival. Overall survival was defined by NCI as the duration of time from the diagnosis date or treatment start date of a certain disease to death from any cause. Patients who were still alive or lost in the record were considered censored. Then risk model was built to predict OS among metastatic prostate cancer patients. Baseline covariates included in this study were age at diagnosis, race, median income quartiles, year of diagnosis, facility type, facility location, Charlson-Deyo comorbidity score, grade group, biopsy Gleason score, PSA(prostate specific antigen) level, T stage and bone metastasis etc. Univariate association for multiple variables was examined, Hazard ratio, parametric and non-parametric p-values were reported. Then, multivariate analysis was performed with Cox proportional hazard model with all clinical

variables, and the final model was determined through stepwise selection. Finally, Nomogram, decision tree and regression-based scoring system were applied to implement risk stratification. All statistical analyses were performed using SAS 9.4 and R 3.6.1.

2.2 Nomogram

Nomogram was a very useful tool for oncology and medicine prognosis estimation, which illustrated through pictorial representation. As Bianco stated, nomogram was like a "graphic calculating device" for physicians. It allowed people easily calculate the rate or likelihood of a clinical event through a straightforward two-dimensional diagram (Bianco, 2006). As for the prognosis of metastatic prostate cancer patients in the current study, 1-year and 2-year survival rate were predicted by nomogram. Clinical covariates selected by Cox proportional hazard model were used in the construction of variable axes in the nomogram. Internal validation was then performed with bootstrap method to see if the model performance of nomogram was acceptable. Calibration estimated the resemblance of the nomogram estimated risk and observed risk, depicted through the calibration plot (Balachandran et al, 2015). The construction of nomogram was done by applying 'rms' package.

2.3 Decision tree

Decision tree had been commonly used as a classification tool since it was easy to understand and had a tree-like representation with branches which resembled human reasoning. Decision tree adaptations in survival analysis were referred as survival tree. It could be used in clinical decision support since it implicitly

performed variable screening and left with most important predictors. In the construction of decision tree, 'Nodes' were predefined variables in the Cox proportional hazard model and were used to control and construct the tree. The reasoning of the prediction started from the root node and ends in the leaf node. Pruning technique would be performed if it was necessary to remove sections of the tree that lacked power for classification and improved predictive accuracy (Bermejo et al, 2015). Due to large sample size, a test sample was used to select the best-sized tree with one third of patients randomly assigned to the testing sample, and two thirds of patients to the training sample with random seed of 2020. The construction of decision tree was done by 'rpart', 'rpart.plot' and 'partykit' packages.

2.4 Regression coefficient-based scoring system

Last method was regression coefficient-based scoring systems. Mehta et al. had found deficiencies in previously developed risk scores, they discussed the mathematical error as using risk ratio-based scoring system, because risk ratios multiplied when assigning weights for Charlson Comorbidity Score (CCS), whereas regression coefficient could be added. So, in this paper, we applied the scoring algorithms elaborated in Mehta's paper which were based on regression coefficients and used that to assign weights to different comorbidity scores. The goal here was to derive a few different forms of CCS and compare their performances in the prediction. Here, Beta/Schneeweiss scoring system, Beta/Sullivan scoring system and HR/Charlson scoring system were implemented (Mehta et al, 2016).

3. RESULT

3.1 Patient characteristics

A total of 6274 eligible patients were included, the demographic characteristics were shown in table 2. All the categorical covariates were described as percentages and frequencies. Age was categorized subjectively as <= 65 years and > 65 years. biopsy Gleason score was classified as 2-6%, 7%, 8-10%. T stage was classified as T1, T2, T3 and T4, etc.

Table 2. *The demographic characteristics of included patients.*

Variable	Level	N (%) = 6274
Age at Diagnosis	<=65	2603 (41.5)
	>65	3671 (58.5)
Race	White	4465 (71.2)
	Black	1176 (18.7)
	Other/Unknown	633 (10.1)
Median Income Quartiles 2008-2012	< $38,000	1223 (19.5)
	$38,000-$47,999	1455 (23.3)
	$48,000-$62,999	1704 (27.2)
	>=$63,000	1874 (30.0)
	Unknown	18
Percent of Patients Without High School Degree 2008-2012	>=21.0%	1176 (18.8)
	13.0-20.9%	1557 (24.9)
	7.0-12.9%	2001 (32.0)
	<7.0%	1525 (24.4)
	Unknown	15
Urban/Rural 2013	Metro	5012 (79.9)
	Urban	954 (15.2)
	Rural	176 (2.8)
	Unknown	132 (2.1)
Primary Payor	Other Government/Not Insured/Unk	987 (15.7)
	Private	1886 (30.1)
	Medicare	3401 (54.2)
Year of Diagnosis	2010-2012	2462 (39.2)
	2013-2015	3812 (60.8)
Facility Type	Non-Academic/Research Program	3302 (52.6)
	Academic/Research Program	2972 (47.4)

Variable	Level	N (%) = 6274
Facility Location	East	2578 (41.1)
	Central/Mountain	2998 (47.8)
	West	698 (11.1)
Charlson-Deyo Score	0	4953 (78.9)
	1	942 (15.0)
	2	255 (4.1)
	>=3	124 (2.0)
Grade	Well/Moderately Differentiated	477 (7.6)
	Poorly/Undifferentiated	5489 (87.5)
	Unknown	308 (4.9)
AJCC Clinical T	T1	2162 (34.5)
	T2	1882 (30.0)
	T3	1006 (16.0)
	T4	1224 (19.5)
PSA	<10	710 (11.3)
	10-20	709 (11.3)
	>20	4855 (77.4)
Gleason	2-6	72 (1.1)
	7	815 (13.0)
	8-10	5387 (85.9)
Risk Group	Low/Intermediate	165 (2.6)
	High	6109 (97.4)
Months of ADT Start from Diagnosis	<=0.20	1429 (22.8)
	>0.20, <=0.53	1604 (25.6)
	>0.53, <=1.02	1604 (25.6)
	>1.02	1637 (26.1)
Metastatic Bone Involvement (2010-2015)	No	616 (9.8)
	Yes	5658 (90.2)
Percent biopsy cores positive (2010-2015)	Mean	84.07
	Median	100.00
Biopsy Cores Positive (2010-2015)	<50%	553 (8.8)
	>=50%	5721 (91.2)

