Change in Prostate Cancer Stage over Time

Fei Zhang[a]
Biostatistics and Data Science, Weill Cornell Medicine College, Cornell University, New York, U.S.A.

Keywords: PSA, Prostate Cancer, Chi-Squared Test, Proportionality Test, Early Diagnosis.

Abstract: Prostate cancer is a form of cancer that occurs in the prostate gland cells among males and it is the second most common type of cancer among males in the US. In early 1990, the American Urological Association (AUA) and the American Cancer Society (ACS) started recommending annual prostate cancer screening with Prostate-specific antigen test (PSA) which is a blood test. In October 2011, the US Preventive Service Task Force (USPSTF) published a final guideline recommending against the use of PSA based screening for prostate cancer. The influence of the use of PSA on the diagnosis rate of prostate cancer, especially in the early stage, has become a hot research topic. The goal of this paper is to determine whether there has been a change in proportion of men diagnosed with localized/regional prostate cancer over time due to the changes in PSA screening recommendations, and whether this change of proportion is associated with other risk factors. This paper uses Chi-Squared test, proportionality test and other methods to analysis data. There was significant difference between the proportion of localized/regional prostate cancer in year 2004 and 2015 as USPSTF recommended against the use of PSA based screening. Age, racial, region and marital status significantly affect the distribution of the proportion of initial stage prostate cancer.

1 INTRODUCTION

Prostate cancer is a form of cancer that occurs in the prostate gland cells among males and it is the second most common type of cancer among males in the US. As per a 2016 CDC annual report, for every 100,000 men in the US, 101 new prostate cancer cases were reported in the year 2014 and of those cases, 19 died(Centers for Disease Control and Prevention). In early 1990, the American Urological Association (AUA) and the American Cancer Society (ACS) started recommending annual prostate cancer screening with Prostate-specific antigen test (PSA) which is a blood test. PSA is made by the prostate gland and high levels of PSA may be indicative of prostate cancer or other non-cancerous conditions. PSA screening was a cheaper and non-invasive alternative to a digital rectal exam which is one main reasons for PSA based screening being recommended even though there was no supported clinical trial evidence for PSA accurate indicator of prostate cancer. There was an alarming increase in the incidence rates as PSA based screening became more common and by 1992 the incidence rate of prostate cancer in the US nearly doubled. Mei Aobing et al(Aobing et al. 2017, Mistry, Cable 2003, Zhao, Huang, Cheng et al. 2014, Kramer, Brown, Prorok, et al. 2013). questioned the sensitivity and specificity of PSA, especially when PSA is between 4.00 ng/mL and 10.00 ng/mL. There is an overlap between SERUM PSA levels in PATIENTS with benign prostatic hyperplasia (BPH)and prostaticcancer (PCa), making it difficult to distinguish benign prostatic hyperplasia from prostate cancer. K. Mistry et al (Mistry, Cable 2003). 's study found that chronic prostatitis, indplacement of urinary ducts, prostate massage, and other conditions can lead to abnormal PSA test results, that is, PSA is a prostate-specific marker rather than a marker of prostate cancer. In October 2011, the US Preventive Service Task Force (USPSTF) published a final guideline recommending against the use of PSA based screening for prostate cancer (USPSTF).

The significant change in incidence rates and diagnosis levels of prostate cancer cases over the last few decades points towards a possibility of overdiagnosis and overtreatment due to these policy changes. There is scope to further study and evaluate the impact of this change in the policy and scientific

[a] https://orcid.org/0000-0003-3199-0521

Zhang, F.
Change in Prostate Cancer Stage over Time.
DOI: 10.5220/0011368700003438
In *Proceedings of the 1st International Conference on Health Big Data and Intelligent Healthcare (ICHIH 2022)*, pages 308-313
ISBN: 978-989-758-596-8

landscape over the years. Early diagnosis, timely operation and effective endocrine therapy can greatly reduce the mortality rate of prostate cancer. Therefore, to study the causes and influencing factors of prostate cancer and determine the susceptible population will provide an important basis for effective prevention, early diagnosis and improved survival rate. The goal of this project is to determine whether there has been a change in proportion of men diagnosed with localized/regional prostate cancer (out of the total number of diagnosed prostate cancer) over time due to the changes in PSA screening recommendations, and whether this change of proportion is associated with age group, the race groups, region groups and marital status. Chang et al (Chang 1996). 's study found that PSA increased with age, and different age groups had different effects on PSA. The older you are, the greater the impact. In addition, the mean prostate volume of all ages also increased with age, and the prostate volume increased with age. There has been a change in proportion of men diagnosed with localized/regional prostate cancer (out of the total number of diagnosed prostate cancer) over time due to the changes in PSA screening recommendations. And this change of proportion is associated with age group, the race groups, region groups and marital status.

