

RESEARCH

Open Access



Hospital-based prostate cancer screening in vietnamese men with lower urinary tract symptoms: a classification and regression tree model

Nguyen Chi Cuong^{1*}, Nguyen Truong Vien³, Nguyen Minh Thien¹, Phan Thanh Hai¹ and Tran Ngoc Dang^{2*}

Abstract

Background Prostate cancer (PCa) is a common disease in men over 65 years of age, and should be detected early, while reducing unnecessary biopsies. This study aims to construct a classification and regression tree (CART) model (i.e., risk stratification algorithm) using multivariable approach to select Vietnamese men with lower urinary tract symptoms (LUTS) for PCa biopsy.

Methods We conducted a case-control study on 260 men aged ≥ 50 years who visited MEDIC Medical Center, Vietnam in 2017–2018 with self-reported LUTS. The case group included patients with a positive biopsy and the control group included patients with a negative biopsy diagnosis of PCa. Bayesian Model Averaging (BMA) was used for selecting the most parsimonious prediction model. Then the CART with 5-fold cross-validation was constructed for selecting men who can benefit from PCa biopsy in steps by steps and intuitive way.

Results BMA suggested five potential prediction models, in which the most parsimonious model including PSA, I-PSS, and age. CART advised the following cut-off points in the marked screening sequence: $18 < \text{PSA} < 33.5 \text{ ng/mL}$, $\text{I-PSS} \geq 19$, and $\text{age} \geq 71$. Patients with $\text{PSA} \geq 33.5 \text{ ng/mL}$ have a PCa risk was 91.2%; patients with $\text{PSA} < 18 \text{ ng/mL}$ and $\text{I-PSS} < 19$ have a PCa risk was 7.1%. Patient with $18 \leq \text{PSA} < 33.5 \text{ ng/mL}$ and $\text{I-PSS} < 19$ have a PCa risk is 70% if $\text{age} \geq 71$; and is 16% if $\text{age} < 71$. In overall, CART reached high predictive value with $\text{AUC} = 0.915$. Sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of CART at the 20% diagnosis probability threshold were 91.5%, 86.2%, 86.9%, 91.2%, and 88.9% respectively; at 80% diagnosis probability threshold were 79.2%, 92.3%, 91.2%, 81.6%, and 85.8% respectively.

Conclusion CART combining PSA, I-PSS, and age has practical use in hospital-based PCa screening in Vietnamese men with lower urinary tract symptoms.

Keywords Prostate cancer, PSA, I-PSS, Vietnamese patients, Classification and regression tree, Bayesian modeling averaging

*Correspondence:
Nguyen Chi Cuong
drcuongmedic@gmail.com
Tran Ngoc Dang
ngocdangytcc@gmail.com

¹Ho Chi Minh City MEDIC Medical Center, Ho Chi Minh City, Vietnam
²Faculty of Public Health, University of Medicine and Pharmacy at Ho Chi Minh City, Ho Chi Minh City, Vietnam
³Pham Ngoc Thach University of Medicine, Ho Chi Minh City, Vietnam



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Prostate cancer (PCa) is common in men, especially in those aged 65 years and older. It has the second-highest incidence/prevalence (i.e., 30.7 per 100 000) and ranks fifth in cancer mortality rate among men (i.e., 7.7 per 100 000) worldwide [1]. In Vietnam, the incidence of PCa is 12.2 per 100 000, and the mortality rate is 5.1 per 100 000 as of 2020 [2]. Approximately 95–98% of PCa cases are adenocarcinomas that develop from adrenal duct cells [3]. PCa treatment depends primarily on the stage of development and the cell and patient characteristics. According to the American Cancer Society, PCa patients diagnosed at the localised or regional stage have a 5-year survival rate of over 90%. However, in the distant stage, the survival rate is only 30% [4]. Therefore, PCa should be detected at an early stage.

Prostate-specific antigen (PSA) is a serine protease in the kallikrein family and considered a tool for the screening and early detection of PCa [5]. It can help detect as early as nine years before having clinical symptoms [6]. There are two types of PCa screening studies using PSA, including population-based and hospital-based (or opportunistic) screenings [7]. The first type of screening deals with testing asymptomatic men with only PSA, and those with elevated PSA are immediately referred to biopsy. However, the latter type of screening involves testing men with some symptoms (e.g., lower urinary tract symptoms) using PSA and other clinical tools. Therefore, all men referred for biopsy in population-based screening are at lower risk of having PCa compared to that of hospital-based screening. PSA only based screening could account for 45–70% of the reduction in PCa mortality [8]; it could also induce the unnecessary biopsies [9]. In a 16 year follow-up of the European Randomized Study of Screening for Prostate Cancer (ERSPC), the unnecessary biopsy was 76% (i.e., 76% of elevated PSA cases have a negative biopsy) [10]. In addition, the optimal cut-off value of PSA for confirming PCa remains to be determined [11] [5]. In particular, even at a low level of PSA (that is, lower than 4 ng/mL), the false negative rate of PCa was high at 15%, whereas, at a high level of PSA (that is, higher than 10 ng/mL), the false positive rate was 50% [5].

In Vietnam, population-based PCa screening using PSA was conducted 12 years ago; it indicated a low prevalence of PCa (2.5%), but a high rate of medium grade lesions. The author also implied that the benefit of a mass screening program for PCa was not proven. Instead, a selective PCa screening in the usual care and at the hospital was superior in Vietnam. In hospital-based screening, combining clinical parameters, PSA, age, and other risk factors improved the prediction of prostate cancer [12–14]. International Prostate Symptom Score (I-PSS) is a screening scale for lower urinary tract symptoms and

is used to screen non-specific prostate gland abnormalities. For PCa screening, the I-PSS scale showed reasonable sensitivity (78%), but the specificity was not high (59.4%) [15]. A previous study showed that PSA screening performance varied with different I-PSS values. Therefore, combining PSA and I-PSS could improve the screening benefits [16]. There is, however, a paucity of such practical multivariable algorithm for hospital-based PCa screening in Vietnam.

