

## A Device for Predicting Prostate Cancer Risk: A Logistic Regression

Prahara Yuri<sup>1\*</sup>, Sungsang Rochadi<sup>2</sup> and Raden Danarto<sup>2</sup>

<sup>1</sup>Department of Urology, Faculty of Medicine, Universitas Indonesia, Cipto Mangunkusumo Hospital, Jakarta, Indonesia

<sup>2</sup>Department of Urology, Faculty of Medicine, Gajah Mada University, Sardjito Hospital, Yogyakarta, Indonesia

\*Corresponding author: Prahara Yuri, Department of Urology, Faculty of Medicine, Universitas Indonesia, Cipto Mangunkusumo Hospital, Jakarta, Indonesia, Tel: +6221-3152892; E-mail: [prahara.yuri@gmail.com](mailto:prahara.yuri@gmail.com)

Received date: October 13, 2016, Accepted date: November 02, 2016, Published date: November 10, 2016

Copyright: © 2016 Yuri P, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Abstract

**Background:** Early detection of prostate cancer is a possible means of decreasing the mortality and increasing the quality of life.

**Methods:** We included 92 patients retrospectively in Sardjito Hospital. Patients received prostate biopsy due to having abnormal serum prostate specific antigen (PSA) level (>4 ng/ml) and DRE. The relationship between the possibility of prostate cancer and the following variables were evaluated including: age, PSA level, prostate volume, DRE finding and family history. By using chi-square analysis, multiple logistic regressions, receiver operating characteristic (ROC) curve were drawn based on the predictive scoring equation to predict the possibility of prostate cancer. Using the predictive equation, we design a nomogram for predicting prostate cancer risk called prostate cancer risk calculator. All analyses were performed with SPSS, version 18.0.

**Results:** We analyzed 92 patients with PSA >4 ng/ml. It showed the relationship between the possibility of prostate cancer and the following variables, including: age ( $p < 0.001$ ), PSA level ( $p < 0.001$ ), DRE finding ( $p < 0.001$ ), family history ( $p < 0.001$ ) and prostate volume ( $p = 0.04$ ). Using a predictive equation, we design a calculator for predicting prostate cancer followed by receiver-operating characteristic curve analysis, it showed the sensitivity 90.4% and specificity 85% in predicting the possibility of prostate cancer.

**Conclusion:** Age, prostate volume, PSA, DRE finding and family history are factors associated prostate cancer. They can be used as independent predictor to predict prostate cancer.

**Keywords:** Early detection; Logistic regression; Prostate cancer risk calculator

### Introduction

Prostate cancer is the second most commonly found cancer in the world and also the sixth most common cause of cancer death in newly found male cancer patient (14%) and from the total death caused by cancer in men in 2008 (6%) [1]. What exactly causes the initiation and progression of prostate cancer is still unidentified at the moment but a number of studies have mentioned that genetic, race, diet and environmental factors play important roles in the development of the above mentioned medical condition [2-4].

In the last 20 years to 30 years, the development of science have discovered factors that can probably aid in identifying low or high risk male individuals regarding the mounting number of prostate cancer with the right and proper screening tool [5]. Prostate specific antigen (PSA), is one of the prostate cancer screening indicator that is used initially. The fall in the incidence of prostate cancer between the year 1992 and 1995 occurred after PSA was used as a screening tool [6]. Nevertheless, the use of PSA in screening for prostate cancer is still controversial because of the high cost, bias, over diagnosis and overtreatment that is preventing clinical expertise from taking the right decision [6,7]. Yang et al. [8] discovered that the use of PSA for screening was ineffective to detect prostate cancer.

In the last decade, a number of nomograms have been developed to predict and help in making decision to perform biopsy on patients with prostate cancer. In general, a number of predictive factors which mainly used for predicting prostate cancer are including total PSA value, digital rectal examination (DRE) and patient's age. However, several other predictive factors may also be enclosed such as race, family history, previous prostate biopsy, prostate volume, negative history of positive findings in prostate biopsy, number of core biopsies and percentage of free-PSA [9].