3.2 Univariate and multivariate association

Overall survival was defined as months from ADT started to death or date of last follow up. Hazards ratios that described the relative risk of the complication based on comparison of event rates were obtained. In Table 3, HR of one covariate level

were relative risk of that particular variable level compared to the reference level. Univariate analysis revealed that Charlson-Deyo Score > 0 ($p < 0.001$), poorly / undifferentiated grade ($p < 0.001$), clinical T3 stage ($p = 0.05$), clinical T4 stage ($p < 0.001$), PSA > 20 ($p < 0.001$), bone metastasis involvement ($p < 0.01$) were significantly associated with shorter OS. (Table 3). To look at HR for these variables specifically. For Charlson-Deyo Score, HR for Charlson-Deyo Score = 1 was 1.36 ($p < 0.001$), meaning that the rate of death in patients with Charlson-Deyo Score = 1 was 1.36 times the rate in group of patients with Charlson-Deyo Score = 0. HR for Charlson-Deyo Score = 2 was 1.79 ($p < 0.001$), so the rate of death in patients with Charlson-Deyo Score = 2 was 1.79 times the rate in group of patients with Charlson-Deyo Score = 0. HR for Charlson-Deyo Score >= 3 was 2.25 ($p < 0.001$), which means rate of death in patients with Charlson-Deyo Score >= 3 was 1.79 times the rate in group of patients with Charlson-Deyo Score = 0, etc.

Table 3. *Univariate association with overall survival*

Covariate	Level	Hazard Ratio (95% CI)	HR P-value	P-value
Facility Type	Non-Academic/Research Program	1.32 (1.23-1.41)	**<.001**	**<.001**
	Academic/Research Program	-	**-**	
Facility Location	East	1.08 (0.96-1.22)	0.192	**<.001**
	Central/Moutain	1.22 (1.08-1.37)	**0.001**	
	West	-	**-**	
Race	Black	1.00 (0.92-1.10)	0.950	**<.001**
	Other/Unknown	0.75 (0.66-0.85)	**<.001**	
	White	-	**-**	

Covariate	Level	Hazard Ratio (95% CI)	HR P-value	P-value
Primary Payor	Other Government/Not Insured/Unk	1.26 (1.13-1.41)	**<.001**	**<.001**
	Medicare	1.35 (1.24-1.46)	**<.001**	
	Private	-	**-**	
Median Income Quartiles 2008-2012	< $38,000	1.11 (1.00-1.23)	**0.044**	**0.035**
	$38,000-$47,999	1.14 (1.04-1.26)	**0.006**	
	$48,000-$62,999	1.09 (1.00-1.20)	0.058	
	>=$63,000	-	**-**	
Percent of Patients Without High School Degree 2008-2012	>=21.0%	1.17 (1.05-1.31)	**0.003**	**<.001**
	13.0-20.9%	1.22 (1.10-1.35)	**<.001**	
	7.0-12.9%	1.21 (1.10-1.33)	**<.001**	
	<7.0%	-	**-**	
Urban/Rural 2013	Urban	1.12 (1.02-1.23)	**0.016**	0.112
	Rural	1.07 (0.87-1.30)	0.524	
	Unknown	1.00 (0.78-1.28)	0.989	
	Metro	-	**-**	
Charlson-Deyo Score	>=3	2.25 (1.81-2.80)	**<.001**	**<.001**
	2	1.79 (1.53-2.09)	**<.001**	
	1	1.36 (1.24-1.49)	**<.001**	
	0	-	**-**	
Year of Diagnosis	2010-2012	1.09 (1.01-1.17)	**0.027**	**0.027**
	2013-2015	-	**-**	
Grade	Poorly/Undifferentiated	1.33 (1.13-1.55)	**<.001**	**<.001**
	Unknown	1.45 (1.18-1.79)	**<.001**	
	Well/Moderately Differentiated	-	**-**	

Covariate	Level	Hazard Ratio (95% CI)	HR P-value	P-value
AJCC Clinical T	T4	1.23 (1.12-1.35)	**<.001**	**<.001**
	T3	0.90 (0.81-1.00)	0.057	
	T2	0.98 (0.89-1.06)	0.574	
	T1	-	-	
PSA	>20	1.22 (1.09-1.37)	**<.001**	**<.001**
	10-20	0.91 (0.78-1.06)	0.207	
	<10	-	-	
Gleason	8-10	1.03 (0.74-1.44)	0.844	**<.001**
	7	0.77 (0.55-1.09)	0.144	
	2-6	-	-	
Risk Group	High	2.33 (1.76-3.09)	**<.001**	**<.001**
	Low/Intermediate	-	-	
Months of ADT Start from Diagnosis	<=0.20	1.69 (1.53-1.86)	**<.001**	**<.001**
	>0.20, <=0.53	1.27 (1.15-1.40)	**<.001**	
	>0.53, <=1.02	1.14 (1.03-1.26)	**0.011**	
	>1.02	-	-	
Biopsy Cores Positive (2010-2015)	<50%	0.73 (0.64-0.83)	**<.001**	**<.001**
	>=50%	z-	-	
Metastatic Bone Involvement, 2010-2015	Yes	1.39 (1.23-1.57)	**<.001**	**<.001**
	No	-	-	
Age at Diagnosis		1.02 (1.02-1.03)	**<.001**	**<.001**
Percent biopsy cores positive (2010-2015)		1.01 (1.00-1.01)	**<.001**	**<.001**