The approach of PCa screening based on machine learning algorithms has only recently been applied. Algorithms including logistic regression, artificial neural networks, random forests, support vector machines, and extreme and light gradient boosting machines have been demonstrated to enhance PCa screening efficiency [13, 17–20]. However, these models do not help make clinical decisions in a step-to-step and intuitive manner. Classification and regression tree (CART) is an approach that allows physicians to apply results of the screening process directly and intuitively [17, 21].

Our study aimed to investigate the association of PSA, I-PSS, epidemiological and behavioural characteristics with PCa and then used these factors to construct a classification and regression tree (CART) algorithm to select Vietnamese men with lower urinary tract symptoms (LUTS) for PCa biopsy. The algorithm is expected to aid in reducing the probability of a negative prostate biopsy (i.e., unnecessary biopsy) while maintaining the ability to reduce PCa mortality for Vietnamese patients.

Methods

Study design and setting

We conducted a case-control study at the MEDIC Medical Center, Ho Chi Minh City, Vietnam. MEDIC is the first and top modernity private medical centre in Vietnam. Every day, more than 4,000 patients visit the centre for examination and treatment. The study was approved by the local institutional ethics committee of the MEDIC Center, and the opinion was signed on 15th July, 2016.

Participants

Our study participants were men aged ≥ 50 years who visited the MEDIC Centre in 2017–2018 with self-reported lower urinary tract symptoms. The inclusion criteria were abnormal lower urinary tract symptoms or enlarged prostate glands identified through DRE or ultrasound images. The exclusion criteria were acute prostatitis or refusal to participate in the study. All patients who meet the selection criteria were prescribed a biopsy. The case group was defined as having a positive biopsy result for PCa, and the control was defined as having a negative biopsy result. Biopsy based on 12-core Transrectal Ultrasound Guided Biopsy of the Prostate [22]. All patients

provided written informed consent before participating in the study.

Sample size

The minimum sample size estimated for each group of case-control studies was 116 patients to provide 90% power and 5% type I error to detect an odds ratio of 2.5. In Vietnam, PSA ≥ 10 ng/mL is considered as the high-risk group of PCa. Therefore, we chose the proportion of PSA ≥ 10 ng/mL equal to 23% in the control group as the proportion of controls with exposure in the sample size formula [23]. Our studies selected 130 patients for each group to ensure larger than the minimum sample size, hence the total patients was 260.

Data collection and variables' definition

We collected epidemiological and behavioural characteristics through interviews using a questionnaire and collected clinical and subclinical information from medical records. Epidemiological characteristics included age, number of children, overweight/obesity (BMI ≥ 23 kg/m² [24]), family history of PCa, existence of urinary tract diseases, history of urinary surgery, benign prostate hyperplasia, and exposure to agrochemicals. Lifestyle behaviours included physical activities (≥ 150 min/week for moderate or vigorous intensity [25]), current tobacco smoking, and heavy drinking (binge drinking (i.e., five drinks or more per occasion) on five or more days in the past month [26]). Food consumption behaviours were determined by the frequency of different food consumption types, including red meat, fruits, vegetables, nuts, vegetable oil, tea, and coffee.

International Prostate Symptom Score (I-PSS) was used to assess seven lower urinary tract symptoms: incomplete emptying, frequency, intermittency, urgency, weak stream, straining, and nocturia. I-PSS Vietnamese version was published by the Vietnamese Ministry of Health and recommended for Benign Prostatic Hyperplasia assessment. Each item of I-PSS was classified on a zero to five scale, reflecting the severity of each symptom [27].

Epidemiological and behavioural characteristics and I-PSS were assessed by only one oncologist for consistency. The oncologist was trained for conducting interviews before joining the study. The questionnaire consisted of ten interviews in a pilot sample for structure and content adaptation.

Serum PSA was quantified by a 2-step immunoassay using light-emitting microparticle technology (CMIA) with Alinity CiCi (Abbott) testing machine system. The testing machine system was calibrated, and quality control was performed at least once every day or when changing the reagent batch [28].

Statistical analysis

Descriptive analysis

Frequency and percentage were used to describe qualitative variables, including overweight/obesity, family history of PCa, existence of urinary tract diseases, history of urinary surgery, benign prostate hyperplasia, and exposure to agrochemicals, lifestyle behaviour, and food consumption behaviour. The median and quartiles were used to describe quantitative variables, including I-PSS, PSA concentration, and age. All descriptive analyses were stratified by the case and control groups.

Univariable logistic regression

A univariable logistic regression model was used to screen independent variables that were likely to be associated with PCa. The I-PSS score, PSA concentration, epidemiological characteristics, lifestyle behaviour, and food consumption behaviour were tested for association with PCa.

Bayesian model averaging (BMA) for model prediction

A BMA approach was used to search for the most parsimonious model for PCa prediction (i.e., minimum explanatory variables and maximum discrimination power) using PSA, I-PSS, epidemiological and behavioural variables. In summary, if there is n variables, there will be 2^n possible models constructing from n variables (not including interactive terms). BMA will construct all possible parsimonious prediction models based on the Bayesian Information Criteria (BIC), and posterior probabilities of these models. The final model with the high practical use in the clinical setting can be chosen based on BMA suggesting and clinical considerations.

CART model for PCa screening

CART was performed using the *rpart* function in the *rpart* package, R language (version 4.0.3). Five-folds cross validation was used to training and testing CART model. All independent variables became CART input in this process. The CART pruning was controlled by the maximum depth of the tree set to 4 to construct a reasonable complexity, the minimum number of observations was allowed to be 10 at each node to ensure sufficient supporting data. Diagnosis values of CART including sensitivity, specificity, positive predictive value, negative predictive value, and accuracy (1 – misclassification error) at the 20%, 50% and 80% probability cut-off were extracted.

