Assessment of the Potential Risk Factors of Prostate Cancer: A Comparative Study

Rayyan Ibrahim Mukhtar Ahmed [*1], Ahmed Mohamed Abdalla Hamdi[2], Altaiyb Omer Ahmed Mohmmed [3]

[1]*(Department of Statistics, Faculty of Science/ Sudan University of Science & Technology, Sudan)*
[2]*(Department of Statistics, Faculty of Science/ Sudan University of Science & Technology, Sudan)*
[3]*(Department of Statistics, Faculty of Science/ Sudan University of Science & Technology, Sudan), P.O BOX 407*

Abstract: *The aim of this study is to compare between chi-square test and Mantel-Haenszel test and to assess the potential risk factors for prostate cancer incidence in order to set priorities for public heath interventions and to reduce the incidence of the disease. This study included patients with prostate cancer who were being treated at the National Center for Radiotherapy and Nuclear Medicine in Khartoum State, Sudan. 250 patients were chosen by interviews and from their medical history. The potential risk factors that significantly related to the incidence of prostate cancer were identified by using chi square test, Fisher exact test and Mantel-Haenszel test. These variables relative to their odds ratios were: prostate specific antigen or PSA (173.1), diseases (6.5), alcohol (8.1), weight (3.6), intake of green vegetables (3.8), animal fat (4.1) and marital status. By comparing the two tests, chi-square test showed the other significant variables that related to the disease, whereas the Mantel-Haenszel test showed that they were insignificant variables. These variables were (intake prostate medication, occupation, cholesterol, blood pressure and family history).*
Keywords: *Incidence, PSA, Risk factor, Odds Ratio, Fisher, Mantel-Haenszel*

I. Introduction:

Prostate cancer is one of the most common male cancers worldwide with over 650,000 men diagnosed each year [1]. Regarding cancer types, prostate cancer is the second most common cancer in men. An estimated 1.1 million men worldwide were diagnosed with prostate cancer in 2012. With an estimated 300,000 deaths in 2012, it is the fifth leading cause of death from cancer in men .In the GLOBOCAN 2012 reports prostate cancer incidence and mortality rates in Africa were reported to be 23.2 and 17.0 per 100,000 respectively [2]. Prostate cancer is considered second among cancers in Sudan [3]. The Estimated number of new cases and deaths per year was 599 and 509, respectively [4]. This high numbers may be the result of late detection, since studies show that 87 percent of men treated when their cancer is diagnosed early can expect to be alive in five years. Prostate Cancer should not be confused with benign prostate hypertrophy (BPH). BPH is the slow enlargement of the prostate gland that occurs in more than half of men over 45, and while it is not malignant condition prostate cancer is present in about 38 percent of men who undergo surgery to ease the symptoms caused by an enlarged prostate [5].

The main objective of this study is to compare between most tests used to analyze categorical data such as contingency tables or stratified tables. Also, this study aimed to assess the variables agreed by the Mantel-Haenszel test and the Chi-square test that affect on the incidence of prostate cancer.

II. The Chi-Square Distribution:

The chi-square distribution is the most frequently employed statistical technique for the analysis of count or frequency data. Karl Pearson described it in 1900 [6]. This is a continuous probability distribution given by positive values that are skewed to the right, because all values are between 0 and ∞. It cannot take on negative values, since it is the sum of values that have been squared [7]. The Pearson's χ^2 test is the most commonly used test for the difference in distribution of categorical variables between two or more independent groups. If we are interested in comparing the proportion of individuals with or without a particular characteristic between two groups, the null hypothesis states that there is no difference between these two proportions. The data can be arranged in a 2×2 contingency table, and will need a larger contingency table to arrange the data if there are more than two groups or the categorical variable of interest can take more than two possible values. The test requires calculating the expected frequencies, as:

$$E_{ii} = \frac{(n_{i.})(n_{.i})}{n} \rightarrow (1)$$

The Pearson χ^2 statistic is calculated as:

$$X^2 = \sum \left[\frac{(O-E)^2}{E} \right] \rightarrow (2)$$

With (r-1)*(c-1) degree of freedom (r: number of rows, c: number of columns) and O is the observed frequency, E is the expected frequency [6]. If the agreement between O and E is sufficiently poor (resulting in a sufficiently large X^2 value,) we will be able to reject H_0. The quantity $\sum[(O-E)^2/E]$ will be small if the observed and expected frequencies are close together and will be large if the differences are large. The computed value of X^2 is compared with the tabulated value χ^2. The decision rule is to reject H_0 if X^2 is greater than or equal to the tabulated value for the chosen value of α [7].

When testing for independence in a contingency table, the χ^2 distribution is used as an approximation to the discrete probability of the observed frequencies. When the total number of observations is small, the estimates of probabilities in each cell become inaccurate and the risk of type I error increases. When the expected frequencies are small, the approximation of the χ^2 statistic can be improved by a continuity correction known as Yates' correction. If the sample size is large, this correction will have little effect on the value of the test statistic [8].

III.   Fisher's Exact Test And Mantel–Haenszel Test:

Sometimes we have data that can be summarized in a 2×2 contingency table, but these data are derived from very small samples. The chi-square test is not an appropriate method of analysis if minimum expected frequency requirements are not met. In this situation the chi-square test should be avoided, and Fisher exact test can be used. It is called exact because if desired, it permits us to calculate the exact probability of obtaining the observed results or results that are more extreme. The Fisher's exact probability test is a test of the independence between two dichotomous categorical variables. It is used to assess the difference between two independent proportions when numbers are small, but cannot be applied to a contingency table larger than a two-dimensional one. For large sample size, the results of χ^2 test, Fisher exact test and Yates corrected χ^2 are similar. Fisher's test and Yates' correction give more conservative results than the conventional χ^2 test [8].

The chi-square test was extended by Nathan Mantel and William Haenszel in 1959 for the situation in which cases and controls have been stratified into subgroups to eliminate possibility of confounding by one or more variables (paper4). In case that the goal of the research is studying the relationship between the status of some disease and the status of some risk factor, another variable(a confounding variable) that may be associated with the (disease, risk factor or both) used to clarify the true relation between two variables. for accomplishing goal the Mantel–Haenszel procedure can be used. The null hypothesis is that there is no association between two variables. The Mantel–Haenszel procedure is appropriate for use with data from prospective and retrospective studies. In the application of the Mantel–Haenszel procedure, case and control subjects are assigned to strata corresponding to different values of the confounding variable. The confounding variable is categorical, with the different categories defining the strata. The Mantel–Haenszel test statistic $\chi^2{}_{MH}$ as:

$$\chi^2{}_{MH} = \frac{\left(\sum_{i=1}^{k} a_i - \sum_{i=1}^{k} e_i \right)^2}{\sum_{i=1}^{k} v_i} \rightarrow (3)$$

The odds ratio can be calculated from:

$$\widehat{OR}_{MH} = \frac{\sum_{i=1}^{k} (a_i d_i / n_i)}{\sum_{i=1}^{k} (b_i c_i / n_i)} \rightarrow (4)$$

Reject the null hypothesis of no association between the disease status and suspected risk factor status in the population if the computed value of $\chi^2{}_{MH}$ is equal to or greater than the critical value of the test statistic, which is the tabulated chi-square value for 1 degree of freedom and the chosen level of significance [7].

IV.   Study Design And Data Collection:

Chi-square test, Fisher exact test and Mantel Haenszel test had been applied to the dataset collected through patient's interviews and from their medical history. The data   consist of 15 independent variables: age, occupation, the state, marital status(maritstat), family history(famhist), eating red meats and animal fat regularly(animfat), eating green vegetables and fruits regularly(greenveg), suffering from overweight(weight), high cholesterol(cholesterol), high blood pressure(bloodpres), ingestion of prostate medication (prostatmed), alcohol consumption(alcohol), smoking, developing one or more of these diseases: "syphilis, gonorrhea, chronic prostatitis, prostate enlargement " (diseases) and prostate specific antigen (PSA). The case–control study was carried out in the National Center for Radiotherapy and Nuclear Medicine in Khartoum state, Sudan. The

sample size was 250 individuals; 150 cases (with prostate cancer) and 100 were controls (without prostate cancer). The outcome variable is (Diagnostic) represented the incidence of prostate cancer (1: Yes and 0: No).