For multivariate association, Cox proportional hazard model was fitted. Dummy variables with reference cell coding were implemented. In this model, black race, Charlson-Deyo Score of 3, poorly/undifferentiated grade, clinical T1 stage, PSA <

10 and Gleason score of 2-6 were coded as the reference levels. Then the final model was determined through stepwise variables selection. All clinical covariates were significantly associated with OS, so the final model was as follows:

$$h(t) = h_0(t) \times \exp(\beta_1 \text{RaceOther} + \beta_2 \text{RaceWhite} + \beta_3 \text{CDScore0} + \beta_4 \text{CDScore1} + \beta_5 \text{CDScore2} + \beta_6 \text{GradeUnknown} + \beta_7 \text{GradeWell} + \beta_8 \text{CLT2} + \beta_9 \text{CLT3} + \beta_{10} \text{CLT4} + \beta_{11} \text{PSA} > 20 + \beta_{12} \text{PSA10-20} + \beta_{13} \text{Gleason7} + \beta_{14} \text{Gleason8-10} + \beta_{15} \text{BoneInvolve} + \beta_{16} \text{Age} + \beta_{17} \text{CorePositive})$$

Where t was the survival time, h(t) was the hazard function determined by covariates, and $h_0(t)$ was the baseline hazard corresponds to the value where all covariates were equal to 0.

Median OS from the start of ADT was 34.4 months [95% confidence interval (CI) 33.2-35.5]. Multivariate analysis revealed that other/unknown race ($p < 0.001$), Charlson-Deyo Score > 0 ($p < 0.001$), unknown grade ($p = 0.009$), clinical T3 stage ($p = 0.03$), clinical T4 stage ($p = 0.001$), PSA > 20 ($p = 0.006$), Gleason score = 7 ($p = 0.026$), bone metastasis involvement ($p < 0.001$), age ($p < 0.001$), positive biopsy cores percentage ($p < 0.001$) were independent prognostic indicators for OS (Table 4).

Table 4. *Multivariate association with overall survival*

Covariate	Level	Hazard Ratio (95% CI)	P-value
Race	Black	1.03 (0.94-1.13)	**0.47**
	Other/Unknown	0.74 (0.65-0.84)	**<.001**
	White	-	

Covariate	Level	Hazard Ratio (95% CI)	P-value
Charlson-Deyo Score	>=3	2.09 (1.68-2.60)	**<.001**
	2	1.61 (1.38-1.89)	**<.001**
	1	1.31 (1.19-1.44)	**<.001**
	0	-	
Grade	Poorly/Undifferentiated	1.18 (0.97-1.43)	**0.097**
	Unknown	1.36 (1.08-1.73)	**0.0097**
	Well/Moderately Differentiated	-	
AJCC Clinical T	T4	1.17 (1.06-1.29)	**0.001**
	T3	0.89 (0.80-0.99)	**0.03**
	T2	0.97 (0.89-1.06)	**0.543**
	T1	-	
PSA	>20	1.18 (1.05-1.32)	**0.006**
	10-20	0.90 (0.77-1.04)	**0.158**
	<10	-	
Gleason	8-10	0.81 (0.56-1.17)	**0.261**
	7	0.66 (0.46-0.95)	**0.026**
	2-6	-	
Metastatic Bone Involvement, 2010-2015	Yes	1.36 (1.20-1.54)	**<.001**
	No	-	
Age at Diagnosis		1.02 (1.02-1.03)	**<.001**
Percent biopsy cores positive (2010-2015)		1.01 (1.00-1.01)	**<.001**

3.3 Risk stratification

3.3.1 Nomogram construction

Nomogram was built with R package “rms”. As shown in Figure 1, the nomogram was built based on clinical variables that were same in the Cox proportional hazard model and constructed with 1 and 2-year overall survival. If we drew an upward vertical line to the “Points” bar in the first line, we could calculate points for each covariate, and based on the sum of all covariates, we could then draw another vertical line from the “Total Points” line to 1 year or 2 year survival line to get a specific survival rate. It was straightforward for a clinical decision making.