Results

Association of epidemiological and behavioural characteristics with prostate cancer

There were total 260 patients (130 in case group vs. 130 in control group) included in the study. The median age

Table 1 Association of epidemiological and behavioural characteristics with Prostate cancer – univariable logistic regression

	Cases (n = 130) No. (%)	Controls (n = 130) No. (%)	OR (95% CI)	p-value
Epidemiological characteristics				
Age (median [25–75 percentile])	71 (64–78)	66 (61–71)	1.07 (1.04–1.11)	<0.001
Overweight or Obesity (BMI ≥ 23)	63 (48.5)	49 (37.7)	1.55 (0.92–2.63)	0.103
Family history of Prostate cancer (yes)	2 (1.5)	5 (3.9)	0.39 (0.04–2.45)	0.447
Existing of urinary tract diseases (yes)	17 (13.1)	20 (15.4)	0.82 (0.39–1.76)	0.723
History of urinary surgery (yes)	10 (7.8)	12 (9.2)	0.83 (0.31–2.18)	0.824
Benign Prostate Hyperplasia (yes)	14 (10.8)	24 (18.5)	0.53 (0.24–1.14)	0.113
Exposed to agrochemicals (yes)	40 (30.8)	20 (15.4)	2.44 (1.29–4.73)	0.005
Lifestyle behaviour				
Physical activity (yes)	68 (52.3)	86 (66.2)	0.65 (0.33–0.95)	0.032
Current tobacco smoking (yes)	62 (47.7)	48 (36.9)	1.56 (0.92–2.64)	0.103
Heavy drinking (yes)	14 (10.8)	12 (9.2)	1.19 (0.53–2.67)	0.680
Food consumption behaviour				
Red meat (≥ 3 times/week)	107 (82.3)	106 (81.5)	1.05 (0.53–2.08)	1.000
Fruits (≥ 3 times/week)	65 (50.0)	83 (63.9)	0.57 (0.33–0.96)	0.033
Vegetables (≥ 3 times/week)	111 (85.4)	120 (92.3)	0.49 (0.19–1.16)	0.114
Nuts (≥ 3 times/week)	9 (6.9)	9 (6.9)	1.00 (0.34–2.95)	1.000
Vegetable oil (≥ 3 times/week)	104 (80.0)	95 (73.1)	1.47 (0.79–2.75)	0.242
Tea (≥ 3 times/week)	56 (43.1)	53 (40.8)	1.10 (0.67–1.80)	0.706
Coffee (≥ 3 times/week)	82 (63.1)	80 (61.5)	1.07 (0.63–1.82)	0.898

Table 2 Association of I-PSS and PSA with prostate cancer - univariable logistic regression

	Cases (n = 130) Median (25–75 percentile)	Controls (n = 130) Median (25–75 percentile)	OR (95% CI)	p
PSA concentration (ng/mL)	89.5 (39.1–100)	12 (8–19)	1.06 (1.05–1.08)	<0.001
I-PSS (total score)	14 (8–19)	6.5 (3–12)	1.14 (1.09–1.18)	<0.001
Incomplete Emptying	4 (1–5)	1 (1–5)	1.27 (1.12–1.44)	
Frequency (every 2 h)	2 (0–2)	0 (0–3)	1.22 (1.08–1.37)	
Intermittency	0 (0–1)	0 (0–1)	1.29 (1.02–1.64)	
Urgency	2 (0–5)	0 (0–1)	1.41 (1.23–1.62)	
Weak Stream	0 (0–2)	0 (0–0)	1.32 (1.12–1.55)	
Straining	0 (0–1)	0 (0–1)	1.16 (0.98–1.38)	
Nocturia	4 (3–5)	2 (1–3)	2.22 (1.79–2.74)	

of cases was significantly higher than the controls (71 vs. 61). Based on univariable logistic regression, the risk factors of PCa included age, exposure to agricultural chemicals. Physical exercise and fruit consumption were noted as protective factors of PCa (Table 1).

Association of I-PSS and PSA with prostate cancer

The PCa odds ratio for 1 ng/mL PSA increase was 1.06 (95% CI, 1.05–1.08). The PCa odds ratio for each I-PSS point increase was 1.14 (95% CI, 1.09–1.18). All the items of I-PSS significantly associated with PCa, except for the “Straining” item with lowest OR=1.16 (95% CI, 0.98–1.38). “Nocturia” item had highest OR with 2.22 (95% CI, 1.79–2.74), and “Urgency” item came to the second with OR=1.41 (95% CI, 1.23–1.62) (Table 2).

BMA for PCa prediction

To determine whether PCa could be predicted by PSA, I-PSS, epidemiological and behavioural variables. There were 27 models suggested from BMA process, among them the best 5 models are shown in Table 3. The most parsimonious model (i.e., minimum explanatory variables and maximum discrimination power) included two variables: I-PSS, and PSA concentration. The second parsimonious model contained I-PSS, PSA, and age. The Area Under the ROC Curve (AUC) of the most parsimonious model was not much different compared to the second parsimonious model (0.931 vs. 0.929). Because age is an important factor for PCa screening and diagnosis in many previous studies [29–31], it is also a critical factor for disease mechanisms from a clinical standpoint. Therefore, we chose the second model with

Table 3 BMA prediction models using I-PSS, PSA, epidemiological, and behavioural characteristics for PCa

