Research Question

What is the correlation between PSA levels and prostate cancer severity in men older than the age of 50 in the US?

Introduction

Prostate cancer runs through my family, some were diagnosed at a young age while others in their late fifties. My grandfather lived with prostate cancer for approximately 10 years before passing away from the metastasis of cancer at the age of 87. When he was diagnosed with stage 2 prostate cancer, his PSA level was 16.10ng/ml. On the other hand, my cousin was diagnosed with stage 3 prostate cancer at the age of 38 and lived in an extremely critical condition while his health was deteriorating for 3 years. At the time of diagnosis, his PSA level was 22.75 ng/ml. The difference in both of these cases made me curious about how a patient's PSA level indicates the severity of his cancer. Therefore, I searched for publicly available data to explore this topic. Using the raw data which I have, I decided to research the correlation between PSA levels and prostate cancer severity in men older than the age of 50, who reside in the US.

Background Knowledge

Prostate cancer is one of the most common types of cancer that affects mainly elderly men (Humphrey *Prostate cancer*). Prostate cancer occurs in the tissue of the prostate; the prostate is a gland that is found under the bladder, an organ which stores and empties the urine in the human body (Anatomy of the urinary system). Cancer is a life threatening disease that causes several implications to a human's physical and mental health. According to Mayo Clinic, prostate cancer is the "second-leading cause of cancer death" for men (*Cancer*). Researchers have proclaimed that the specific cause behind prostate cancer remains unknown, however, they have identified multiple factors (e.g. race, age, diet) that increase the risk of prostate cancer. In general, prostate cancer starts when a genetic mutation in the DNA of healthy prostate cells occurs (*Prostate cancer*). Genetic mutations infiltrate our cell cycle's control factors, causing normal prostate cells to multiply very rapidly and uncontrollably, which leads to the production of cancerous cells in the prostate (*What causes prostate cancer?*). Usually, the human body will attempt to eliminate cells with mutated DNA, before they become cancerous cells, however, with age the body loses its ability to do so effectively. Therefore, men of older age are at a much higher risk of prostate cancer (*What is cancer?*).

Prostate-Specific Antigen (PSA) is a protein that is produced by normal prostate cells and cancerous malignant cells in the prostate. PSA is an enzyme that plays a significantly important role in fertility and the liquefaction of seminal fluid. A PSA test is a blood test which examines the blood levels of the substance that thins the ejaculate shortly after ejaculation. A PSA test is measured in nanograms of PSA per milliliter (ng/ml). A PSA test is most often used to detect prostate cancer, whenever prostate enlargement is diagnosed in a rectal examination (*Prostate-specific antigen (PSA) test*).

Data Source

The data set used in this investigation is raw data collected by the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. The raw data was collected between 1995 and 2018. The number of patients participating in this study was 76, 677. The data set was obtained from the Cancer Data Access System of the

National Cancer Institute. The patients participating in this data collection process will remain unknown and all data will be used ethically in accordance with the signed Data Transfer Agreement.

Hypothesis

Using the raw data from National Cancer Institute, it can be hypothesized that there is a strong positive correlation between PSA levels and the severity of prostate cancer in men older than the age of 50, who reside in the US. As PSA levels increase, the stage of prostate cancer a patient has will also increase, because at higher PSA levels, more PSA proteins are being produced by cancerous prostate cells, thus, prostate cancer stage will be higher.

Variables

The independent variable in this investigation is PSA levels at the time of prostate cancer diagnosis. PSA levels used in this investigation will range from 0 ng/ml to 100 ng/ml. Any PSA reading above 100 ng/ml is excluded, because that is an unrealistically high PSA level, and there was only a small number of patients who exceeded 100 ng/ml (HALO Diagnostics; Lojanapiwat, Bannakij). A PSA level above 10 ng/ml is usually considered to be very high. Older men with PSA levels below 10 ng/ml have a lower risk of being diagnosed with prostate cancer with a probability of 1:4. While men with PSA levels above 10 ng/m have a 50% chance of having or developing prostate cancer (HALO Diagnostics; *Prostate cancer screening tests*). The dependent variable is prostate cancer severity, which is measured by the stage of cancer (stages 1 to 4). For reference, below is a table (Fig. 1) which describes the nature of each of the prostate cancer stages.

Stage 1 Prostate Cancer	The tumor is on a small part of the prostate. Cancer is slowly progressing. Healthy cells and cancer cells are difficult to distinguish.
Stage 2 Prostate Cancer	The tumor is localized to the prostate. The tumor grows significantly. Cancer has a higher chance of spreading and growing beyond the prostate. Healthy cells and cancer cells can be differentiated. The tumor can be felt during rectal examinations.
Stage 3 Prostate Cancer	The tumor grows significantly faster. Cancer has advanced within the prostate; tumors affect a larger area of the prostate. Cancer will likely spread outside of the prostate. Healthy cells and cancer cells are easily distinguishable.
Stage 4 Prostate Cancer	Cancer has spread to the bones, lymph nodes, and other organs in the body. Healthy cells and cancer cells are well differentiated. Tumors are very large and occupy a greater area of other body parts, not only the prostate.