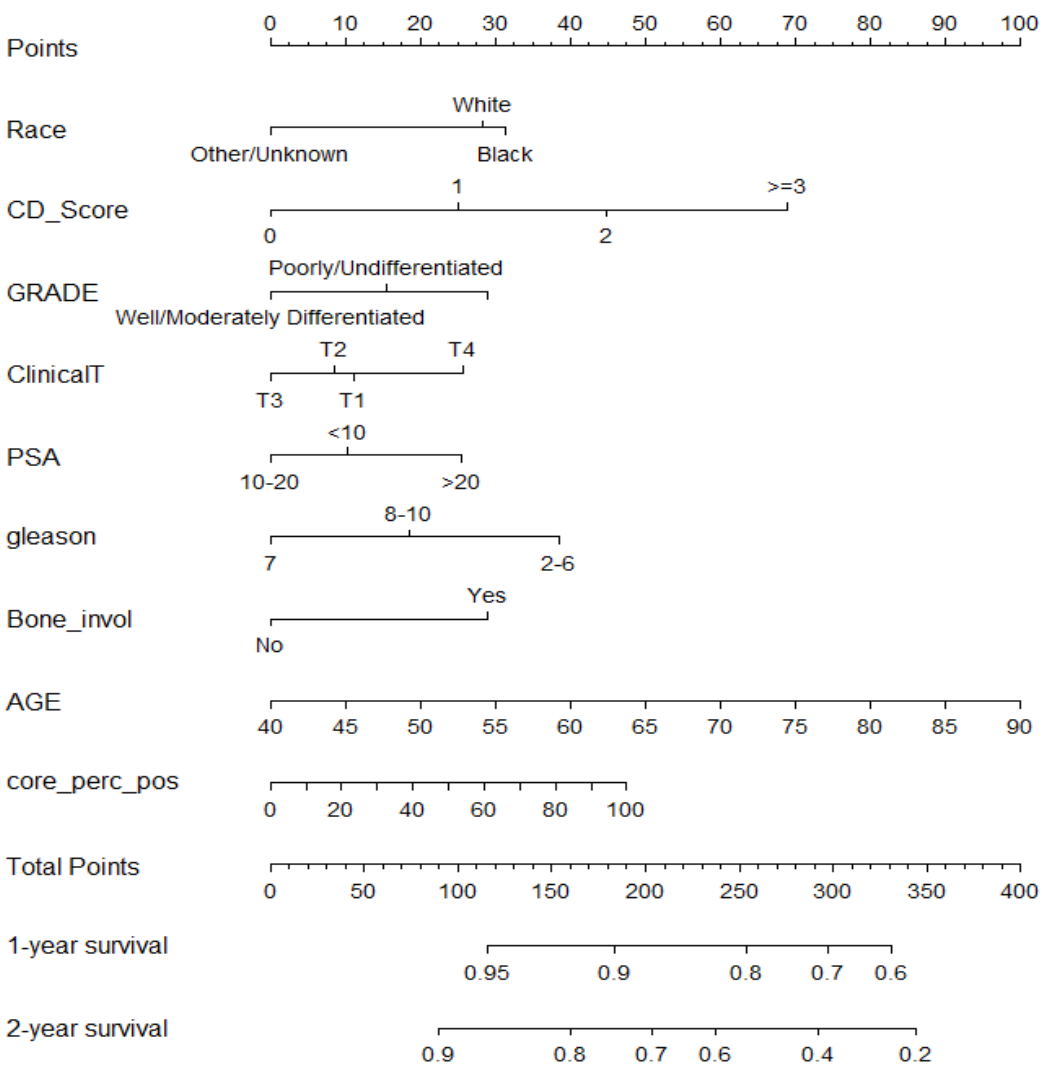

Figure 1. Nomogram for metastatic prostate cancer

3.3.2 Nomogram validation

The nomogram was then internally validated. Concordance index (C-index) was 0.614, C-index was used widely to validate the predictive ability of a survival model.

A higher C-index meant our model predicted higher probabilities of survival for higher observed survival times. Calibration curves for the probability of OS at 1-year and 2-year were shown in Figure 2 and Figure 3. The black dotted line was when nomogram survival outcomes equal to observed outcomes, which stood for the perfect calibration. The blue line connected by dots was the actual calibration. We could tell from Figure 2 on 1-year outcome that the result was not very precise when the survival rates were less than 0.80, and overall calibration for 2-year survival rate was better than 1-year outcome comparing two figures.

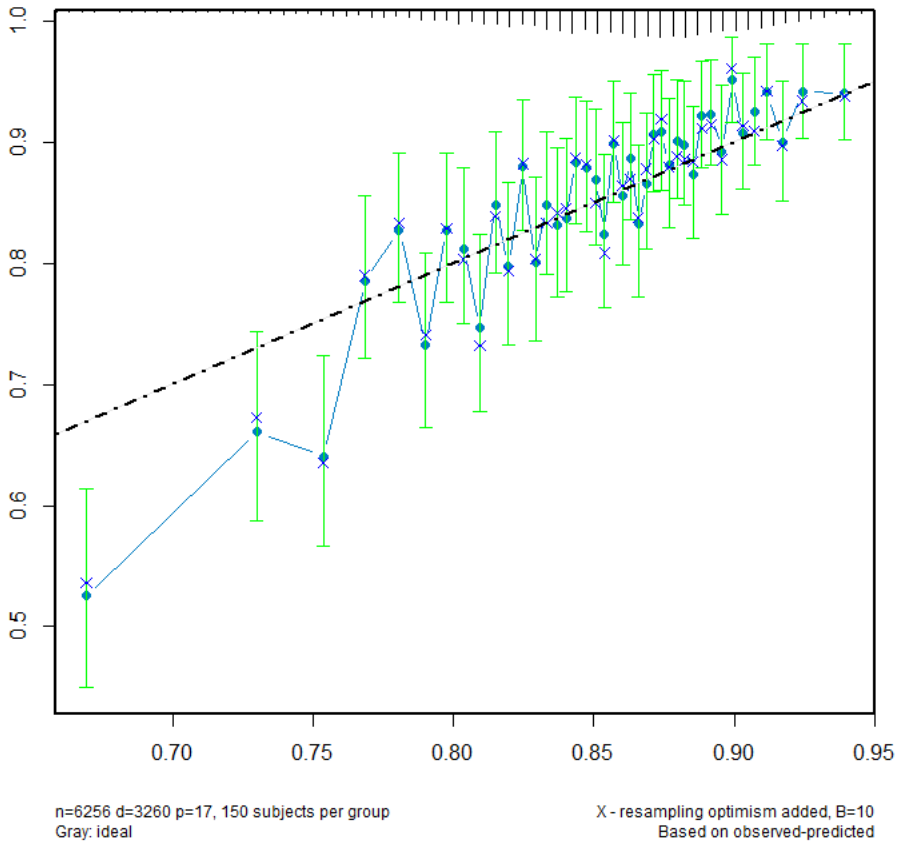

Figure 2. Calibration for 1-year survival for metastatic prostate cancer patients