V.   Methodology

The Mantel-Haenszel and the Pearson's χ^2 tests had been used to determine the variables most related to outcome variable (the incidence of prostate cancer). If chi-square assumptions had not been met, the Fisher exact test was used. Study subjects consisted of patients who are treated for prostate cancer during one year (2015 – 2016), everyday (317 days). The data of 250 individuals were collected through personal interviews and from their medical history. The risk factors which significantly associated with the outcome variable were identified by using NCSS11. Both tests agreed that these variables: (PSA, diseases, alcohol, weight, greenveg, animfat and maritstat) were significantly related to the disease.

VI.   Results Of Chi-Square Test:
Table1: Chi-Square Test of Independence:

Risk factor	Person's chi-square (χ^2 value)	Yates's Correction (χ^2 value)	df	Fisher exact (Prob. level)	Prob. Level	Reject H_0 at α = 0.05?
PSA	213.8670 †	210.0101	1	0.00000	0.00000	Yes
Age	194.9433	191.2548	1	0.00000	0.00000	Yes
Maritstat	103.5506 †	100.5623	1	0.00000	0.00000	Yes
Diseases	90.4197	87.9527	1	0.00000	0.00000	Yes
State	46.7504	*	3	*	0.00000	Yes
Alcohol	33.3068	31.6459	1	0.00000	0.00000	Yes
Prostatmed	30.0011	28.3393	1	0.00000	0.00000	Yes
Occupation	27.6701	26.2024	1	0.00000	0.00000	Yes
Greenveg	27.0563	25.7204	1	0.00000	0.00000	Yes
Cholesterol	24.4971 †	22.8479	1	0.00000	0.00000	Yes
Bloodpress	21.6627	20.1640	1	0.00000	0.00000	Yes
Animal fats	13.4549	12.3917	1	0.00035	0.00043	Yes
Famhist	9.6032	8.7500	1	0.00196	0.00310	Yes
Weight	6.6964	5.9730	1	0.00995	0.01453	Yes
Smoking	0.0431	0.0061	1	0.89692	0.93792	No

* Test computed only for 2×2 tables.

† WARNING: At least one cell had a value less than 5.

Chi-square test demonstrated that all variables were significantly associated with the outcome variable (diagnostic), (except the smoking variable was insignificant). In Table1, risk factors were arranged by their importance. The most important independent variable was PSA (χ^2 = 213.87, Yates's = 210.00) with p-value (0.00000). In case that the 2×2 contingency tables had cell counts less than 5, the Fisher exact test had been used to prove that there was association between the independent variables and the prostate cancer incidence instead of χ^2 test. So; the variables of PSA, marital status and cholesterol are associated to the variable (diagnostic) with fisher p-value (0.00000). These variables were illustrated in the table with the sign (†) associated with the value of χ^2. The second risk factor was age, there was strong association between age and outcome variable (χ^2 = 194.94, Yates's = 191.25) with p-value (0.00000). When the age became greater than or equal 50, this indicated that the individuals had strong risk factor for prostate cancer incidence. The results of this test showed that there were no association between smoking and the disease (χ^2 = 0.0431, Yates's = 0.0061) with p-value (0.938).