Fig. 1. Description of All Four Prostate Cancer Stages (Prostate cancer - stages and grades)

Inclusion Criteria

There are multiple variables in this lab report that are controlled to ensure all the results are reliable and as precise as possible. A clinical consideration that is made is the PSA level that the prostate cancer patients

have. PSA levels in older men have a high range of PA levels that range from 0.00 to 1000.00, therefore in this IA the data has been filtered in order to exclude prostate cancer patients with PSA levels above 100.00 ng/ml.

Another consideration made is the type of cancer that patients have. It is extremely important that all the patients must have only prostate cancer, because this IA is investigating the correlation between PSA levels and prostate cancer stage without consideration for any other medical condition; otherwise this investigation would be invariable (*PSA Test*).

To add on, the age of the prostate cancer patient must be taken into consideration, therefore this investigation is focusing on men aged 50 and older. Men of any age can be diagnosed with prostate cancer, however, with an increase in age the risk of prostate cancer significantly increases. According to WebMD, 66 years old is the average age of men when they are first diagnosed with prostate cancer (*McMillen, Matt*).

Lastly, a geographical consideration made is the location in which patients live; for this investigation, all the patients reside in the US. Prostate cancer is the second most common type of cancer in men in the US, therefore, a wide range of data will be provided for a thorough analysis of the data. In 2022, approximately 268,500 men have been diagnosed with prostate cancer (*McMillen, Matt*).

Variable	Inclusion Criteria
Clinical (PSA level)	All patients have a PSA level between 0 ng/ml and 100 ng/ml.
Clinical (Prostate cancer patient)	All patients have prostate cancer.
Demographics (Age)	All patients are above the age of 50.
Geographic (Location)	All patients live in the United States of America.

Fig. 2. Inclusion Criteria Table

Methodology

In this investigation, a thorough analysis was conducted using IBM's SPSS (a software used for statistical analysis) and MS Excel, in order to effectively analyze the data and investigate the correlation between both variables. The data was collected from multiple medical facilities across various states, therefore, the uncertainty of the tools used during prostate biopsies and PSA tests were not reported, thus, the National Cancer Institute did not provide me with an uncertainty value for the raw data. Unfortunately, this weakens the results of the research, because an uncertainty is a "quantitative measurement of variability in the data" (Anthony Carpi and Anne E. Egger). However, according to an article by Discovery of Sound in the Sea, determining uncertainty of a "scientific process when it affects health" is very difficult (Scientific uncertainty).

The statistical analyses incorporated in this investigation are: descriptive statistics, R^2 value, and an ANOVA test. I calculated several descriptive statistical values, such as mean PSA levels per cancer stage, standard deviation (not a statistical test), and more, in order to provide multiple analyses on the data, the correlation between both variables, and the overall trends of the data. A scatter plot graph was used to plot the mean PSA levels for each stage and a linear trendline was added to illustrate the general increase of PSA levels. In addition, error bars were calculated using standard deviation and were used to interpret the spread of data and its statistical significance. Furthermore, the R^2 value was used to analyze the type and strength of the

correlation present between PSA levels and prostate cancer stages. Lastly, the f-calculated, f-critical, and p values from the ANOVA test were used to determine whether the data is statistically significantly different.

The table below (figure 3) displays the 6 groups created for this investigation, to categorize the various PSA level ranges. The groupings in Fig. 3 were organized based on past scientific research (Lojanapiwat et al.; HALO Diagnostics *Understanding PSA levels & the PSA test*).

PSA Level Ranges (ng/ml)	Groupings
0.00 < PSA Level ≤ 2.50	Normal
2.50 < PSA Level ≤ 4.00	Elevated
4.00 < PSA Level ≤ 10.00	Functionally high
10.00 < PSA Level ≤ 20.00	High risk of prostate cancer
20.00 < PSA Level ≤ 50.00	Abnormally high
50.00 < PSA Level ≤ 100.00	Extreme risk of prostate cancer

Fig. 3. The Categories For PSA Level Groupings & The Corresponding PSA Levels Measured in ng/ml.

To calculate the distribution of PSA levels among all four cancer stages, Fig. 4 and Fig. 5 were used. To illustrate the calculations, below is a sample calculation that was used to obtain the percentage of stage 2 prostate cancer patients who had a functionally high PSA level before diagnosis.

Sample Calculation:

% of Patients = (Number of patients with stage 2 prostate cancer who had a functionally high PSA level ÷ Total number of patients in stage 2) **X** 100

= (4568 ÷ 6815) **X** 100
= 67.03% of patients in stage 2 had a functionally high PSA before diagnosis of prostate cancer.