2 METHOD

2.1 Data

The Surveillance, Epidemiology, and End Results (SEER: https://seer.cancer.gov/) database which is maintained by the National Cancer Institute (NCI), was used for this study. The database extract contains 28% of all cancer cases in the US, diagnosed between 2004 and 2015. The data was collected from 18 different population-based registries and contains incidence as well as survival records of patients by patient ID. The dataset is deidentified and is compliant with HIPAA regulations regarding protection of patient privacy and intended use of the data for research purposes. There is a total of about 1.94 million patient records out of which 230,326 records are associated with prostate cancer. Each record contains a patient ID, registry ID, year and month of diagnosis, age at diagnosis, histology (stage) and other demographic information such as race, birth year, sex. A check for duplicates yielded 31 patients with 2 diagnoses at different points of time, only initial records were as the evaluation is primarily based on the initial prostate cancer diagnosis.

2.2 Statistical Analysis

Since the raw data contained 133 columns, it was essential to select the relevant variables, filter the essential categories, and create categorical variables for age groups as well as diagnosis dates (before or after PSA final guidelines). In the cleaning process, male patients diagnosed with prostate cancer in interested stages were selected from the original SEER dataset. Duplicate records, patients with age under 50 years old, and unrelated variables and stages were removed. Then The clean data was then grouped cases by race, region, marital status, and age group to yield number of cases as a function of time (month-year). In order to test the difference in proportion of localized relative to the overall cases in different categories (race, age group, geographical region, marital status), we used proportionality test for two groups and Chi-squared test of independence in case of multiple groups. These non-parametric tests were used after making sure that the following requirements were met. Variables are categorical (binary) in nature: Prostate Cancer Stage(Localized/Regional and Distant). All cases belonged to a single population. Data management and statistical analyses were performed using R, version 4.1.1.

3 RESULTS

In general, the widespread adoption of PSA-based prostate cancer screening caused a stage migration toward earlier stage of prostate cancer at diagnosis (95% CI 0.0347 - 0.0451) during the early 2000s. Thus, there was significant difference between the proportion of localized/regional prostate cancer in year 2004 and 2015 as USPSTF recommended against the use of PSA based screening. The dotted line marks the time when USPSTF released their final guidelines (October 2011) and after which, there is a sharp linear decrease in the proportion of localized/regional cases detected in the following years (Figure 1).

	PSA Screening Status							
	PSA Screening				No PSA Screening			
Variables	Proportion	pvalue	95% CI	Statistics	Proportion	pvalue	95% CI	Statistics
Race		<0.0001	(-0.0206,-0.0137)	119.19		0.0027	(-0.0144,-0.0028)	1542.10
Black	94.00%				91.90%			
White	95.70%				92.80%			
Age		<0.0001	(-0.0508,-0.0459)	1904.40		<0.0001	(-0.0907,-0.0807)	9.01
50-70	97.10%				95.20%			
above 70	92.30%				86.70%			
Region		0.0281		9.09		<0.0001		39.88
Midwest	95.70%				93.40%			
Northeast	95.50%				92.60%			
South	95.60%				93.20%			
West	95.30%				92.00%			
Marital Status		<0.0001		1228.70		<0.0001		713.91
Divorced	93.40%				90.20%			
Married	96.10%				93.30%			
Separated	93.10%				90.30%			
Single	93.30%				88.20%			
Widowed	87.80%				80.50%			
Unmarried	90.00%				90.60%			

Figure 1: Proportion of US Males Diagnosed with Initial Stage Prostate Cancer Over Time.