Presently, calculators to predict prostate cancer are available across the globe which are called Prostate Cancer Prevention Trial-Risk Calculator (PCPT-RC), European Randomized Study of Screening for Prostate Cancer-Risk Calculator (ERSPC-RC) and Indonesian Prostate Cancer Risk Calculator (IPCRC) [2]. All of these tools are using two different populations but the identified factors can help to decide whether a particular is required to undergo biopsy. These factors include total PSA value, PSA velocity, PSA density, DRE, family history, race, patients age, prostate volume, ultrasonography (USG), and abnormal Magnetic Resonance Imaging (MRI), previous history of  $\alpha$ -reductase inhibitors consumption and biopsy. These calculators can assist specialists and general physician to determine the next steps to be taken, the possibility of performing biopsy and also the frequency of follow-up [2,7,10], making decision, determining the possibility to perform prostate biopsy.

The aim of this study is to develop a tool such as a calculator that could be used to predict the occurrence of prostate cancer in Sardjito Hospital, thus, helping clinician in, and scheduling appropriate outpatient visit for patients with high risk of prostate cancer.

## Materials and Methods

### Study population

This is a retrospective-analysis study using a case control design. We collected and reviewed them in order to evaluate the possibility of prostate cancers based on patient's age, PSA serum concentration, prostate volume, digital rectal examination, and family history in Sardjito Hospital Yogyakarta before 2015. We excluded patients below 40 years old and volume below 10 ml. PSA was measured using PSA Enzyme Immunoassay using PSA monoclonal antibody. The prostate was measured in three dimensions, and its volume was estimated using a modification of the prolate ellipsoid formula and recorded in  $cm^3$  ( $0.523$  [length (cm)  $\times$  width (cm)  $\times$  height (cm)]) by TAUS/TRUS. DRE was classified as normal or abnormal (any prostatic nodule or induration). The biopsy specimens were examined for the presence of cancer and were categorized using the Gleason score by a pathologist. All variables data were collected from medical record. High risk PCa was defined as clinical stage  $>T2b$  and/or Gleason score 7 and/or PSA  $>10.0$  ng/ml.

### Statistical analysis

The relationship between the possibility of prostate cancer and its variables were evaluated. The association of each factor with its diagnosis was assessed by simple logistic regression analysis. Multiple logistic regression analysis with backward selection was used to determine which factors were independent predictors of PCa in the model-building set. A prediction equation for prostate cancer prediction was developed based on the final logistic regression model. Multiple logistic regression analysis with a backward variable selection procedure was used to determine which factors were independent predictors of prostate cancer in the model-building set. PSA level, prostate volume, and age were log-transformed prior to analysis. A prediction equation for prostate cancer prediction was developed based on the final logistic regression model. We calculated a predictive

scoring equation to predict the possibility of PCa using chi-square analysis, Kolmogorov-Smirnov test, multiple logistic regression and receiver operating characteristic (ROC) curve. We regarded a P value  $<0.05$  as statistically significant.

The logistic model is as follows:

$$\ln(\text{odds}): \quad A_0 + \beta_1(\text{lpsa-lpsac}) + \beta_2(\text{lvol-lvolc}) + \beta_3(\text{lage-lagec}) + \beta_4(\text{DRE}) + \beta_5(\text{HYS})$$

Odds is defined as  $p/(1-p)$  where p (in the case of this example) is the probability to detect prostate cancer.  $\ln(\text{odds})$  is often abbreviated as logit.  $A_0$  (the model constant) and,  $\beta_1$  to  $\beta_5$  are the parameters to be estimated by fitting the model to the dataset. The latter four parameters are associated with the predictors PSA, prostate volume, age and DRE findings, respectively.  $\text{lpsa}$  is the  $^2\log(\text{PSA})$ ,  $\text{lpsac}$ =the mean value of  $^2\log(\text{PSA})$ .  $\text{lvol}$  is the  $^2\log(\text{prostate volume})$ ,  $\text{lvolc}$ =the mean value of  $^2\log(\text{vol})$ . Although, arbitrary in principle base 2 logarithms were used as they enable the interpretation of PSA and prostate volume in terms of doubling which is more pragmatic than using for example base 10 logarithms which would necessitate interpretation in terms of 10-fold increases. DRE is the outcome of the digital rectal exam, 0 is assumed to correspond to normal, 1 to abnormal finding. 4HYS is the family history and 1 for positive. All analyses were performed with SPSS, version 18.0.