Cancer Stage	Number of Patients
1	25
2	6815
3	688
4	217

Fig. 4. The Total Number of Prostate Cancer Patients In Each Cancer Stage

Cancer Stage	Normal	Elevated	Functionally High	High Risk of Prostate Cancer	Abnormally High	Extreme Risk of Prostate Cancer
1	8	5	9	3	0	0

2	516	534	4568	929	230	38
3	32	56	402	129	56	13
4	8	6	69	55	46	33

Fig. 5. The Number of Prostate Cancer Patients From Each Stage in Each of The PSA Level Groupings

Investigation & Results

The table below (figure 6) presents the distribution of PSA levels from the 6 different groups, among the 4 prostate cancer stages. All the tables and graphs were constructed using the raw data that was filtered and sorted in accordance with the inclusion criteria. Microsoft Excel was used to construct the graphs in Fig. 7. For reference, the full raw data table can be found in the appendix.

Color	PSA Level Groupings
	Normal
	Elevated
	Functionally high
	High risk of prostate cancer
	Abnormally high
	Extreme risk of prostate cancer

Fig. 6. A Key Indicating The Significance of Each Color Used In The Graph in Fig. 7

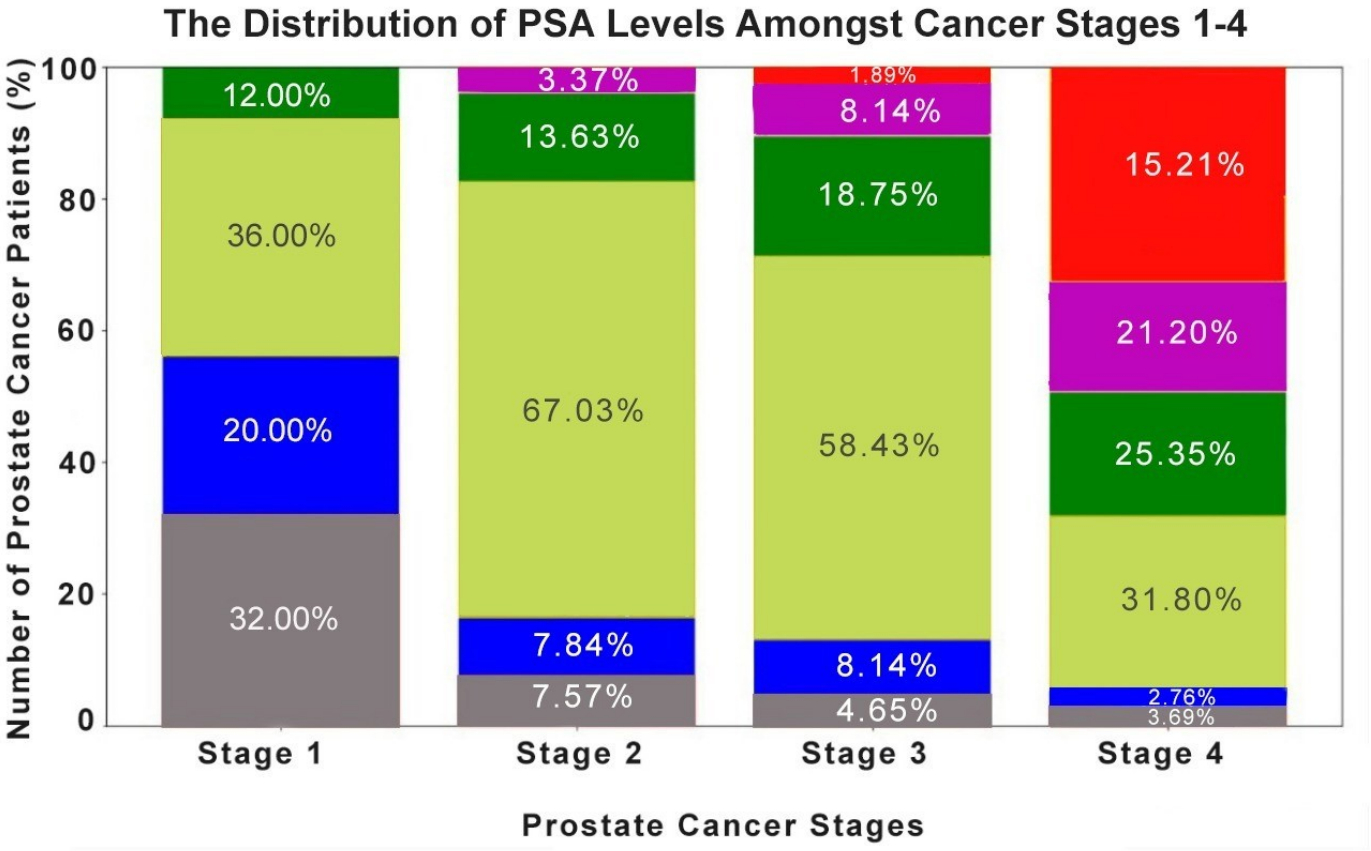

Fig. 7. Cumulative Bar Graph Representing The Distribution of PSA Levels Amongst Cancer Stages. The y-axis represents the % of prostate cancer patients with a specific PSA level. Each stacked bar represents a different PSA level grouping.