Chi-square test was accurate and suitable for most variables tested and which were classified into two categories or more. The variable "state" had 4 categories, this explains that why the Yates's correction and Fisher exact were inapplicable. So; the chi square test was used to prove the association to the disease. This variable is illustrated in the table with a sign (*)

VII.   Results Of Mantel-Haenszel Tests:
Table2: Mantel-Haenszel Statistics Section:

Risk factor	χ^2- value	df	Prob. Level	estimated odds ratio	Lower 95.0% C.L.	upper 95.0% C.L.	Reject H_0 at α = 0.05?
PSA	121.11	1	0.000000	173.0926	69.1265	433.4235	Yes
Age	54.68	1	0.000000	56.1608	19.3090	163.3455	Yes
Maritalstat	26.25	1	0.000000	0.0197	0.0044	0.0886	Yes
Diseases	13.77	1	0.000207	6.5452	2.4261	17.6574	Yes
Alcohol	10.05	1	0.001523	8.0659	2.2188	29.3216	Yes
Weight	5.33	1	0.021021	3.6003	1.2129	10.6870	Yes
Greenveg	4.75	1	0.029287	3.8013	1.1440	12.6305	Yes

Animal fats	4.05	1	0.044048	4.0880	1.0382	16.0963	Yes
Occupation	3.79	1	0.051448	4.2533	0.9909	18.2556	No
Cholesterol	2.77	1	0.096053	3.9273	0.7843	19.6670	No
Prostatmed	1.69	1	0.194113	0.5032	0.1784	1.4189	No
Familyhist	0.71	1	0.397929	1.7424	0.4809	6.3139	No
Smoking	0.30	1	0.586767	1.3525	0.4553	4.0176	No
Bloodpress	0.09	1	0.761938	0.8011	0.1909	3.3629	No

Mantel-Haenszel (M-H) statistics was illustrated in Table2, it showed that the following risk factors:(PSA, Age, marital status, diseases, alcohol, weight, intake of green vegetables , intake animal fat) were significantly associated with prostate cancer incidence (the outcome variable). Other variables that were not related to the outcome variable were: occupation, family history, cholesterol, blood pressure, intake of prostate medication and smoking with (p-value$>$ 0.05), the variable of state had been excluded; it consists of 4 categories. The Age variable is classified into two categories (strata), stratum1 represented age under 50 and stratum 2 represented age equal to or above 50. The highest (M-H) odds ratio was 173.1 with χ^2=121.11 and one degree of freedom (for the variable PSA) this implies that men with abnormal PSA were more susceptible to the disease when the age is greater than or equal 50. So, PSA was the most important risk factor. In Table2 risk factors were ranked by their importance.

VIII.   Discussion Of Results:

There was an agreement between the results of the chi-square test and Mantel-Haenszel test. The agreed risk factors were: PSA, age, diseases, alcohol, weight, intake of green vegetables, animal fats and marital status. Abnormal PSA increases the risk of the disease with very large odds ratio (173.1) and χ^2- values 121.11 and 210.0101 form Table1 and Table2, respectively. So, PSA was most important variable in this study. Other studies had reached similar results, Ernesto P. Esteban et al. [9]; conducted an analytical study of 218 Japanese patients. They had first developed a theoretical framework to study PSA dynamics for BPH and prostate cancer patients. This analytical study then was applied to obtain monograms for a better understanding of the relationship among PSA and tumor volume in Japanese men with proven BPH or proven prostate cancer. This novel approach which does not neglect PSA contribution due to BPH may provide new information useful for a better diagnostic and prognosis of prostatic diseases or localized prostate cancer. They provided a relationship among PSA, age, and tumor volume. Another study provided by, Swanson KR, et al. [10], developed a mathematical model for the dynamics of serum levels of PSA as a function of the tumor volume. Their model results show good agreement with experimental observations and provide an explanation for the existence of significant prostatic tumor mass despite a low-serum PSA. This result can be very useful in enhancing the use of serum PSA levels as a marker for cancer growth.