## Results

In this study, we analyzed 92 patients with PSA  $>4$  ng/ml. From the analysis we had found that the average for age, PSA value, prostate volume and International Prostate Symptom Score (IPSS) are 69.39 years old; 35.97 ng/ml, 51.86  $cm^3$  and 21.05 respectively. In terms of the age of patients, there is no apparent difference between high risk and low risk groups of patients of developing prostate cancer (69.08 vs. 69.8 years old). Patients in low risk groups on average are older and have higher IPSS score when compared to high risk groups (57.685 vs. 47.25 years old and 21.57 vs. 20.6, respectively). Additionally, the average PSA value is higher in high risk groups when compared to low risk groups (52.27 vs. 16.08 ng/ml respectively). Roughly about 39.2% of patients in high risk groups have abnormal DRE findings whereas about 50% of them have positive family history for prostate cancer (Table 1).

Variables	All (n=92)	Low Risk Groups	High Risk Groups	p-value
Age (y), X/median	69.39/71	69.8/71.5	69.08/71	<0.001*
PSA (ng/ml), X/median	35.97/16	16.08/10.95	51.27/23.50	<0.001*
IPSS, mean/median	21.04/21	21.57/21	20.6/21	<0.001*
Prostate volume (ml), X/median	51.86/48.8	57.85/55.24	47.25/44.50	0.04*
		<b>Low risk groups (%)</b>	<b>High risk groups (%)</b>	
Abnormal digital rectal examination	43.5	4.3	39.2	<0.001#
Positive family history	57.6	7.6	50	<0.001#

#chi-square analysis, \*Kolmogorov-Smirnov test

**Table 1:** Characteristic of research variables.

The existing variables are divided into several categories (such as DRE and family history) and continuous variables (such as PSA, patient's age and prostate volume). These variables are analyzed as predictive factors such as age, PSA, patient's age and prostate volume were then analyzed using 2-log transformed and centered for construct a better prediction models.

Lpsa is  $^2\log(\text{PSA})$ , Lage is  $^2\log(\text{AGE})$ , lagec is the mean value of  $^2\log(\text{AGE})$ . Logistic regression analysis has shown lpsac=4.4, lvolc=5.6, lagec=6.1,  $\beta_0=3.8$ ,  $\beta_1=0.2$ ,  $\beta_2=2.3$ ,  $\beta_3=1.8$ ,  $\beta_4=2.2$  and  $\beta_5=4.6$  (Table 2).

Predictive factors	OR	95%CI	p-value
AGE, 2-log centered	4.11	1.02-16.5	0.046
PSA, 2-log centered	5.57	1.16-26.76	0.032
Prostate volume, 2-log centered	3.24	1.01-15.2	0.024
Digital rectal examination, 1/0	6.33	1.36-29.54	0.019
Family history, 1/0	12.03	3.09-46.89	<0.001

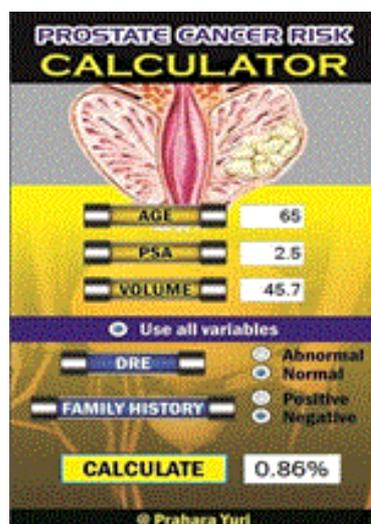
PSA=Prostate specific antigen (1=Positive; 0=Negative);  
Family history (1=Positive; 0=Negative)