The cumulative bar graph in Fig. 7, shows the distribution of PSA levels among the four cancer stages. In this study, 32% of stage 1 cancer patients had normal PSA levels, and only 20% of stage 1 cancer patients had elevated PSA levels. As expected, 67.03% of stage two cancer patients and 58.43% stage three cancer patients have functionally high PSA levels. This makes sense, because stage two is neither a “safe” nor an aggressive form of prostate cancer, thus, the majority of patients affected with stage 2 cancer, will likely have high PSA levels, but only a minority are expected to exceed 10 ng/ml, which becomes extremely dangerous and alarming. As the groups shift to the categories with PSA levels above 10 ng/ml, a significant decrease in stage 1 and 2 cancer patients is witnessed. For instance, in the group with a high risk of prostate cancer, the highest percentage of cancer patients are from stage 4, 25.35%. A similar pattern repeats for the group with abnormally high PSA levels; stage 4 has the highest percentage of prostate cancer patients, 15.21%, while for stage 1 the percentage of cancer patients drops to 0%.

Statistical Tests

In this investigation, multiple descriptive statistics tests were conducted in order to accurately investigate the presence of a correlation between PSA levels and prostate cancer severity (i.e. cancer stages). It is inferable from Fig. 9 that as PSA levels increase cancer stage increases, because the mean PSA levels are continuously increasing. In addition, in Fig. 9 the mean PSA level for stage 4 increases greatly in comparison

to stages 1 to 3, which have more consistent increases between stages. Despite standard deviation not being a statistical test, it was also calculated and interpreted. Figure 8 shows the formula used by Excel. The sample standard deviation formula was used (STDEV.S). As shown in Fig. 9, stage 1 has a standard deviation value of 4.30 ng.ml, while for stage 4 it is 24.51 ng/ml. The standard deviation values indicate that as prostate cancer stage increases the data becomes more spread out. In addition, the low standard deviation values for stages 1, 2, and 3, indicate that the data provided for PSA levels is very reliable. Overall, these descriptive statistics suggest that there is a positive correlation between PSA levels and prostate cancer stage.

$$s = \sqrt{\frac{\sum(X - \bar{x})^2}{n - 1}}$$

Fig. 8. Sample Standard Deviation Formula

Descriptive Statistics of PSA Levels for Each Prostate Cancer Stage				
	Stage 1	**Stage 2**	**Stage 3**	**Stage 4**
Mean PSA Level	4.84	7.70	10.41	23.38
Median PSA Level	3.60	6.00	6.61	12.70
Standard Deviation of PSA Levels	4.30	7.22	11.64	24.51

Fig. 9. Descriptive Statistics Test Results for PSA Level of Each Cancer Stage

The graph in Fig. 10, clearly shows a positive correlation between PSA levels and prostate cancer stage in men older than 50, and a high coefficient of determination. The R^2 value of 0.8461 suggests that the positive correlation between both variables is strong, thus, as PSA levels increase, cancer stages increase. The error bars were created using the standard deviation values for each stage. Overall, the standard deviation error bars don't overlap entirely, however there is some overlapping for stages 1, 2, and 3, indicating that the data is statistically significant, meaning the relationship between both factors is real and not due to chance.