Model	Variable	OR (95% CI)	p	R ² (%)	AUC	BIC	Posterior probability
1	IPSS	1.12 (1.06–1.18)	<0.001	49.2	0.931	199.9	0.364
	PSA	1.06 (1.04–1.08)	<0.001				
	Intercept	0.03 (0.01–0.08)	<0.001				
2	IPSS	1.11 (1.05–1.17)	<0.001	50.3	0.929	201.4	0.174
	PSA	1.06 (1.04–1.07)	<0.001				
	Age	1.06 (1.01–1.10)	0.047				
	Intercept	0.01 (0.00–0.04)	<0.001				
3	IPSS	1.13 (1.07–1.19)	<0.001	50.1	0.929	205.5	0.120
	PSA	1.06 (1.04–1.08)	<0.001				
	Fruits (≥ 3 times/week)	0.50 (0.23–1.06)	0.069				
	Intercept	0.05 (0.02–0.12)	<0.001				
4	IPSS	1.12 (1.06–1.19)	<0.001	49.7	0.931	203.6	0.057
	PSA	1.06 (1.04–1.08)	<0.001				
	Overweight or Obesity	1.68 (0.80–3.53)	0.174				
	Intercept	0.03 (0.01–0.07)	<0.001				
5	IPSS	1.12 (1.06–1.18)	<0.001	49.7	0.931	203.7	0.056
	PSA	1.06 (1.04–1.08)	<0.001				
	Vegetable oil (≥ 3 times/week)	1.84 (0.75–4.52)	0.186				
	Intercept	0.02 (0.01–0.07)	<0.001				

three variables as the best model to use in the clinical setting (Table 3). This final model is also in light with the final model suggested by CART algorithm (details shown below).

CART for PCa screening

CART was deployed with all independent variables input for PCa screening, and the final model is shown in (Fig. 1).

The results indicated that PSA, I-PSS, and age played important roles in PCa screening. CART advised the following cut-off points in the marked screening sequence: $18 < \text{PSA} < 33.5$ ng/mL, $\text{I-PSS} \geq 19$, and $\text{age} \geq 71$. Patients with $\text{PSA} \geq 33.5$ ng/mL have a PCa risk was 91.2%; patients with $\text{PSA} < 18$ ng/mL and $\text{I-PSS} < 19$ have a PCa risk was 7.1%. Patient with $18 \leq \text{PSA} < 33.5$ ng/mL and $\text{I-PSS} < 19$ have a PCa risk is 70% if $\text{age} \geq 71$; and is 16% if $\text{age} < 71$.

In overall, CART reached high predictive value with $\text{AUC} = 0.915$. Sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of CART at the 20% diagnosis probability threshold were 91.5%, 86.2%, 86.9%, 91.2% and 88.9% respectively; at 80% diagnosis probability threshold were 79.2%, 92.3%, 91.2%, 81.6%, and 85.8% respectively (Table 4).

Discussion

Epidemiological and behavioral characteristics

The study included 260 observations at the Medic Center HCMC with 130 in the case group and 130 in control group. Based on univariable logistic regression, the risk factors of PCa included age, exposure to agricultural

chemicals and protective factor included physical exercise and fruit consumption.

Previous studies found that farming and exposure to agricultural chemicals are risk factors for PCa but not for all agricultural chemicals [29, 30]. Exposure to a few specific pesticides including fonofos, malathion, terbufos, and azinphos-methyl, dimethoate associated with PCa [31–33]. Genomic analysis showed pesticides might interact with genetic variants in pathways related to neurotransmission release in PCa patients [34, 35]. Therefore, the relationship between exposure to agricultural chemicals and PCa is plausible. Further epidemiological and mechanism studies are needed to identify the relationships of PCa with specific agricultural chemicals, in particular in the Vietnamese context.

Although our study initially found a relationship between physical exercise and PCa, there is a lack of evidence in the literature. Recent review and meta-analyses reveal that the association between regular physical activity and a low risk of prostate cancer remains elusive [36, 37]. Given also many general health benefits of physical activity, there is the need to clarify the role of physical activity in association with PCa in further studies.

The association of fruit consumption with PCa was shown in our study and recent studies [38, 39]. Total fruit intake significantly reduced PCa risk. However, our study did not analyze fruit subtypes. A previous study found that citrus fruit consumption is associated with PCa, but other fruit subtypes are not associated [38]. This relationship might be due to the anti-carcinogenic properties of vitamins and phytochemicals in citrus fruits [40, 41].