Also age was an important risk factor for prostate cancer incidence, with χ^2- values 191.2548 and 54.68 by using chi-square test and Mantel-Haenszel test respectively. Men above the age of 50 years were exposed to the disease 56.2 times than those who were younger, Table2. There were some studies that confirm validity of this study. Carter et al. [11], showed that 50% of men between 70 and 80 years of age showed histological evidence of malignancy. At that time risk of 42% for developing histological evidence of prostate cancer in 50-year-old men had been calculated. In men at this age, however, the risk of developing clinically significant disease was only 9.5%, and the risk of dying from prostate cancer was only 2.9%.

 Marital status had effect to the disease with (p-value$<$ 0.05) in both tables. Table2 illustrated that married men were more susceptible than unmarried. However Tyson MD et al. [12] found different results. They used Multivariate Cox regression techniques to study the relationship between marital status and prostate cancer and overall mortality. They concluded that marital status was an independent predictor of prostate cancer-specific mortality and overall mortality in men with prostate cancer. Unmarried men have a higher risk of prostate cancer-specific mortality compared to married men of similar age, race, stage, and tumor grade.

The variable of developing one or more of these diseases: "syphilis, gonorrhea, chronic prostatitis and prostate enlargement" was significantly related to the outcome variable. The values of chi-square test and Mantel-Haenszel test were 87.95 (p-value 0.0000) and 13.77 (p-value 0.0002). Men who suffered from one or more of the above diseases were exposed to prostate cancer 6.5 times who did not suffer. Sutcliffe S. et al. [13] conducted a study about gonorrhea, syphilis, clinical prostatitis, and the risk of prostate cancer. They were asked participants to report their history of gonorrhea, syphilis, and clinical prostatitis by mailed questionnaire. Of the 36,033 participants in this analysis, 2,263 were diagnosed with prostate cancer. No association was observed between gonorrhea [adjusted relative risk "RR" was 1.04; 95% confidence interval "CI",(0.79, 1.36)] or syphilis [RR was 1.06; 95% CI,(0.44, 2.59)] and prostate cancer. They were also observed association between clinical prostatitis and prostate cancer (RR, 1.08; 95% CI, 0.96-1.20), although a significant positive association was observed among younger men (<59 years) screened for prostate cancer (RR, 1.49; 95% CI, 1.08-2.06; (p-(interaction) = 0.006). Saiful Miah et al. [14] collected data from an online search and contemporary data

presented at international urological congresses. They found a relationship between benign prostatic hyperplasia (BPH) and prostate cancer.

In this study, the effect of alcohol consumption on prostate cancer incidence was observed. Men who consumed alcohol were exposed to the disease 8.1 times than those who did not, Table2. Clair Brunner et al. [15] found little evidence that variants in alcohol metabolizing genes were associated with prostate cancer diagnosis. The obesity or overweight was significantly related to the disease with χ^2- values 5.97 and 5.33, table 1 and 2 respectively. Men with obesity were more susceptible to the disease 3.6 times than normal men, with p-value less than 0.05 in both tables. Regarding with this point, a previous study [16] found that the obesity was a risk factor for cancer of prostate. Between 1960 and 1980 mortality data were collected from 6,763white men through a questionnaire on cohort members. Overweight men had a significantly higher risk of fatal prostate cancer than men near their desirable weight. The predicted relative risk of fatal prostate cancer was 2.5 for overweight men.

Intake of fruits and vegetables was important factor to decrease the risk of the disease incidence. This study showed that men who did not intake fruits and vegetables were more susceptible to the disease 3.8 times more than who did. The p-value was 0.000 and 0.0293 from Table1 and 2 respectively. Victoria et al. [17], evaluated the association between prostate cancer risk and intake of fruits and vegetables in 1338 patients with prostate cancer among 29361 men and Cox proportional hazards models were used. They demonstrated that intake of fruits and vegetables decreased the incidence of prostate cancer.

This study found association between animal fat intake and occurrence of prostate cancer. Men who did intake animal fat were exposed to the disease 4.1 times more than who did not. A previous study [18], clarified that the role of animal fat on the incidence of prostate cancer and indicated that it may act by shortening the latency period of the disease.