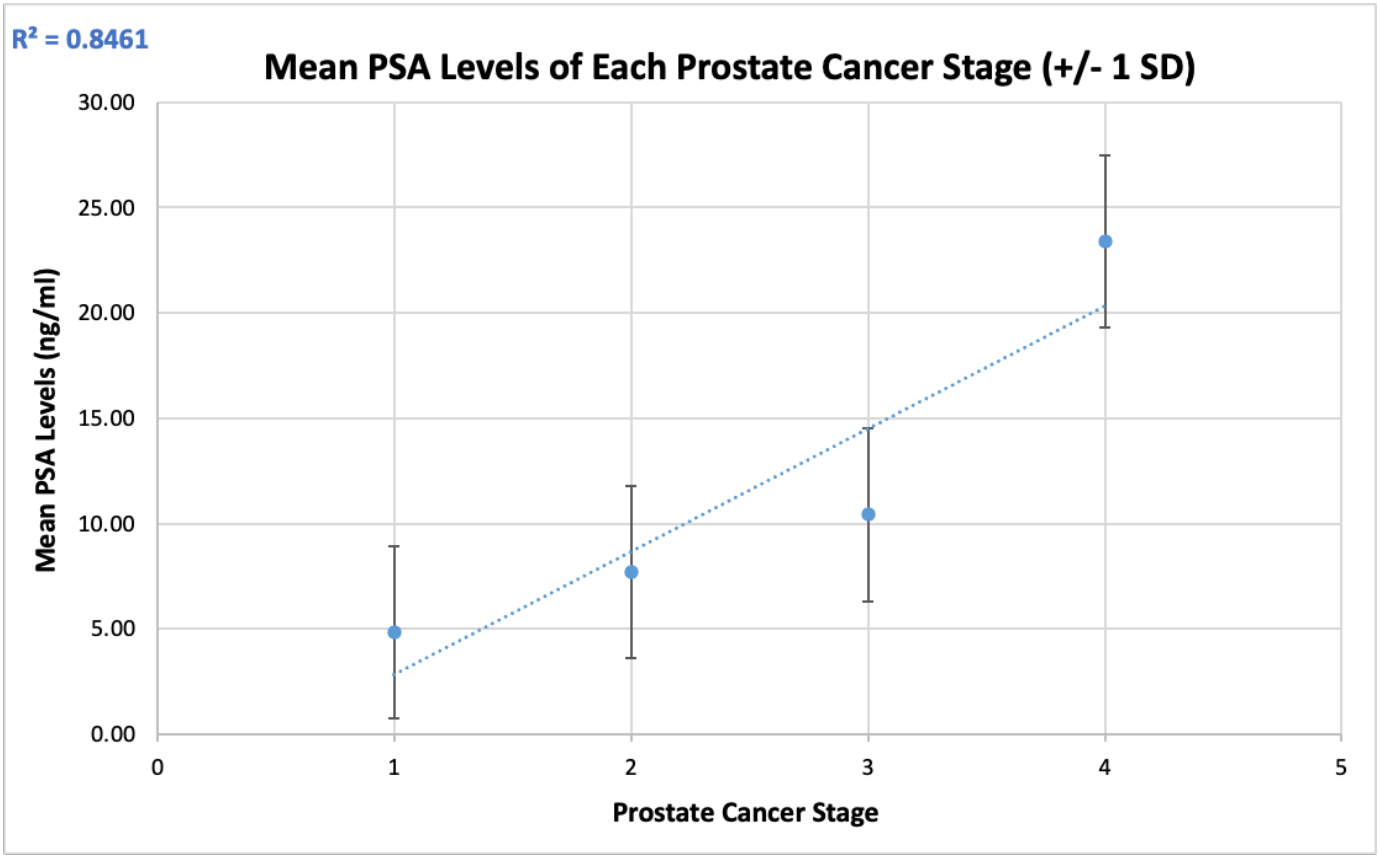

Fig. 10. Graph Displaying Mean PSA Levels In Each Prostate Cancer Stage

For further investigation of the positive correlation between PSA levels and prostate cancer stages, an ANOVA test was conducted. In Fig. 11 the P-value is approximately $4.97x10^{-5}$, F calculated value is 88,266, and the F critical value is 3.84. The P-value is less than 0.05, and the F critical value is very close to 3.9, suggesting the results are statistically significantly different (Bargen, Von). In addition, a very large F calculated value proposes that the data is of high variation, thus, the results are statistically significantly different (F-value (ANOVA) definition - isixsigma). Therefore, the P, F calculated, and F critical values obtained in this ANOVA test, further indicate that there is a positive correlation between PSA levels and prostate cancer stage.

ANOVA						
Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**F critical**
Between Groups	163090452	1	163090452	88266	4.96693E-05	3.8421
Within Groups	28905667	15644	1848			
Total	191996119	15645				

Fig. 11. ANOVA Statistical Test Results

Conclusion

Coming back to the research question: What is the correlation between PSA levels and prostate cancer severity in men older than the age of 50 in the US? It is evident from this investigation that the processed data supports the hypothesis: there is a positive correlation between PSA levels and prostate cancer stages in men aged 50 and older. The R^2 value of 0.8461, the P-value of $4.97x10^{-5}$, the F critical value of 3.84, and the F-value of 88,266, indicate that the relationship between PSA levels and prostate cancer stage is not due to chance, rather there is a relationship. In addition, the values are indicative of a positive correlation between

both variables, thus, supporting the hypothesis. The values for mean PSA level for each stage further support the hypothesis of this investigation, because as mean PSA levels increase, prostate cancer stage increases. Furthermore, there is a very large difference between the mean PSA levels of stage 1 prostate cancer and stage 4 prostate cancer, which supports the concept that higher PSA levels increase prostate cancer stage.