**Table 2:** Multivariable Logistic regression analysis.

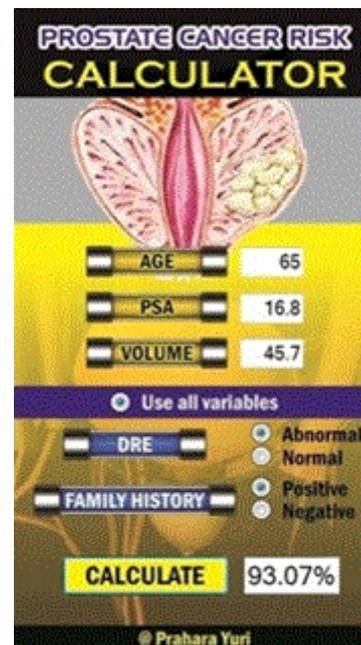
A logistic formula is thus constructed as:

$$\text{Logit} = 3.8 - 0.2 (\text{lpsa} - 4.4) - 2.3 (\text{lvol} - 5.6) - 1.8 (\text{lage} - 6.1) - 2.2 (\text{DRE}) - 4.6 (\text{HYS})$$

From multivariate analysis, we found that the sensitivity and specificity of this instrument to be 90.4% and 85% respectively.



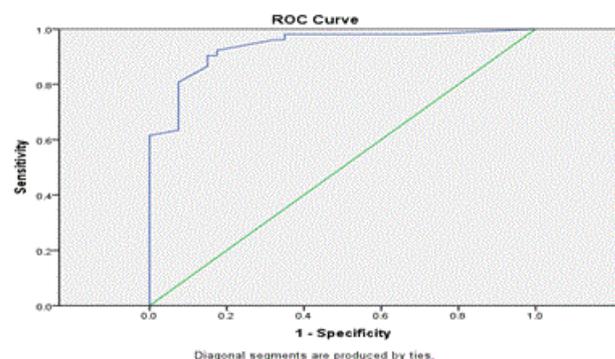
**Figure 1:** Prostate cancer risk calculator male patient, 65-year-old with PSA value of 2.5 ng/ml, and prostate volume 45.7 cm<sup>3</sup>. There are no abnormal findings during digital rectal examination and there is no family history of prostate cancer. The calculated risk for prostate cancer of this patient is 0.86%.



**Figure 2:** Prostate cancer risk calculator. Male patient, 65-year-old with PSA value of 16.8 ng/ml and prostate volume 45.7 cm<sup>3</sup>. DRE finding is abnormal and this patient also has positive family history of prostate cancer. The calculated risk for prostate cancer is 93.07%.

The cut off value was 48%. It means that the patients have the possibility of a positive biopsy result if this calculator showed value more than 48 % and it could be considered for urologist to perform prostate biopsy.

A prostate cancer predicting calculator is developed based on the results of logistic regression analysis of the predictive factors as stated in Table 2. This newly developed calculator is then applied using Microsoft Visual C#. An Example of the application of this calculator can be seen in case 1 and case 2 below (Figures 1-3).



**Figure 3:** ROC curve of prediction in having a positive biopsy.

## Discussion

The evaluation of risk factors of patients with prostate cancer is very beneficial to determine treatment course, patient's consultation and especially in confirming clinical diagnosis. The risk factors such PSA is inadequate as prostate cancer predictor due to its low specificity whereas DRE has a low sensitivity. Therefore, combination of these risk factors on every individual in the development of prostate cancer diagnosis are mandatory in order to providing better results in determining diagnosis and deciding whether it is required to perform prostate biopsy [8,11]. Recently, there has been a rapid increase of the incidence of prostate cancer in Indonesia due to an increase in PSA screenings even though the incidence in Asia is lower than in Western countries [3].

Henrnandez et al. [9] in their multivariate analysis showed that the increase in age, total PSA, low free PSA, abnormal DRE, African-American, positive family history, a large number of previous core biopsies and previous positive history of biopsy increases the risk for prostate cancer as well as for obtaining high-grade biopsy results. Up until now, there have been no clear grounds on when to start or to stop screening. Although PSA-based screening can reduce PCa-specific mortality, population-based PCa screening programs are not yet acceptable to many because of the high numbers of unnecessary tests and the detection of PCa that would never cause any harm (over diagnosis) [10].

A number of studies have been found that ERSPC-RC has a good predictive and strength and is more accurate in predicting the risk of developing prostate cancer when compared to PCPT-RC [1,3,10,12]. Nevertheless, no studies have been conducted to validate the application of these instruments in Asian countries. Lee et al. [3] had found that ERSPC-RC increased prostate cancer predictive capability when compared to PCPT-RC and PSA in a cohort study conducted in Korea. These differences, however, have not shown a statistically significant benefit in implementing these screening tools in a Korean Health centers.