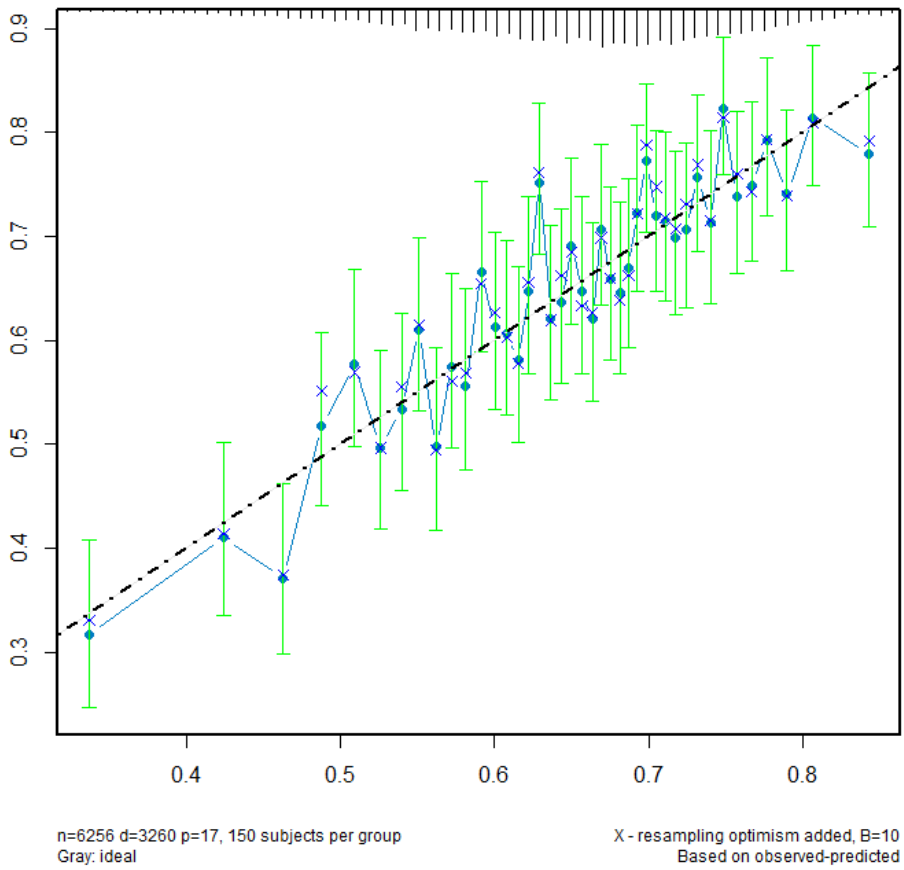

Figure 3. Calibration for 2-year survival for metastatic prostate cancer patients

3.3.3 Decision tree

Decision tree was built with R package "rpart". Tree methods worked by recursive binary partitioning of survival covariates and formed into some small regions, which were called Nodes. Complexity parameter equaled to 0.03 was implemented, this meant any split that did not decrease the overall lack of fit by a factor of 0.03 was not attempted. Based on the original decision tree fitted in Figure 4, all patients were divided into 5 cohorts and each cohort had a respective Kaplan-Meier curve. The first split was based on age < 76.5 or age >= 76.5, separating the

patients with Node 2 and Node 7. Then the splits continued with Charlson-Deyo Score, percent of positive biopsy cores, etc. Took Node 7 as an example, Node 7 was based on Charlson-Deyo Score, which meant for those older than 76.5 years old, 735 patients with Charlson-Deyo Scores of 0 (Node 8) were separated from those who had Charlson-Deyo Scores greater than 0 (Node 9), and based on their K-M plots, patients in Node 8 had a better survival rate compared to patients who ended up in Node 9. Tree pruning was a technique to determine the optimal size for a tree and to remove unnecessary splits. However, because of the uniqueness of survival data, censored responses typically did not have within-node homogeneity, making pruning of the tree unevaluatable (Zhou & McArdle, 2015). Decision tree structure also implied interactions between the covariates. Split by age followed by the split of Charlson-Deyo Score indicating possible interaction between age and Charlson-Deyo Score. Besides the predictors showed in the decision tree, other clinical variables were not selected to be covariates, which did not agree with our Cox proportional hazard model where all clinical variables were included. Since only predictors that were considered as the best splits were selected and had to meet certain criteria to boost the overall performance of the tree (Zhou & McArdle, 2015). Lastly, because survival data are scaled internally and exponentially, so the predicted rate in the Nodes constantly equaled to 1. Thus the accuracy reported by predict() function in R were not applicable (Therneasu & Atkinson, 2019). Overall, since the decision tree excluded too many clinical variables that were statistically significant in the Cox proportional hazards model, the fit was not very ideal in this case.