When comparing the results of this investigation with published scientific reports, the results are fairly similar, which further supports the hypothesis that there is a positive correlation between PSA levels and prostate cancer stages. An investigation that is similar to the study investigated in this paper, is a research report published in the National Library of Medicine, where the researchers investigated the correlation between PSA levels and the aggressiveness, diagnosis, and metastasis of prostate cancer. The investigation used data from 1116 patients who had a prostate biopsy and PSA level test, and the report included multiple statistical tests. After the investigation, the researchers found that "the data showed a strong correlation of PSA level with tumor diagnosis, tumor aggressiveness, and bone metastasis." (Lojanapiwat, Bannakij et al).

Evaluation

The main strength of this investigation is the reliability of the data used. The raw data was provided by a well established and trusted organization, National Cancer Institute, which guarantees that the data is accurate and was not tampered with. In addition, the data used was collected from approximately 7800 prostate cancer patients of various ethnicities, thus, enhancing the reliability and accuracy of the data. Furthermore, this investigation included a number of descriptive statistical values, which were used in a statistical analysis of the results obtained and highlighted a clear relationship between PSA levels and prostate cancer stage.

Despite the strengths of this investigation, there were few limitations and weaknesses. For example, the severity of prostate cancer was measured using cancer stages instead of using a more in-depth method, such as gleason scores. Such a limitation impacts the accuracy of the results, because a cancer stage only states a number from 1 to 4. Yet, within each stage there is a range of severities, which was not taken into account in this investigation. The investigation could be improved by incorporating other measurement tools. For instance, Gleason scores compare the condition of cancer cells to healthy cells; the more differentiated the cells are, the higher the severity of prostate cancer (Gleason grading system). In addition, the Tumor, Nodes, and Metastasis (TNM) grading system could be used to measure the severity of prostate cancer. The TNM grading system is able to identify the exact size of the tumor, how much has the cancer metastasized, the location that the cancer has spread to, and if the cancer has spread to nearby lymph nodes (*Prostate cancer - stages and grades*).

Another limitation is that the only country involved in this investigation was the US, for the reasons stated earlier. Such a limitation would make it more difficult to claim a final judgment as to whether there is a correlation between PSA levels and cancer stage, because results are not including men from other countries. Of course, the US is a very large country, thus, the data obtained and the results concluded are not as generalized, because the sample size was sufficient. To counter this limitation, the investigation could use data from multiple countries in different regions of the world.

Extensions

When further investigating the correlation between PSA levels and prostate cancer severity in men older than 50, there are various factors that could be implemented in the research to obtain results of higher importance. An interesting factor that could be investigated is race. Scientific research shows that men in the black

community have higher PSA levels and have a significant risk of being diagnosed with aggressive or metastatic prostate cancer. The reasons behind why race impacts the severity of prostate cancer remain undetermined (*Prostate cancer*). A potential research question for this new investigation could be: what is the retrospective link between race and prostate cancer severity in men older than 50? This investigation would have to include older men of various races, with an equal sample size of men for all races. Investigating this potential RQ would provide accurate results regarding which races are at a higher risk of being diagnosed with prostate cancer and the range of cancer severities they might experience.

Another factor that would be interesting to investigate is the prostate patient's age, nutrition, lifestyle, hormone medication, and smoking habits. All the variables listed may have effects on increasing PSA levels, thus, investigating the correlation between one or more of the listed variables with PSA levels, would provide medical research facilities and practitioners with valuable data, regarding the factors that increase the risk of prostate cancer ("What Factors Can Falsely Elevate or Lower Your Prostate-Specific Antigen (PSA)?").

Furthermore, an investigation into the reluctance of men worldwide to be examined by a urologist could be studied. When men are hesitant about checking up on their health, they increase their risk of being diagnosed with a more progressive stage of cancer, because the treatment would have been prolonged. A study conducted by American Journal of Men's Health, found that most men dislike digital rectal examinations mainly due to the discomfort it causes. In addition, multiple men feel as though getting a DRE is emasculating due to the procedure. Therefore, a potential research question could be: to what extent does the reluctance of men to undergo a digital rectal examination increase the risk of being diagnosed with progressive prostate cancer? (Winterich et al.).

Lastly, researching the retrospective link between a prostate cancer patient's socio-economic status and education level could be a fascinating topic. Men who come from a lower socio-economic background, are likely to be diagnosed with an advanced prostate cancer stage and experience a higher severity of prostate cancer, because they have less money available to undergo frequent medical examinations to check up on their health. In addition, the health care package they have access to is often of less quality, thus, the tests and medical resources available might be limited. The same applies to men with a lower education level, because they have not been taught about health check ups, living a healthy lifestyle, ways to fight prostate cancer while undergoing treatment, and have a bi-annual digital rectal examination to ensure their prostate is healthy. A potential research question could be: what is the retrospective link between socio-economic status, education, and prostate cancer severity?. Investigating this RQ will educate men of lower socio-economic status and education levels about regular check ups with a urologist, to ensure their body if free of cancer, their prostate is at a suitable size, and their PSA levels are normal (Seikkula, H.A., Kaipia, A.J., Ryynänen, H., Seppä, K., Pitkäniemi, J.M., Malila, N.K. and Boström, P.J. (2018)).