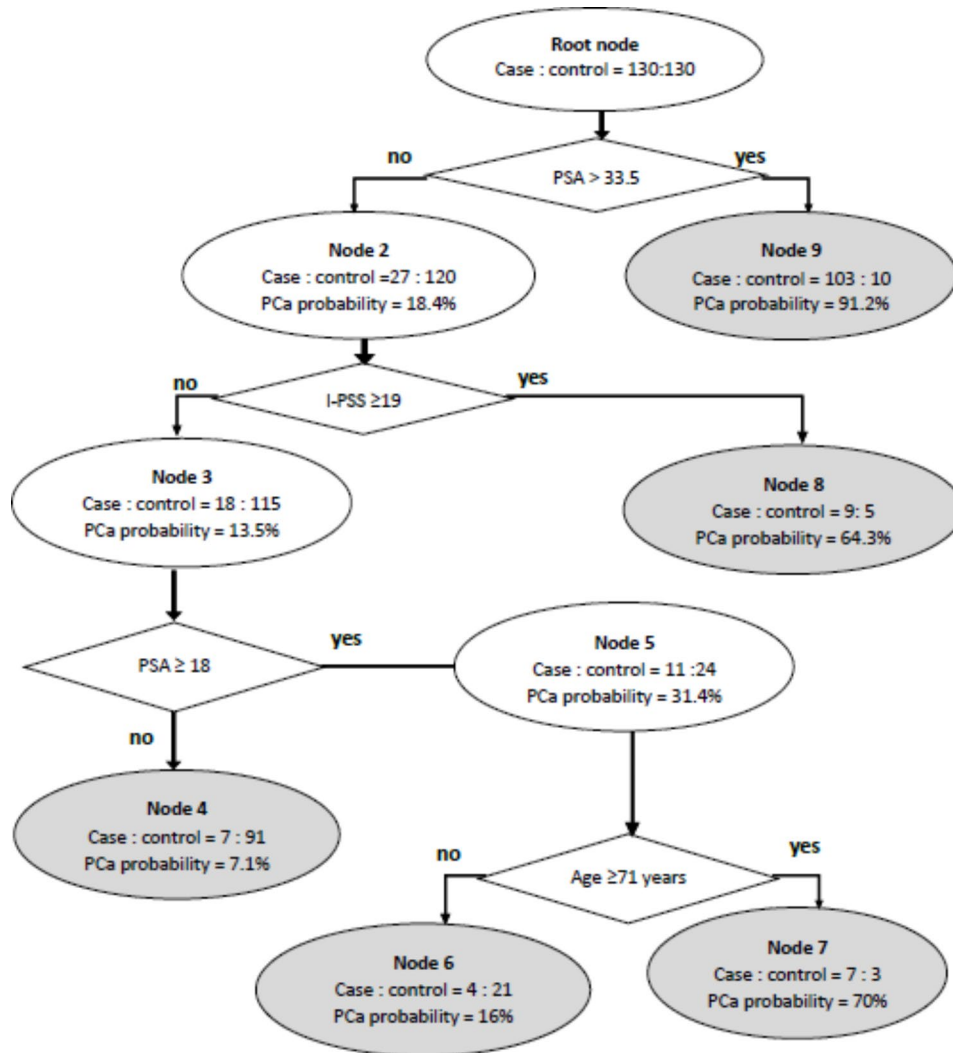


Fig. 1 Trained CART in prostate cancer screening

Table 4 Diagnosis values of CART in PCa screening

	Probabil- ity=20%	Probabil- ity=50%	Probabil- ity=80%
Sensitivity	0.915	0.915	0.792
Specificity	0.862	0.862	0.923
Positive predictive value	0.869	0.869	0.912
Negative predictive value	0.912	0.912	0.816
Accuracy	0.889	0.889	0.858

However, the causal relationship remains unclear because most findings are based on a cross-sectional study.

CART model suggested that age is most important in epidemiological and behavioral characteristics. Therefore, age, PSA, and IPSS were combined in PCa screening CART model. Our study showed that 50% of patients were older than 71 years in cases and older than 66 in controls. PCa is a disease that commonly occurs in the elderly men. Previous studies found that 75–80% of new cases occur in men aged over 65 years [42, 43]. Another

study in the United States had an average participant’s age was 66 years [44]. A study by European Association of Urology showed that PCa rarely occurred in men younger than 50 years; it also indicated that the median age of PCa patient was 70 years [45]. Giwerzman et al. showed that age was the closest risk factor of PCa [46]. Similarly, our study detected age as an independent risk factor of PCa. According to the Bayesian Model Averaging process, the PCa risk increased by 6% each year of age increased.

The role of I-PSS in PCa screening.

The Prostate Symptom Score (I-PSS) with seven recommended questions became an international standard to assess the symptoms of urination dysfunction in patients during the previous month. This scale is able to monitor changes in symptoms over time or after an intervention. A symptom severity assessment with an I-PSS scale is an important part of the initial evaluation, diagnosis, prediction, and monitoring of response

to treatment [47, 48]. Our study noted that I-PSS was used to detect the symptoms of PCa ($p < 0.001$) in both univariable and multivariable analyses. A cohort study by Martin et al. [49] detected an association between I-PSS and PCa. For overall PCa, men with $I-PSS \geq 20$ had a 2.26-fold increased risk of PCa compared to those with no symptoms. For localised PCa, men with $I-PSS \geq 20$ had a 4.6-fold increased risk of PCa compared to those with no symptoms [49]. A study by Hosseini et al. [15] showed an association between I-PSS and PCa. The mean I-PSS score of the PCa group was 16.05 and higher than that of the non-PCa group, with a mean I-PSS score of 6.84. The prevalence of patients with $I-PSS \geq 20$ in the PCa group was 30.3%, which was higher than that in the non-PCa group. The sensitivity and specificity of I-PSS at cut-off $I-PSS \geq 20$ were 78% and 59%, respectively [15]. Our study and data have provided evidence about the relationship of PCa screening value with I-PSS. In Vietnam, according to the Ministry of Health, I-PSS has not yet been recommended for PCa initial screening; however, I-PSS has been recommended for benign hypertrophy of prostate – a disease that has many symptoms similar to early-stage PCa symptoms. Our study recommended using I-PSS for initial screening for any patient who has self-reported lower urinary tract symptoms.

The role of PSA in PCa screening

Prostatic specific antigen (PSA) is an antigens- proteolytic protein that is secreted by prostate cells and excreted into the glandular microducts, which are largely poured into the sperm through the crystalline ducts, and smaller portions are poured into the serum and lymphatic secretions. PSA increases in PCa, prostate benign proliferation, and prostate inflammation after the procedure (cystoscopy, catheterisation of urethral, prostate massage, after a prostate biopsy within 4 weeks, after ejaculation within 48 h). PSA decreases by 50% when taking 5 alpha-reductase inhibitors with a continuous period of over 6 months [50].