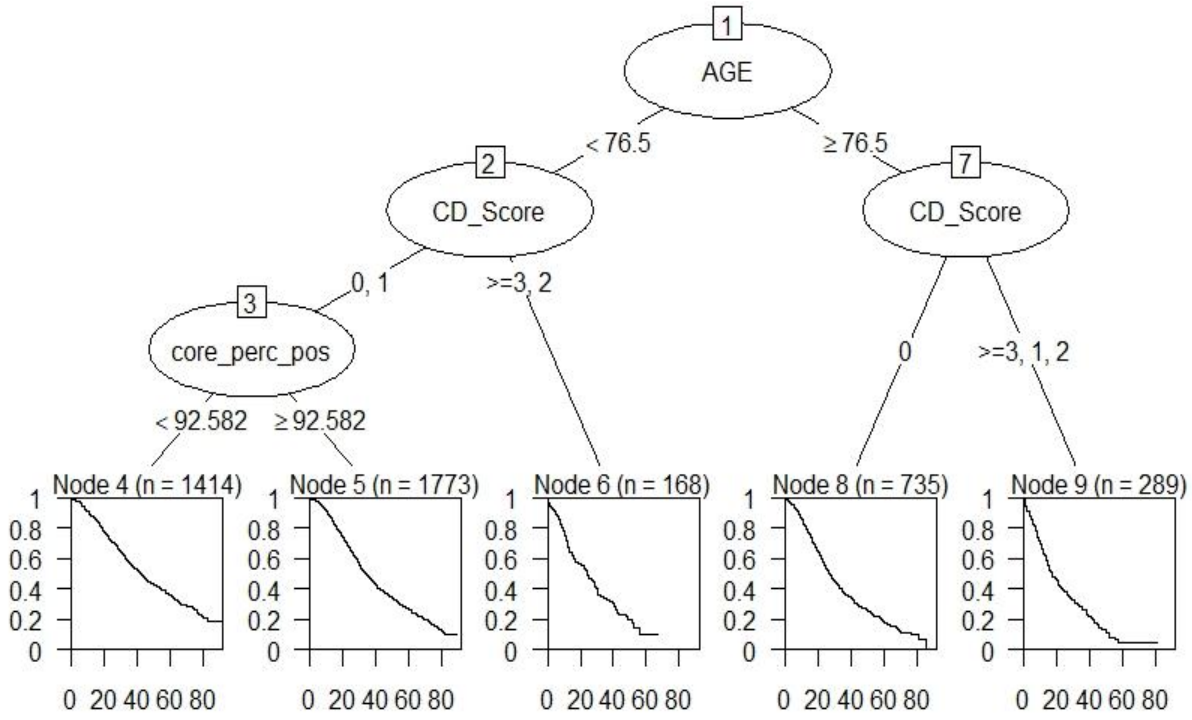

Figure 4. Decision tree for metastatic prostate cancer patiens

3.3.4 Regression coefficient-based scoring system

Beta/Schneeweiss scoring system, Beta/Sullivan scoring system and HR/Charlson scoring system were implemented. Table 1 in Mehta's article (2016) was used as the reference for the calculations presented in Table 5 of this paper. Specifically, weights for Beta/Scheneeweiss scores were increased by 1 unit if beta coefficient increased by 0.3 unit. Beta/Sullivan scores were calculated from regression coefficients that were divided by the smallest absolute value of beta coefficient and rounded to the nearest integer. HR/Charlson scores were the hazard ratios of regression coefficients. Reference groups in the model were switched when retrieving the beta estimates so that all HR were greater than 1. Table 5 displayed weights for metastatic prostate cancer predictors using three scoring systems.

Scores derived from Beta/Schneeweiss and HR/Charlson were close to each other, while the scores calculated from Beta/Sullivan looked somewhat odd in this case. It was because the smallest absolute value of beta coefficient here was 0.005, so the scores were large when regression coefficients were divided by 0.005.

Table 5. *Deriving weights for comorbidity score using three scoring systems*

Disease category	Beta estimate	Beta/ Schneeweiss	Beta/ Sullivan	HR/ Charlson
Race Black	0.3356	1	66	1
Race White	0.3017	1	59	1
CD Score >=3	0.7360	2	144	2
CD Score = 1	0.2685	1	53	1
CD Score = 2	0.4787	2	94	2
Grade unknown	0.3101	1	61	1
Grade poorly/undifferentiated	0.1654	1	32	1
Clinical T1	0.1191	0	23	1
Clinical T2	0.0919	0	18	1
Clinical T4	0.2760	1	54	1
PSA < 10	0.1099	0	22	1
PSA > 20	0.2730	1	54	1
Gleason = 2 - 6	0.4128	1	81	2
Gleason = 8 - 10	0.1985	1	39	1

Bone metastasis involved	0.3093	1	61	1
Age	0.0214	0	4	1
Core percent positive	0.0051	0	1	1
Maximum possible comorbidity score	NA	14	866	20

Among all 6274 metastatic prostate cancer patients, for Beta/Schneeweiss scores, minimum score was 1, mean score equaled to 4.933, maximum score was 8, and 1st, 2nd, 3rd quartiles were 4, 5, 5 respectively. About 70% patients had a Beta/Schneeweiss score >= 5. Thus, Beta/Schneeweiss score was not a well stratification tool in this case. For Beta/Sullivan scores, minimum score was 66, mean score equaled to 268.5, maximum score was 477, and 1st, 2nd, 3rd quartiles were 241, 268, 303 accordingly. Based on Beta/Sullivan scores, we divided the patients into 4 equal-sized cohorts by score quartile and visualized their differences with Kaplan-Meier plots in Figure 5. We could tell that the first cohort of patients who had beta/Sullivan score from 66 to 241 had most higher survival rates. Patients from cohort 2 and 3 had similar results, whereas patients who had beta/Sullivan score from 303 to 477 had the least survival rates. Besides that, C-indices were derived to validate if the scoring algorithm had predictive ability, and C-indices for 4 cohorts were 0.611, 0.573, 0.578, 0.618 respectively which were acceptable. Therefore, although scores calculated from beta/Sullivan scoring algorithms looked odd, it could be used as a risk stratification tool for metastatic

prostate cancer patients. Lastly, for HR/Charlson scoring system, minimum score was 3, mean score equaled to 7.7, maximum score was 11, and 1st, 2nd, 3rd quartiles were 7, 8, 8 respectively. There were approximately 64% patients had a HR/Charlson score >= 8, hence it was not proper to be applied as a classification method. In a conclusion, Beta/Sullivan scoring system was the best among three regression coefficient-based scoring systems and could be used to develop a risk index for prognostic assessment in metastatic prostate cancer patients.