Bibliography

Bargen, Von. "Statistics." *BIOLOGY FOR LIFE*, Von Bargen, 2 Jan. 2017, www.biologyforlife.com/statistics.html.

byyuvaliami, Posted. "IB Biology Internal Assessment (23/24)." *Chiron*, 14 Aug. 2020, yuvalamichay.wordpress.com/2020/07/14/ib-biology-internal-assessment-23-24/.

Gupta, Nishi, et al. "Mutations in the Prostate Specific Antigen (PSA/KLK3) Correlate with Male Infertility." *Scientific Reports*, Nature Publishing Group UK, 11 Sept. 2017, www.ncbi.nlm.nih.gov/pmc/articles/PMC5593825/.

HALO Diagnostics. “Understanding PSA Levels & the PSA Test.” *HALO Diagnostics*, 20 Sept. 2022, www.halodx.com/blog/understanding-psa-levels-the-psa-test-halo-dx/.

Li, Yanping, et al. “Impact of Healthy Lifestyle Factors on Life Expectancies in the US Population.” *Circulation*, U.S. National Library of Medicine, 24 July 2018, www.ncbi.nlm.nih.gov/pmc/articles/PMC6207481/.

Lojanapiwat, Bannakij, et al. “Correlation and Diagnostic Performance of the Prostate-Specific Antigen Level with the Diagnosis, Aggressiveness, and Bone Metastasis of Prostate Cancer in Clinical Practice.” *Prostate International*, Asian Pacific Prostate Society (APPS), Sept. 2014, www.ncbi.nlm.nih.gov/pmc/articles/PMC4186957/.

McMillen, Matt. “Prostate Cancer in Older People.” Edited by Sabrina Felson, *WebMD*, WebMD, 12 May 2022, www.webmd.com/prostate-cancer/prostate-cancer-late-in-life-diagnosis.

“Prostate Cancer - Stages and Grades.” *Cancer.Net*, 6 Jan. 2022, www.cancer.net/cancer-types/prostate-cancer/stages-and-grades.

“PSA Test.” *Mayo Clinic*, Mayo Foundation for Medical Education and Research, 22 June 2021, www.mayoclinic.org/tests-procedures/psa-test/about/pac-20384731.

“Prostate Cancer.” *Mayo Clinic*, Mayo Foundation for Medical Education and Research, 1 Oct. 2022, www.mayoclinic.org/diseases-conditions/prostate-cancer/symptoms-causes/syc-20353087.

“Cancer.” *Mayo Clinic*, Mayo Foundation for Medical Education and Research, 27 Apr. 2021, www.mayoclinic.org/diseases-conditions/cancer/symptoms-causes/syc-20370588#:~:text=Cancer%20often%20has%20the%20 ability,cancer%20screening%2C%20treatment%20and%20prevention.

Ellis, Mary Ellen. “What Causes Prostate Cancer and What Are the Risk Factors?” *Healthline*, Healthline Media, 4 July 2018, www.healthline.com/health/prostate-cancer-risk-factors#causes.

“What Causes Prostate Cancer?” *American Cancer Society*, American Cancer Society, 1 Aug. 2019, www.cancer.org/cancer/prostate-cancer/causes-risks-prevention/what-causes.html.

“What Is Cancer?” *National Cancer Institute*, www.cancer.gov/about-cancer/understanding/what-is-cancer#:~:text=Cancer%20is%20a%20disease%20caused,ar e%20also%20called%20genetic%20changes.

“What Factors Can Falsely Elevate or Lower Your Prostate-Specific Antigen (PSA)?” *GoodRx*, GoodRx, www.goodrx.com/conditions/prostate-cancer/elevated-psa-levels.

“PSA Test.” *Mayo Clinic*, Mayo Foundation for Medical Education and Research, 22 June 2021, www.mayoclinic.org/tests-procedures/psa-test/about/pac-20384731.

Seikkula, H.A., Kaipia, A.J., Ryynänen, H., Seppä, K., Pitkäniemi, J.M., Malila, N.K. and Boström, P.J. (2018), The impact of socioeconomic status on stage specific prostate cancer survival and mortality before and after introduction of PSA test in Finland. Int. J. Cancer, 142: 891-898. https://doi.org/10.1002/ijc.31109.

Winterich, Julie A, et al. “Masculinity and the Body: How African American and White Men Experience Cancer Screening Exams Involving the Rectum.” *American Journal of Men's Health*, U.S. National Library of Medicine, Dec. 2009, www.ncbi.nlm.nih.gov/pmc/articles/PMC2662681/.