There were two known common risk calculators for screening in the world; ERSPC-RC and PCPT-RC. ERSPC-RC is a better prediction tool of prostate cancer after biopsy than the PCPT-RC [9,13]. Several studies showed that the performance of the PCPT-RC for predicting prostate cancer is superior to the prediction accuracy of PSA testing alone [9,14] and ERSPC [13]. However, in several studies, PCPT-RC has been shown as not universally applicable in the population of men with elevated PSA (above 3.0 ng/mL) [15,16].

The PCPT-RC may overestimate the risk of finding prostate cancer [17]. This result could be due to that the PCPT-RC model was fitted on a population of primarily healthy men with PSA less than 3.0 ng/mL and above 55 years of age [18]. The accuracy of the PCPT-RC on such a healthy population of men are not ruled out by the current validation study since no cohorts of this type were included [16,17].

Both risk calculators tended to overestimate the risk of prostate cancer in the present study. Compared with the two risk calculators, the overestimation of the ERSPC-RC was lower than that of the PCPT-RC. The overestimation might be explained due to the differences between the cohorts as the source of the development for each risk calculators, even though the calibration plot was affected by multiple factors including several variables. Interestingly, the analysis of consistency of accuracy between the ERSPC-RC and PSAD showed that the ERSPC-RC more consistently predicts prostate cancer than PSAD. Despite this result, we cannot assume that the ERSPC-RC is

more useful than PSAD because there was no significant difference between the AUC values between the ERSPC-RC and PSAD [3].

Asia populations were also have a normogram to predict prostate cancer such as Indonesian Prostate Cancer Risk Calculator [2] and Korean Prostate Cancer Risk Calculator (KPCRC) [19]. From the logistic regression analysis as a model for IPCRC, there were some differences compared to other risk calculator. Coefficient for PSA in IPCRC was 0.62 lower than ERSPC, of 1.1 and PCPT 0.85. It was similar with prostate volume of 0.04 in IPCRC lower than ERSPC of 1.36. But IPCRC had higher coefficient in DRE findings of 3.99 than ERSPC of 0.8 and PCPT 0.91. The differences was due to low incidence of PCa in our population and most of our patients came in a more severe conditions and were not suitable for screening (eg. had urinary retention) [2]. The ROC analysis of IPCRC showed high sensitivity and specificity in predicting prostate cancer with area under curve (AUC) 0.938 (95%CI 0.93-0.95) in our study population. The AUC was higher than the PCPT (AUC 0.70) and the ERSPC (AUC 0.79) [15]. This indicated that IPCRC might be better in differentiating patient with PCa and BPH, but further validation and comparison in a larger population is still needed [2].

The effect of DRE and family history in our study in the clinical setting were stronger. This difference may be explained by interobserver variation for DRE outcome, and by the fact that in the clinical cohort more advanced PCa were found (PT2) than in the screening.

In this study, we added the family history that wasn't included in IPCRC. This study has several limitations. First, it has fewer numbers of samples in developed this calculator. Indonesian prostate cancer risk calculator was developed from 1957 men in Indonesian but PCPT was developed from 18,882 men while ERSPC developed from 6,288 men. The result may have been influenced by the heterogeneity of patients and biopsy technique. This device will be of limited use for general and finally, validation and comparison to other predictor devices is still needed.

## Conclusion

Family history, DRE, PSA level and age are the risk factors that are associated with the incidence of prostate cancer and can be considered to be independent factors to predict prostate cancer. For future studies, a prospective study with larger sample size may be conducted to evaluate the applicability of this calculator.