3.1 Age Groups

From 2004 to 2015, the proportion of localized/regional prostate cancer for age group greater than 70 years was less than that of 50-70 years age group. Moreover, the proportions of both the age groups were stagnant during the time period between 2004 and October 2011, however there was significant decrease in proportions of both age groups since October 2011(Table1). The two-proportion z-test showed that there were differences in proportion of men diagnosed with localized/regional disease by age groups: 50-70 and over 70 ($\chi 2$ statistic: 327.07, 95% CI: -0.06 - -0.056). The influence of PSA screening guideline change on the proportion was different for both age groups (with PSA screening: 95% CI -0.0508 - - 0.0459, without PSA screening: 95% CI -0.0907 - -0.0807). The plot shows that the age group of individuals over 70 years experienced a steeper decrease in proportions as compared to the 50-70 years age group.

Table 1: Summarized Results of Statistical TestsConducted on All Groups.

	PSA Screening Status							
	PSA Screening				No PSA Screening			
Variables	Proportion	pvalue	95% CI	Statistics	Proportion	pvalue	95% CI	Statistics
Race		<0.0001	(-0.0206,-0.0137)	119.19		0.0027	(-0.0144,-0.0028)	1542.10
Black	94.00%				91.90%			
White	95.70%				92.80%			
Age		<0.0001	(-0.0508,-0.0459)	1904.40		<0.0001	(-0.0907,-0.0807)	9.01
50-70	97.10%				95.20%			
above 70	92.30%				86.70%			
Region		0.0281		9.09		<0.0001		39.88
Midwest	95.70%				93.40%			
Northeast	95.50%				92.60%			
South	95.60%				93.20%			
West	95.30%				92.00%			
Marital Status		<0.0001		1228.70		<0.0001		713.91
Divorced	93.40%				90.20%			
Married	96.10%				93.30%			
Separated	93.10%				90.30%			
Single	93.30%				88.20%			
Widowed	87.80%				80.50%			
Unmarried	90.00%				90.60%			

3.2 Race Groups

During the period from 2004 to 2015, the proportion of localized/regional cases decreased for both major race groups (black and white) after October 2010. The proportion among white males was greater than black males and two-proportion z-test showed that the difference in the overall proportion by race was statistically significant ($\chi 2$ statistic: 117.92, 95% CI 0.012 - 0.018). The influence of PSA screening guideline change in the proportion among black and white patients was different (with PSA screening:

95% CI -0.0508 - -0.0459, without PSA screening: 95% CI -0.0907 - -0.0807).

3.3 Region Groups

All the regions (Midwest, Northeast, South, and West) showed a decrease from 2010 with slight variations in trends. Chi-squared test showed that there is an association between the US regions and the proportions ($\chi 2$ statistic: 35.384, df = 3, p-value < 0.0001). So, there is sufficient evidence to state that the influence of PSA screening on proportion of localized/regional stage varies between different US geographical regions.

3.4 Marital Status

Marital status was divided as Divorced, Married, Separated, Single, Unmarried or domestic partner and Widowed. In the period between 2004 and 2015, there were slightly variance and experience decrease since 2010 in marital status. The unmarried group has some outliers on the lower end of the proportion range in the time trend analysis, however smoothing shows similar trend lines among all marital status. The married group has the highest proportion of regional stage prostate cancer cases from 2004 and 2015, while widowed group had the least. Chi-squared test of independence shows that there is an association in the proportion of men diagnosed with localized/regional disease with marital status: Divorced, Married, Separated, Single, Unmarried or domestic partner and Widowed ($\chi 2$ statistic: 1928.10, df = 5, p-value < 0.0001). Additionally, the influence of PSA screening on proportion of localized/regional stage depends on marital status (with PSA screening: $\chi 2 = 9.088$, df = 3, p-value = 0.02814, without PSA screening: $\chi 2 = 39.878$, df = 3, p-value < 0.0001).

4 DISCUSSION

Many articles have confirmed that age, residence, race, and marital status have a significant impact on the diagnosis of prostate cancer. The incidence of various tumors is very different in different countries in the world. Even the incidence of different regions in the same country is also very different. For example, the country with the highest incidence of gastric cancer is Japan, the incidence of colorectal cancer is the highest in the United States, and Sweden has the highest incidence of prostate cancer. For different regions, the probability of occurrence of each type of cancer is different in each region, which may be affected by local eating habits, weather, air quality, water quality and other external environmental factors.