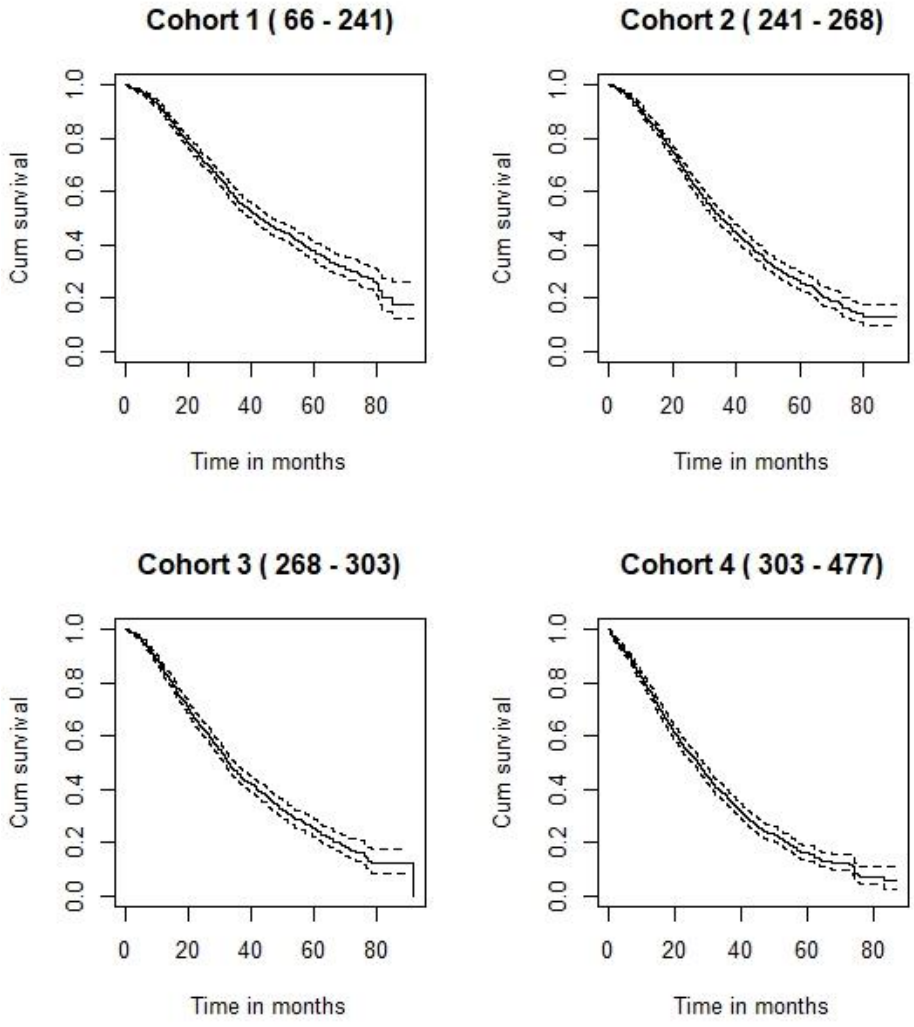

Figure 5. Kaplan-Meier plots of 4 cohorts divided by Beta/Sullivan Scoring system

4. DISCUSSION

Prostate cancer is the sixth most fatal cancer for men worldwide (Jemal et al., 2011). National Comprehensive Cancer Network has stated the importance of risk stratification to guide germline testing and relevant treatment in the prostate cancer guidelines (NCCN Guidelines, 2018). This study explores the association between some clinical variables and survival outcome of 6274 patients with metastatic prostate cancer. The result suggests that race, age, Charlson-Deyo Score, clinical stage, PSA, bone metastasis involvement, positive biopsy cores percentage are independent factors for OS. Then we implement three different risk stratification methods with some crucial clinical variables. There have already existed numerous standards to do risk stratification towards prostate cancer. The most common one is the three-stratum D'Amico classification which came out in the late 1990s. It is applied widely by many international organizations and groups as a clinical guideline. This standard relies heavily on PSA, Gleason score and clinical stage on classifying patients with low, intermediate, and high risk (Mohler et al. 2014). However, as Epstein et al. stated, this guideline might lead to overtreatment or undertreatment due to possible within-group heterogeneity (Reese, 2012). This study demonstrates three ways to do risk stratification, first one is nomogram. Nomogram have been well developed, and widely used in the clinical setting, thus it had been inserted into NCCN prostate cancer guidelines (Susman, 2003). Nomogram in this study has a C-index of 0.614, which should be applicable in further validation. Second one is decision tree. Pantic et al. (2020) carried CART (classification and regression tree analysis) in their study with

prognosis for prostate cancer patients and discovered decisive variables are PSAD and age. In contrast, the important Nodes in this study are age and Charlson-Deyo Score. We think the difference might be caused by different data source. Of three stratification methods in this study, Charlson-Deyo Score has played a crucial role in prostate cancer prognosis. It does not only significantly associate with shorter OS, but also appears a lot in three stratification methods: Charlson-Deyo Score takes great points in nomogram, and plays an important role as a split Node in the decision tree, regression coefficient-based scoring systems are also impacted by it. Of three regression coefficient-based scoring systems, Beta/Sullivan Scoring algorithm is the best stratification tool.

We acknowledge that there were inherent limitations in this study. We only include bone metastasis involvement cases in our study due to the lack of other metastasis cases; there were unbalanced sample size of different races etc., these conditions could influence whether we should include such prognostic markers in the model. For decision tree, the result is not ideal. We believe this is normal since 'rpart' package used to develop decision tree is not perfectly transferrable to longitudinal survival data. Besides that, there are limited software packages available for decision tree built with survival data and are still in development stage. Overall, we will apply some external validations of these risk stratification methods and compare it with other methods. Our goal is to further complete such investigations and to promote the use of these methods in clinical decision-making.