In our study, PSA shown a significant associated with PCa and is an important predictor for PCa in both BMA and CART algorithm. Currently, all guidelines of the American and European Urogenital Societies use PSA cut-off levels ranging from 2 ng/mL to 4 ng/mL in order to make prostate biopsy decision [51]. Meanwhile, researchers chose $PSA > 4$ ng/mL as the cut-off level to ensure high sensitivity in screening [52–54]. According to Vietnam Ministry of Health guideline, $PSA > 4$ ng/mL has been recommended for selecting patients with lower urinary tract symptoms for a further clinical assessment for PCa diagnosis. The cut-off value of PSA for referring biopsy, however, is not determined. Previous studies showed that only using PSA for PCa screening before biopsy could tend to the high probability of a negative

biopsy out of elevated PSA cases (high proportion of unnecessary biopsy). In PCa patients, only 65–75% of cases have $PSA > 4$ ng/mL; 35% of the remaining PSA cases remain at a normal level [55]. The study by Thompson et al. [56] in U.S. on cancer screening with 2950 men over 50 years old showed that 15.2% of patients had $PSA < 4$ ng/mL got prostate cancer, as well as 14.9% of the negative prediction group with a Gleason score of ≥ 7 [56]. Wright et al. [57] found that a PSA threshold of > 4 ng/mL detected more cancer cases but increased unnecessary biopsy cases [57]. Morgan et al. [28] noted that the sensitivity of reached 98.2% at PSA cut-off level of > 4 ng/mL, and the sensitivity at $PSA > 10$ ng/mL was 91% with a specificity of 54% [28]. In some cases, the serum PSA values in the PCa and non-PCa groups overlapped, especially when PSA levels were 4–10 ng/mL. The PSA value in this range was called “diagnostic gray zone,” according to Shariat and Karakiewicz [58].

In hospital-based PCa screening, combining PSA, clinical parameters, age, and other risk factors could reduce the rate of unnecessary biopsy while maintaining the ability to reduce PCa mortality [12–14].

CART value for PCa screening

To remedy the inherent limitations of PSA in PCa screening, we used a combination of PSA with I-PSS and the main risk factors of PCa to build the CART model. Based on CART, patients with a PSA cut-off level > 33.5 ng/mL have a PCa risk of up to 91.2%. Patients with $I-PSS < 19$, and $PSA < 18$ ng/mL were at 7.1% risk. CART overcomes the limitations of using only I-PSS or PSA for screening. Other machine learning algorithms have been used in PCa screening in previous studies and have reached higher values than PSA only. In a study by Babaian et al. [59], a neural network algorithm for PCa screening showed an improved value compared to using only PSA, the specificity of the neural network was not good (lower than 65%) [59]. A study by Satoshi et al. [17] showed that artificial neural network, random forest, and support vector machine improved overall value when compared to only PSA; however, sensitivity and specificity were usually lower than 80% [17]. Our CART algorithm with three variables, PSA, I-PSS, and age, showed a relatively high predictive power ($AUC = 0.915$). In addition, CART algorithm could also support physicians to make clinical decisions in a step-to-step and intuitive manner; hence it has practical use in a daily clinical setting. At 20% diagnosis probability threshold, CART showed a high negative predictive value (91.2%), and at 80% diagnosis probability threshold, CART also had a high positive predictive value (91.2%). Therefore, we recommended 20% diagnosis probability threshold for negative prediction and 80% diagnosis probability threshold for referring prostate biopsy. Any other patients with a probability of

PCa range from >20% to <80%, further tests including the digital rectal examination (DRE), PSA re-test after a month, and transrectal ultrasonography (TRUS) should be considered to reduce unnecessary biopsy while keeping the ability to diagnose PCa early.

The study has some limitations. First, we lack other tests such as DRE, TRUS, biomarkers that can contribute to making biopsy decisions [5, 7]. Second, the CART model has not yet been tested in different populations for the validity and reliability of the algorithm. Finally, we could not estimate the overdiagnosis rate of PCa in the study. It warrants further study in the near future to overcome these limitations.

Conclusion

CART advised the following cut-off points in the marked screening sequence: $18 < \text{PSA} < 33.5$ ng/mL, $\text{I-PSS} \geq 19$, and $\text{age} \geq 71$. Patients with $\text{PSA} \geq 33.5$ ng/mL have a PCa risk was 91.2%; patients with $\text{PSA} < 18$ ng/mL and $\text{I-PSS} < 19$ have a PCa risk was 7.1%. Patient with $18 \leq \text{PSA} < 33.5$ ng/mL and $\text{I-PSS} < 19$ have a PCa risk is 70% if $\text{age} \geq 71$; and is 16% if $\text{age} < 71$. In overall, CART reached high predictive value with $\text{AUC} = 0.915$. Sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of CART at the 20% diagnosis probability threshold were 91.1%, 86.9%, 86.2%, 91.5% and 88.9%; at 80% diagnosis probability threshold were 81.6%, 91.2%, 92.3%, 79.2%, and 85.8%. I-PSS, PSA, and age had importance role in PCa screening. CART combining PSA, I-PSS, and age has practical use in hospital-based PCa screening in Vietnamese patients.

Acknowledgements

We thank the MEDIC Medical Center for providing us the data. We appreciate the participant of all patients in the study.

Author contributions

Nguyen Chi Cuong^{1*}, Nguyen Tuong Vien², Nguyen Minh Thien¹, Phan Thanh Hai¹, Tran Ngoc Dang^{2*}.

TND, NCC came up with ideas and framework for the study. NMT, PTH, NCC collected the hospital data. TND, NTV, NCC analyzed and interpreted the data. TND, NTV, NCC was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Funding

The authors declare that they have received no funding for this study.