From the analysis of internal reasons, the incidence of cancer may be related to mental state, mental quality, happiness index, personal physical fitness and so on. Many researchers have proven through genetics that people of different regions and races have different genes for prostate cancer susceptibility, and the order of these genes is also inconsistent, which will fundamentally affect the prevalence and incidence of cancer. Nan Di et al (nan, Yun 2019, Li 2003) found that the differences in the genotype and allele frequency distribution of susceptibility genes between different races in prostate cancer caused the abnormal incidence of prostate cancer, which can directly participate in the development of prostate cancer. Occurrence and development. There are obvious differences in the incidence of prostate cancer among people of different races and regions, and the incidence varies dozens of times. Studies by foreign scholars have shown that there are obvious differences in the incidence of prostate cancer among different ethnic groups in the United States, such as Indians, African Americans, Mexican Americans and Asian Americans. Studies by domestic scholars have shown that there are obvious differences in the distribution of your genotypes under the front ranks of different ethnic groups, which may affect the hormone levels and biological effects of different individuals. VDR genes and androgen-related gene polymorphisms have obvious racial types, and they are different from each other. The incidence of prostate cancer is the same in different races.

Genetic factors are undoubtedly the main factors affecting the incidence of prostate cancer, and the differences in genetic gene sequences between different races are the main factors contributing to the huge differences in the incidence of prostate cancer among different races. Those studies' results are consistent with this paper.

Besides, a study found a significant increase in the incidence of prostate cancer among Asian immigrants. It suggests that factors such as geography and dietary habits may play a role in the development of prostate cancer. Chuiguo Huang (Huang 2018) used multi-factor Cox regression analysis, survival analysis and other methods to confirm that related factors such as age, race, marital status, PSA concentration, T stage in TNM staging, tumor tissue grading, and the use of different interventions are affecting the Gleason score of 8. Separate

independent risk factors for the prognosis of prostate cancer patients.

In addition, prostate-specific antigen (PSA), as the most valuable tumor marker for prostate cancer, only has the specificity of prostate tissue but not the specificity of prostate cancer. Various prostate tissues (including normal tissues, benign hyperplasia tissues and cancer tissues) are PSA can be secreted, leading to its lack of specificity and sensitivity in the diagnosis of early prostate cancer. For a long time, clinicians have used total PSA = 4.0 ng/ml as the threshold for screening prostate cancer (PCa) and non-prostate cancer and has been widely used. However, a large number of studies have shown that in patients with tPSA≤4.0 ng/ml, the incidence of prostate cancer is not low; and among patients with tPSA>4.0 ng/ml, 75% of patients do not have prostate cancer. Therefore, the use of a single PSA indicator with fixed threshold value to diagnose prostate cancer has a higher false positive rate and false negative rate. Many new studies have shown that the original threshold of the PSA method should be adjusted according to the patient's actual physical condition and past medical history. For example, big data analysis and machine learning methods can be used to obtain a new PSA threshold for early warning PCa in the T2DM population (the original threshold 4.0 ng/ml), and calculate its sensitivity and specificity. Probability function fitting is used to estimate the distribution of PSA levels in the overall population, support vector machines are used to calculate new thresholds, and receiver operating characteristic (ROC) curves are used to test its diagnostic efficacy.

This article only analyzed several risk factors. However, change in proportion of men diagnosed with localized/regional prostate cancer over time due to the changes in PSA screening recommendations may associated with other factors other than age group, the race groups, region groups and marital status. Besides, this article can use survival analysis, multivariate statistical analysis and other models to study more risk factors and influencing factors of prostate cancer in the future. MIC can be used to carry out factor correlation analysis on the large and complex medical data of major hospitals, and then obtain more accurate relationships and visual images from the complex data. New models or other methods can be used to further analyze the impact of the policy of discontinuing the PSA screening recommendations.