IX.   Conclusion

The variables agreed by the Mantel-Haeszel test and the Chi-square test that affect on the incidence of prostate cancer shown above were (PSA, age, diseases, alcohol, weight, intake of green vegetables, animal fat and marital status). The smoking was insignificantly related to the disease as shown by both tests with p-value 0.937 and 0.587, Table1 and Table2 respectively. Chi-square test showed the other significant variables that related to the disease, whereas the Mantel-Haenszel test showed that they were insignificant variables. These variables were (intake prostate medication, occupation, cholesterol, blood pressure and family history).

References

[1]   https://www.healthknowledge.org.uk/public-health-textbook/disease-causation-diagnostic/2b-epidemiology-diseases-phs/cancers/prostate-cancer. http://globocan.iarc.fr/old/Factsheets/cancers/prostate-new.asp.

[2]   Report of the Sudanese Federal Ministry of Health for 2016 https://pcafrica.wordpress.com/site-map/sudan/data/ http://training.seer.cancer.gov/prostate/intro/   [National cancer institute & SEER Training Modules]

[3]   John Ludbrook, Is there still a place for Pearson's chi-square test and Fisher exact test in surgical research, *ANZ journal of surgery, 81(12)*, 2011; 923-926.

[4]   Wayne W. Daniel, *Biostatistics* (United States Of America, Wiely, 2009), 9th edition.

[5]   Saumya Panda, Module on Biostatistics and research methodology for the Dermatologist, *IJD, 61(1)*, 2016; 385-392.

[6]   E.P. Ernesto, D. Giovanni, R. Jaileen and L.M. Stephanie, An analytical study of prostate-specific antigen dynamics, *computational and mathematical methods in medicine, Volume 2016 , Article ID 3929163*, 2016, 6 pages

[7]   Kristin R. Swanson, Lawrence D. True, Daniel W. Lin, Kent R. Buhler, Robert Vessella, and James D. Murray, A Quantitative Model for the Dynamics of Serum Prostate-Specific Antigen as a Marker for Cancerous Growth, *Am J Pathol, 158(6)*, 2001, 2195–2199.

[8]   Carter HB, Piantadosi S, Isaacs JT, Clinical evidence for the implication of the multistep development of prostate cancer, *J Urol, 143(4)*, 1990, 742-6.

[9]   Tyson MD[1], Andrews PE, Etzioni DA, Ferrigni RG, Humphreys MR, Swanson SK, Castle EK, Marital status and prostate cancer outcomes. *Can J Urol. 20(2)*, 2013 Apr, 6702-6.

[10]   Sutcliffe S, Giovannucci E, De Marzo AM, Leitzmann MF, Willett WC, Platz EA., Gonorrhea, syphilis, clinical prostatitis, and the risk of prostate cancer. ,*Cancer Epidemiol Biomarkers Prev, 15(11)*, 2006 Nov, 2160-6.

[11]   Saiful Miah and James Catto, BPH and prostate cancer risk, *Indian J Urol. 30(2)*, 2014 Apr-Jun, 214–218

[12]   Clair Brunner et al. Alcohol consumption and prostate cancer incidence and progression: A Mendelian randomization study, *Int J Cancer, 140(1)*, 2017 Jan, 175–85.

[13]   David A. Snowdon, Ronald L.Philips, Warren Choi, Diet, obesity, and risk of fatal prostate cancer, *Am J Epidemiol; 120 (2)*, 1984, 244-250.

[14]   Victoria A. Kirsh, Ulrike Peters, Susan T. Mayne, Amy F. Subar, Nilanjan Chatterjee, Christine C. Johnson , Prospective Study of Fruit and Vegetable Intake and Risk of Prostate Cancer, *J Natl Cancer Inst , 99 (15)*, 2007, 1200-1209

[15]   Le Marchand, Kolonel, Wilkens, Myers, Hirohata , Animal fat consumption and prostate cancer: a prospective study in Hawaii, *Epidemiology.5(3)*, 1994 May, 276-82.