Data Availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval

This study's ethical was approved by the Biomedical Research Ethics Committee of MEDIC Medical Center. All patients provided written informed consent before participating in the study. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 29 March 2021 / Accepted: 4 October 2022

Published online: 29 October 2022

References

- Cornford P, Bellmunt J, Bolla M, Briers E, Santis MD, Gross T, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part II: Treatment of Relapsing, Metastatic, and Castration-Resistant Prostate Cancer. *Eur Urol*. 2017;71:630–42.
- International Agency for Research on Cancer. Vietnam cancer factsheet. The *Glob Cancer Obs*. 2020.
- Sata F, Umemura T, Kishi R. The epidemiology of prostate cancer-recent trends in prostate cancer incidence and mortality. *Gan To Kagaru Rhyoho*. 2001;28(2):184 - 269.
- Noone AM, Howlader N, Krapcho M, et al SEER Cancer Statistics, Review. 1975–2015 http://seer.cancer.gov/csr/1975_2015/ based on November 2017 SEER data submission.: National Cancer Institute; 2018.
- American Cancer Society. Prostate cancer: Early Detection, Diagnosis, and Staging 2019 [Available from: <https://www.cancer.org/cancer/prostate-cancer/detection-diagnosis-staging/>].
- Parker C. Active surveillance: towards a new paradigm in the management of early prostate cancer. *Lancet Oncol*. 2004;5:101–6.
- Remmers S, Roobol MJ. Personalized strategies in population screening for prostate cancer. *Int J Cancer*. 2020;147(11):2977–87.
- Etzioni R, Tsodikov A, Mariotto A, Szabo A, Falcon S, Wegelin J, et al. Quantifying the role of PSA screening in the US prostate cancer mortality decline. *Cancer Causes Control*. 2008;19(2):175–81.
- Bell N, Connor Gorber S, Shane A, Joffres M, Singh H, Dickinson J, et al. Recommendations on screening for prostate cancer with the prostate-specific antigen test. *CMAJ*. 2014;186(16):1225–34.
- Hugosson J, Roobol MJ, Månsson M, Tammela TLJ, Zappa M, Nelen V, et al. A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer. *Eur Urol*. 2019;76(1):43–51.
- Hosseini SY, Moharramzadeh M, Ghadian AR, Hooshyar H, Lashay AR, Safarinejad MR. Population-based screening for prostate cancer by measuring total serum prostate-specific antigen in Iran. *Int J Urol*. 2007;14(5):406–11.
- Liu J, Dong B, Qu W, Wang J, Xu Y, Yu S, et al. Using clinical parameters to predict prostate cancer and reduce the unnecessary biopsy among patients with PSA in the gray zone. *Sci Rep*. 2020;10(1):5157.
- Aladwani M, Lophatananon A, Ollier W, Muir K. Prediction models for prostate cancer to be used in the primary care setting: a systematic review. *BMJ Open*. 2020;10(7):e034661.
- Agnihotri S, Mittal RD, Kapoor R, Mandhani A. Raising cut-off value of prostate specific antigen (PSA) for biopsy in symptomatic men in India to reduce unnecessary biopsy. *Indian J Med Res*. 2014;139(6):851–6.
- Hosseini M, Ebrahimi SM, SeyedAlinaghi S, Mahmoodi M. Sensitivity and specificity of international prostate symptom score (IPSS) for the screening of Iranian patients with prostate cancer. *Acta Medica Iranica*. 2011;49(7):451–5.
- Üçer O, Müezzinoğlu T, Çelen İ, Temeltaş G. The effect on the sensitivities of PSA and PSA-age volume score of IPSS and nocturia in predicting positive prostate biopsy findings. *Afr J Urol*. 2017;23(2):116–9.
- Nitta S, Tsutsumi M, Sakka S, Endo T, Hashimoto K, Hasegawa M, et al. Machine learning methods can more efficiently predict prostate cancer compared with prostate-specific antigen density and prostate-specific antigen velocity. *Prostate Int*. 2019;7(3):114–8.
- Lee J, Yang SW, Lee S, Hyon YK, Kim J, Jin L, et al. Machine Learning Approaches for the Prediction of Prostate Cancer according to Age and the Prostate-Specific Antigen Level. *Korean J Urol Oncol*. 2019;17(2):110–7.
- Takeshi U, Motohide U, Atsunari K, Akira N, Kazutoshi F, Yasushi M, et al. A novel model to predict positive prostate biopsy based on serum androgen level. *Endocrine-related Cancer*. 2018;25(1):59–67.
- Ahn JH, Lee JZ, Chung MK, Ha HK. Nomogram for prediction of prostate cancer with serum prostate specific antigen less than 10 ng/mL. *J Korean Med Sci*. 2014;29(3):338–42.
- Heidegger I, Fritz J, Klocker H, Pichler R, Bektic J, Horninger W. Age-Adjusted PSA Levels in Prostate Cancer Prediction: Updated Results of the Tyrol Prostate Cancer Early Detection Program. *PLoS ONE*. 2015;10(7):e0134134.
- Tillier C, Cornford P, De Lauw G, Kaur K, Kinsella J, Klinge IC, et al. Evidence-based Guidelines for Best Practice in Urological Health Care: Transrectal