Lojanapiwat, Bannakij et al. “Correlation and diagnostic performance of the prostate-specific antigen level with the diagnosis, aggressiveness, and bone metastasis of prostate cancer in clinical practice.” Prostate international vol. 2,3 (2014): 133-9. doi:10.12954/PI.14054

"Prostate-Specific Antigen (PSA) Test." *National Cancer Institute*, NHI, 11 Mar. 2022, www.cancer.gov/types/prostate/psa-fact-sheet.

"Anatomy of the Urinary System." Anatomy of the Urinary System | Johns Hopkins Medicine, 19 Nov. 2019, www.hopkinsmedicine.org/health/wellness-and-prevention/anatomy-of-the-urinary-system#:~:text=from%20the%20ureters.-,Bladder.,empty%20urine%20through%20the%20urethra.

"Gleason Grading System." Mount Sinai Health System, www.mountsinai.org/health-library/special-topic/gleason-grading-system.

F-Value (ANOVA) Definition - Isixsigma. www.isixsigma.com/dictionary/f-value-anova/.

Anthony Carpi, Ph.D., and Ph.D. Anne E. Egger. "Uncertainty, Error, and Confidence: Process of Science." *Visionlearning*, Visionlearning, Inc., 12 Feb. 2017, www.visionlearning.com/en/library/Process-of-Science/49/Uncertainty-Error-and-Confidence/157.

"Scientific Uncertainty." *Discovery of Sound in the Sea*, 18 Dec. 2018, dosits.org/decision-makers/scientific-uncertainty/.

Appendix

Link To Full Data Transfer Agreement (legal authorization for use of data)

Sample of Data Transfer Agreement:

Please complete the information below:	
CDAS PROJECT NUMBER:	PLCO-1071
PROJECT TITLE:	Looking for the retrospective link between cigarette habits in men and prostate cancer
RECIPIENT:	American international School of riyadh
RECIPIENT LEAD INVESTIGATOR:	Tamara yaghi

The National Cancer Institute (NCI) and the RECIPIENT hereby enter into this Agreement for the transfer of data collected in the course of the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (DATA) to RECIPIENT through NCI's Cancer Data Access System (CDAS). Collectively or individually, NCI and RECIPIENT shall also be referred to as Parties or Party. This Agreement is effective as of the date of the last signature below (Effective Date).

In consideration of NCI providing DATA to RECIPIENT, RECIPIENT hereby agrees to the following terms and conditions:

1. DATA WILL NOT BE USED TO TREAT OR DIAGNOSE HUMAN SUBJECTS. RECIPIENT will use DATA in compliance with all applicable local, state, and/or federal laws and regulations, including but not limited to those for the protection of human subjects.

2. RECIPIENT must not use DATA for any study other than the approved Research Plan, attached as **Attachment 1**, unless RECIPIENT obtains the written consent of NCI by way of a new approved application through CDAS or by written and signed amendment to this Agreement. RECIPIENT grants NCI the right to publicly disclose the Research Plan, including titles, summaries or any other information contained therein as well as the names and contact information for the investigators conducting the research.

3. The DATA will be used solely by RECIPIENT LEAD INVESTIGATOR and RECIPIENT's faculty, employees, fellows, students, and agents that have a need to use, or provide a service in respect of, the DATA in connection with the Research Plan and whose obligations for using the DATA are consistent with the terms of this Agreement.

4. The DATA will not be further distributed to others without NCI's written consent. The RECIPIENT shall refer any request for the DATA to NCI.

5. Personally identifiable information will not be provided. If DATA being provided are coded, RECIPIENT will not request the key to the code. RECIPIENT must not attempt to learn the identity of or to contact the human subjects from which DATA were obtained, their physicians, or the collection sites for DATA. In the event that personally identifiable information is inadvertently transferred, RECIPIENT agrees to immediately destroy the personally identifiable information and report the circumstances to NCI. The DATA may be protected by the Federal Privacy Act and/or a Certificate of Confidentiality.

[Link To Full Raw Data Spreadsheet]

Sample Raw Data:

Sorted & Filtered Raw Data Table Displaying Prostate Cancer Stage, PSA Level, PSA Groupings		
Cancer Stage	**PSA Level (ng/ml)**	**PSA Group**
1	0.2	normal
1	0.36	normal
1	0.62	normal
1	0.77	normal
1	0.8	normal
1	0.9	normal
1	1	normal
1	1.37	normal
1	2.6	reasonable
1	2.7	reasonable
1	2.73	reasonable
1	3.1	reasonable
1	3.6	reasonable
1	4	suspicious
1	5.6	suspicious
1	5.6	suspicious
1	5.6	suspicious
1	6.35	suspicious
1	7.8	suspicious
1	7.84	suspicious
1	8.4	suspicious
1	8.6	suspicious
1	10.1	dangerous
1	14.21	dangerous
1	16.1	dangerous
2	0.05	normal
2	0.2	normal
2	0.3	normal
2	0.3	normal
2	0.3	normal
2	0.3	normal
2	0.32	normal
2	0.39	normal