5. References

Balachandran V.P., Gonen M., Smith J.J.& DeMatteo R.P. (2015), Nomograms in Oncology – More than Meets the Eye. Lancet Oncol. 16(4): e173-e180. doi: 10.1016/S1470-2045(14)71116-7

Bermejo P, Vivo A, Tarraga P.J. & Rodriguez-Montes J. (2015). Development of Interpretable Predictive Models for BPH and Prostate Cancer. Clin Med Insights Oncol. 9: 15-24. doi: 10.4137/CMO.S19739

Bianco F.J. (2006). Nomograms and Medicine. European Urology 50. 884 – 886.
Doi: 10.1016/j.eururo.2006.07.043

Bilimoria K.Y., Stewart A.K., Winchester D.P. & Ko C.Y. (2007). The national cancer data base: A powerful initiative to improve cancer care in the United States. Annals of Surgical Oncology. 15(3): 683-690. doi: 10.1245/s10434-007-9747-3

Christopher J. S., Yu-Hui C., Michael C., Glenn L., David F. J., et al. (2015). The New England Journal of Medicine, 373:737-746. doi: 10.1056/NEJMoa1503747

D'Amico AV, Whittington R, Malkowicz SB, Fondurulia J, Chen M-H, Kaplan I, et al. (1999). Pretreatment nomogram for prostate-specific antigen recurrence after radical prostatectomy or external-beam radiation therapy for clinically localized prostate cancer. J Clin Oncol.17:168–172.

Epstein JI, Zelefsky MJ, Sjoberg DD, Nelson JB, Egevad L, Magi-Galluzzi C, et al.(2016). A contemporary prostate cancer grading system: a validated alternative to the Gleason score. Eur Urol. 69:428–435.

Gravis G, Boher J-M, Joly F, …Fizazi k. (2016). Androgen Deprivation Therapy (ADT) Plus Docetaxel Versus ADT Alone in Metastatic Non castrate Prostate Cancer: Impact of Metastatic Burden and Long-term Survival Analysis of the Randomized Phase 3 GETUG-AFU15 Trial. European Urology, Volume 70, issue 2, Pages 256-262. Retrieved from: https://doi.org/10.1016/j.eururo.2015.11.005

Grivas P, Robins D, Hussain M. (2013) Predicting Response to Hormonal Therapy and Survival in Men with Hormone Sensitive Metastatic Prostate Cancer. Crit Rev Oncol Hematol. 85(1); 82-93.
doi: 10.1016/j.critrevonc.2012.05.003

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. (2011). Global cancer statistics. CA Cancer J Clin; 61:69-90[Erratum, CA Cancer J Clin 2011;61:134.]

Kruppa J., Ziegler A., & Konig I.R. (2012). Risk estimation and risk prediction using machine-learning methods. Hum Genet. 131(10): 1639-1654, doi: 10.1007/s00439-012-1194-y

Lei J, Liu., Wei Q, Song T, Yang L, Meng Y & Han P. (2015). Androgen-deprivation therapy alone versus combined with radiation therapy or chemotherapy for nonlocalized prostate cancer: a systematic review and meta-analysis. Asian Journal of Andrology. 18(1): 102-107.
doi: 10.4103/1008-682X.150840

Mheta H.B., Mehta V, Girman C.J., Adhikari D & Johnson M.L. (2016). Regression coefficient-based scoring system should be used to assign weights to the risk index. Journal of Clinical Epidemiology. November 2016, Volume 79, Page 22-28.
doi: https://doi.org/10.1016/j.jclinepi.2016.03.031

Mohler J.L., Kantoff P.W., Armstrong A.J., Bahnson R.R., Cohen M, D'Amico A.V., et al. (2014). Prostate cancer. J Natl Compr Canc Netw. Version 2.12:686–718.

NACHC (National Association of Community Health Centers) (2019). Value Transformation Frameworl Action Guide. Retrieved from http://www.nachc.org/wp-content/uploads/2019/03/Risk-Stratification-Action-Guide-Mar-2019.pdf

National Cancer Institute (2017). Retrieved from https://www.cancer.gov/types/metastatic-cancer

NCCN Clinical Practice Guidelines in Oncology (2018). Prostate Cancer. National Comprehensive Cancer Network website. Retrieved from

www.nccn.org/professionals/physician_gls/pdf/prostate.pdf.

Pantic D, Stojadinovic M, & Stojadinovic M. (2020). Decision Tree Analysis for Prostate Cancer Prediction in Patients with Serum PSA 10 ng/ml or Less, *Serbian Journal of Experimental and Clinical Research*, *21*(1), 43-50. doi: https://doi.org/10.2478/sjecr-2018-0039

Reese A.C., Pierorazio P.M., Han M., Partin A.W. (2012). Contemporary evaluation of the National Comprehensive Cancer Network prostate cancer risk classification system. Urology. 80:1075–1079. doi: https://doi.org/10.1016/j.urology.2012.07.040

Sumi S, Vinayak M, Felix Y.F, Paul L.N. (2017). Prostate International; 5(3): 89-94. doi: 10.1016/j.prnil.2017.04.003

Susman E.D., Phend C. (2003). Nomograms Inserted Into NCCN Prostate Cancer Guidelines. Oncology Times: July 25th. Volume 25, issue 14, p28-31.
doi: 10.1097/01.COT.0000289304.22023.f9

Therneasu T.M., Atkinson E.J. (2019). An introduction to Recursive Partitioning Using the RPART Routines. Mayo Foundation. Section 8.4. Derived from:
https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf

Yang D, Mahal Br, Vinayak M & Neil.E. (2018). Androgen Deprivation Therapy and Overall Survival for Gleason 8 Versus Gleason 9-10 Prostate Cancer. European Urology. Volume75, Issue1, Pages 35-41.
doi: https://doi.org/10.1016/j.eururo.2018.08.033

Zhou Y, McArdle J. J. (2015). Rationale and Applications of Survival Tree and Survival Ensemble Methods. Psychometrika. 2015 Sep; 80(3): 811-833. doi: 10.1007/s11336-014-9413-1.