- Ultrasound Guided Biopsy of the Prostate. The European Association of Urology Nurses; 2019.
23. Vu Le C, Dao OQ, Khac Tran LN. Mass screening of prostate cancer in Vietnam: current status and our opinions. *Urol Oncol*. 2010;28(6):673–6.
 24. World Health Organization. The Asia-Pacific perspective: redefining obesity and its treatment: Sydney : Health Communications Australia; 2000 2000.
 25. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The Physical Activity Guidelines for Americans. *JAMA*. 2018;320(19):2020–8.
 26. Abuse Substance. Key substance use and mental health indicators in the United States: Results from the 2019 National Survey on Drug Use and Health. 2020.
 27. Barry MJ, Fowler FJ Jr, O'Leary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK, et al. The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. *J Urol*. 1992;148(5):1549–57. discussion 64.
 28. Morgan TO, Steven O, Jacobsen J, McCarthy WF, Jacobson DJ, McLeod DG, et al. Age specific ranges for serum prostate specific antigen in Black men. *New Engl J Med*. 1996;335:304–10.
 29. Ragin C, Davis-Reyes B, Tadesse H, Daniels D, Bunker CH, Jackson M, et al. Farming, reported pesticide use, and prostate cancer. *Am J Mens Health*. 2013;7(2):102–9.
 30. Parent ME, Déry M, Siemiatycki J. Does exposure to agricultural chemicals increase the risk of prostate cancer among farmers? *Mcgill J Med*. 2009;12(1):70–7.
 31. Pardo LA, Beane Freeman LE, Lerro CC, Andreotti G, Hofmann JN, Parks CG, et al. Pesticide exposure and risk of aggressive prostate cancer among private pesticide applicators. *Environ Health*. 2020;19(1):30.
 32. Band PR, Abanto Z, Bert J, Lang B, Fang R, Gallagher RP, et al. Prostate cancer risk and exposure to pesticides in British Columbia farmers. *Prostate*. 2011;71(2):168–83.
 33. Koutros S, Beane Freeman LE, Lubin JH, Heltshe SL, Andreotti G, Barry KH, et al. Risk of total and aggressive prostate cancer and pesticide use in the Agricultural Health Study. *Am J Epidemiol*. 2013;177(1):59–74.
 34. Koutros S, Berndt SI, Hughes Barry K, Andreotti G, Hoppin JA, Sandler DP, et al. Genetic susceptibility loci, pesticide exposure and prostate cancer risk. *PLoS ONE*. 2013;8(4):e58195.
 35. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet*. 2018;50(7):928–36.
 36. Shephard RJ. Physical Activity and Prostate Cancer: An Updated Review. *Sports Med*. 2017;47(6):1055–73.
 37. Benke IN, Leitzmann MF, Behrens G, Schmid D. Physical activity in relation to risk of prostate cancer: a systematic review and meta-analysis. *Ann Oncol*. 2018;29(5):1154–79.
 38. Perez-Cornago A, Travis RC, Appleby PN, Tsilidis KK, Tjønneland A, Olsen A, et al. Fruit and vegetable intake and prostate cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Int J Cancer*. 2017;141(2):287–97.
 39. WCRF/AICR. World Cancer Research Fund International/American Institute for Cancer Research Continuous Update Project Report: Diet, Nutrition, Physical Activity, and Prostate Cancer. 2018 [Available from: <https://www.wcrf.org/wp-content/uploads/2021/02/prostate-cancer-report.pdf>].
 40. Liu RH. Potential synergy of phytochemicals in cancer prevention: mechanism of action. *J Nutr*. 2004;134(12 Suppl):3479s–85s.
 41. Key TJ. Fruit and vegetables and cancer risk. *Br J Cancer*. 2011;104(1):6–11.
 42. Halpern EJ. *Epidermiology and histopathology of prostate cancer*. In: Halpern EJCM, editor. *Imaging of the prostate 1st ed*. Martin Dunitz, London-New York; 2002. pp. 19–25.
 43. Reiter RE, de Kernion JB. *Epidermiology. Etiology and Prevention of Prostate Cancer*. In: Walsh PC, Retik AB, Vaughan ED, editors. *Campbell's Urology*. 8th ed. USA: Saunder Company; 2002. pp. 3003–14.
 44. De Sousa S, Marcia P, dA, Evandro L, et al. Correlation Between PSA Level, Gleason Score and Bone Scan in Staging of Prostate Cancer Patients. *ALAS-BIMN J*. 2005;7(27):1–6.
 45. European Urology EAU. *Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent—Update 2013*. Published by Elsevier B.V. 2014. pp. 124 – 37 p.
 46. Giwercman A, Richiardi L, Kaijser M, Ekblom A, Akre O. Reduced risk of prostate cancer in men who are childless as compared to those who have fathered a child: a population based case-control study. *Int J Cancer*. 2005;115(6):994–7.
 47. de la Rosette J, et al. *Guidelines on the benign prostatic hyperplasia*. European Association of Urology; 2007.
 48. Oelke M, et al. *Guidelines on Management of male lower urinary tract symptoms (LUTS), incl. Benign prostatic obstruction (BPO)*. European Association of Urology; 2012.
 49. Martin RM, Vatten L, Gunnell D, Romundstad P, Nilsen TI. Lower urinary tract symptoms and risk of prostate cancer: the HUNT 2 Cohort, Norway. *Int J Cancer*. 2008;123(8):1924–8.
 50. Duc NH, et al. The role of prostate specific antigen (PSA) in the early diagnosis of prostate cancer. *Med Ho Chi Minh City*. 2009;13(1):5–9.
 51. Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med*. 2009;360(13):1320–8.
 52. Carroll P, Coley C, McLeod D, et al. Prostate-specific antigen best practice policy - part I: early detection and diagnosis of prostatic cancer. *Urology* 2001;57(2):217 - 24.
 53. Carter HB. J.D P. Prostate-specific antigen testing for early diagnosis of prostate cancer: formulation of guidelines. *Urology*. 1999;54(5):780–6.
 54. Catalona WJ, Partin AW, Slawin KM, Brawer MK, Flanigan RC, Patel A, et al. Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: a prospective multicenter clinical trial. *JAMA*. 1998;279(19):1542–7.
 55. Caplan A. A. K. Prostate-specific antigen and the early diagnosis of prostate cancer. *Am J Clin Pathol*. 2002;117:104–8.
 56. Thompson IM, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level < or = 4.0 ng per milliliter. *N Engl J Med*. 2004;351(14):1470.
 57. Wright EJ, Fang J, Metter EJ. Prostate specific antigen predicts the long-term risk of prostate enlargement: results from the Baltimore Longitudinal Study of Aging. *J Urol*. 2002;167(6):2484–8.
 58. Shariat SF, Karakiewicz PI. Screening for prostate cancer in 2007: the PSA era and its challenges are not over. *Eur Urol*. 2008;53(3):457–60.
 59. Babaian RJ, Fritsche H, Ayala A, Bhadkamkar V, Johnston DA, Naccarato W, et al. Performance of a neural network in detecting prostate cancer in the prostate-specific antigen reflex range of 2.5 to 4.0 ng/mL. *Urology*. 2000;56(6):1000–6.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.