## References

1. Zhu X, Albertsen PC, Andriole GL, Roobol MJ, Schröder FH, et al. (2012) Risk-based prostate cancer screening. *Eur Urol* 61: 652-661.
2. Yuri P, Wangge G, Abshari F, Satjakoesoemah AI, Perdana NR, et al. (2015) Indonesian Prostate Cancer Risk Calculator (IPCRC). *Acta Med Indones* 47: 95-103.
3. Lee DH, Jung HB, Park JW, Kim KH, Kim J, et al. (2013) Can western based online prostate cancer risk calculators be used to predict prostate cancer after prostate biopsy for the Korean population? *Yonsei Med J* 54: 665-671.
4. Yuri P, Hendri AZ (2015) Association between tumor-associated macrophages and microvessel density on prostate cancer progression. *Prostate* 3: 2-7.
5. Kransse R, Roobol M, Schröder FH (2008) A graphical device to represent the outcomes of a logistic regression analysis. *Prostate* 68: 1674-1680.

6. Payne H, Cornford P (2011) Prostate-specific antigen: An evolving role in diagnosis, monitoring, and treatment evaluation in prostate cancer. *Urol Oncol Semin Orig Investig* 29: 593-601.
7. Stricker PD, Frydenberg M, Kneebone A, Chopra S (2012) Informed prostate cancer risk-adjusted testing: a new paradigm. *BJU Int* 110: 30-34.
8. Yang JB, Jeong BC, Seo SI, Jeon SS, Choi HY, et al. (2010) Outcome of prostate biopsy in men younger than 40 years of age with high prostate-specific antigen (PSA) levels. *Korean J Urol* 51: 21-24.
9. Hernandez DJ, Han M, Humphreys EB, Mangold LA, Taneja SS, et al. (2009) Predicting the outcome of prostate biopsy: Comparison of a novel logistic regression-based model, the prostate cancer risk calculator, and prostate-specific antigen level alone. *BJU Int* 103: 609-614.
10. Roobol MJ, Schroder FH, Bangma CH, Steyerberg EW (2012) A risk calculator for prostate cancer risk 4 years after negative screening findings: ERSPC Rotterdam. *Eur Urol* 11: 145.
11. Sfoungaristos S, Perimenis P (2012) PSA density is superior than PSA and Gleason score for adverse pathologic features prediction in patients with clinically localized prostate cancer. *J Can Urol Assoc* 6: 46-50.
12. Roobol MJ, van Vugt HA, Loeb S, Zhu X, Bul M, et al. (2012) Prediction of prostate cancer risk: The role of prostate volume and digital rectal examination in the ERSPC risk calculators. *Eur Urol* 61: 577-583.
13. Liang Y, Messer JC, Loudon C, Jimenez-Rios MA, Thompson IM, et al. (2013) Prostate cancer risk prediction in a urology clinic in Mexico. *Urol Oncol Semin Orig Investig* 31: 1085-1092.
14. Nguyen CT, Yu C, Moussa A, Kattan MW, Jones JS (2010) Performance of Prostate Cancer Prevention Trial Risk Calculator in a Contemporary Cohort Screened for Prostate Cancer and Diagnosed by Extended Prostate Biopsy. *J Urol* 183: 529-533.
15. Trotter G, Roobol MJ, Lawrentschuk N, Boström PJ, Fernandes KA, et al. (2011) Comparison of risk calculators from the Prostate Cancer Prevention Trial and the European Randomized Study of Screening for Prostate Cancer in a contemporary Canadian cohort. *BJU Int* 108: 1-8.
16. Liang Y, Ankerst DP, Feng Z, Fu R, Stanford JL, et al. (2013) The risk of biopsy-detectable prostate cancer using the prostate cancer prevention trial risk calculator in a community setting. *Urol Oncol Semin Orig Investig* 31: 1464-1469.
17. Ngo TC, Turnbull BB, Lavori PW, Presti JC (2011) The prostate cancer risk calculator from the prostate cancer prevention trial underestimates the risk of high grade cancer in contemporary referral patients. *J Urol* 185: 483-487.
18. Eyre SJ, Ankerst DP, Wei JT, Nair PV, Regan MM, et al. (2009) Validation in a Multiple Urology Practice Cohort of the Prostate Cancer Prevention Trial Calculator for Predicting Prostate Cancer Detection. *J Urol* 182: 2653-2658.
19. Park JY, Yoon S, Park MS, Cho DY, Park HS, et al. (2011) Initial biopsy outcome prediction in Korean patients-comparison of a novel web-based Korean prostate cancer risk calculator versus prostate-specific antigen testing. *J Korean Med Sci* 26: 85-91.