5 CONCLUSIONS

A statistical analysis of the SEER dataset helped us understand the effects of healthcare policies on prostate cancer diagnosis levels over the years. The overall proportion of cases with localized/regional prostate cancer show a slight increase between 2004 and mid of 2008 which is due to the prominence of PSA-based prostate cancer screening. In 2008, the overall proportion started declining due to rising awareness of overdiagnosis of initial stage (localized/regional) of prostate cancer from PSA screenings. Additionally, during this time (2008) the USPSTF began recommending men over the age of 75, against PSA screening tests which led to a steeper decline in the proportion of initial stage cases in the above 70 years age category. In October 2011, the USPSTF issued a draft recommending against PSA screening for other age groups as well and due to this a steep linear decline in the proportion of initial stage prostate cancer can be seen across all categories: race, region, marital status, and age group. Statistical tests were conducted to determine if the proportions between the groups (within each category) were significantly different. Table 1 shows that the p-value for the proportionality tests was consistently less than the level of significance (0.05), due to which we could reject the null hypothesis (proportions are equal). Thus, we can conclude that the proportions of initial stage (localized/regional) prostate cancer cases were significantly different between age groups (50-70 and above 70) as well as racial groups (black and white). It was also found that the proportions of men diagnosed with initial stage of prostate cancer were statistically different between 2004 and 2015, based on the sample size. In order to obtain a detailed analysis, it was important to look at the proportion trends by other factors like marital status and regions in the US. Histogram of proportionvalues for different regions showed that the distributions had different variances, and Chi-Squared test indicated the total proportions for the regions are statistically different. Finally, marital status also significantly affected the distribution of the proportion of initial stage prostate cancer. Due to the differences in domestic and foreign policies, the current domestic research in China focuses on other research based on PSA screening recommendations. The main research directions are as follows. Factors affecting prostate cancer. Whether the threshold value of the detection index needs to be adjusted according to the patient's actual situation such as past medical history and how to adjust. In addition to the current detection methods and indicators taken into account, do you need to add

other indicators to make the detection results more reliable and effective, avoid biased results, and reduce unnecessary testing for patients. At present, few scholars or institutions have studied the influence of the existence of PSA screening on the efficiency of prostate cancer diagnosis. This article fills this loophole very well, and hope that this article encourages more scholars to study the detection method itself. With the rapid development of computers, statistical learning and artificial intelligence deep learning algorithms are gradually being integrated with medicine. Use artificial intelligence, big data complex analysis and other emerging computing methods to explore new research methods based on medical observation data, and then better diagnose. It is also possible to analyze the influencing factors of prostate cancer from a new perspective. In addition to the four factors mentioned in the article, as well as many genes that are currently being studied, there are actually many factors that can be analyzed. Humans are social animals and are affected by various factors, such as psychological factors, diet, and water sources.

REFERENCES

Centers for Disease Control and Prevention. United States cancer statistics: 1999–2014 cancer incidence and mortality data. https://nccd.cdc.gov/uscs/

Chuiguo Huang. A nomogram for analyzing prognostic features in patients with Gleason 8 prostate tumor[D]. The Second Clinical College of Zhengzhou University: Department of Urology, 2018

Di nan, Zhizhong Yun. Research progress on ethnic differences and susceptibility to prostate cancer genome[J]. Journal of Clinical Medical, 2019,6(06):194-195.

Jiangping Chang. Effects of gland volume and age on prostate specific antigen in benign prostate hyperplasia[J]. Journal of Clinical Urology, 1996(4): 207-209.

KramerB S, Brown ML, Prorok PC, et al. Prostate cancer screening: What we know and what we need to know [J]. Annals of internal Medicine, 2013,119 (9) : 914-923.

Mei Aobing, et al. The correlation study and clinical guidance of serum EPCA-2 and PSA in the diagnosis of early prostate cancer[J]. Guizhou Medical Journal, 2017(9): 917 - 920.

Mistry K, Cable G. Meta-analysis of prostate-specific antigen and digital Rectal examination for prostate carcinoma: A meta-analysis [J]. Journal of the American Board of Family Practice, 2003,16 (2) : 95-101.

Ming Li. The incidence of prostate cancer and associated factors[J]. China Cancer, 2003(12): 4-7.

U.S. Preventive Services Task Force (USPSTF). Rockville, MD: U.S. Dept. of Health & Human Services, Agency for Healthcare Research and Quality

Zhao R, Huang Y, Cheng G, et al. Developing a follow-up Strategy for Patients with PSA Ranging from 4 to 10ng/mL via a New Model to Reduce Unnecessary Prostate Biopsies [J]. Plos One, 2014,9 (9): e106933-e106933.