Validation of the prostate health index in a predictive model prostate cancer

A. Sanchís-Bonet[a,*], M. Barrionuevo-González[b], A.M. Bajo-Chueca[c], L. Pulido-Fonseca[a], L.E. Ortega-Polledo[a], J.C. Tamayo-Ruiz[a] y M. Sánchez-Chapado[a,c]

[a] *Departamento de Urología, Hospital Universitario Príncipe de Asturias, Alcalá de Henares (Madrid), España*

[b] *Departamento de Bioquímica y Análisis Clínicos, Hospital Universitario Príncipe de Asturias, Alcalá de Henares (Madrid), España*

[c] *Departamento de Biología de Sistemas, Facultad de Medicina, Universidad de Alcalá de Henares, Alcalá de Henares (Madrid), España*

∗ Corresponding author. E-mail address: asanchisbonet@yahoo.es (A. Sanchís-Bonet)

Abstract

Objectives: To validate and analyse the clinical usefulness of a predictive model of prostate cancer that incorporates the biomarker "[-2] pro prostate-specific antigen" using the prostate health index (PHI) in decision making for performing prostate biopsies.

Material and methods: We isolated serum from 197 men with an indication for prostate biopsy to determine the total prostate-specific antigen (tPSA), the free PSA fraction (fPSA) and the [-2] proPSA (p2PSA). The PHI was calculated as p2PSA/fPSA×√tPSA. We created 2 predictive models that incorporated clinical variables along with tPSA or PHI. The performance of PHI was assessed with a discriminant analysis using receiver operating characteristic curves, internal calibration and decision curves.

Results: The areas under the curve for the tPSA and PHI models were 0.71 and 0.85, respectively. The PHI model showed a better ability to discriminate and better calibration for predicting prostate cancer but not for predicting a Gleason score in the biopsy ≥ 7. The decision curves showed a greater net benefit with the PHI model for diagnosing prostate cancer when the probability threshold was 15-35% and greater savings (20%) in the number of biopsies.

Conclusions: The incorporation of p2PSA through PHI in predictive models of prostate cancer improves the accuracy of the risk stratification and helps in the decision-making process for performing prostate biopsies.

KEYWORDS: Prostate cancer; Prostate health index; Predictive models; Decision curve analysis; Prostate biopsy

Resumen

Objetivos: Validar y analizar la utilidad clínica de un modelo predictivo de cáncer de próstata que incorpora el biomarcador "[-2] proantígeno prostático específico" a través del índice de salud prostática (PHI) en la toma de decisión para realizar una biopsia de próstata.

Material y métodos: Se aisló suero de 197 varones con indicación de biopsia de próstata para la determinación del antígeno prostático específico total (tPSA), fracción libre de PSA (fPSA) y [-2] proPSA (p2PSA); el PHI se calculó como p2PSA/fPSA×√tPSA. Se crearon 2 modelos predictivos que incorporaban variables clínicas junto a tPSA o a PHI. Se evaluó el rendimiento de PHI usando análisis de discriminación mediante curvas ROC, calibración interna y curvas de decisión.

Resultados: Las áreas bajo la curva para el modelo tPSA y el modelo PHI fueron de 0,71 y 0,85, respectivamente. PHI mostró mejor capacidad de discriminación y mejor calibración para predecir cáncer de próstata, pero no para predecir un grado de Gleason en la biopsia ≥ 7. Las curvas de decisión mostraron un beneficio neto superior del modelo PHI para el diagnóstico de cáncer de próstata cuando el umbral de probabilidad está entre 15 y 35% y un mayor ahorro (20%) en el número de biopsias.

Conclusiones: La incorporación de p2PSA a través de PHI a los modelos predictivos de cáncer de próstata mejora la exactitud en la estratificación del riesgo y ayuda en la toma de decisión sobre realizar una biopsia de próstata.

PALABRAS CLAVE; Cáncer de próstata; Índice de salud prostática; Modelos predictivos; Curvas de decisión; Biopsia de próstata.

Introduction

Serum prostate-specific antigen (PSA) was the first marker used for the early diagnosis of prostate cancer (PCa); however, it is not very specific, with the consequences that this entails.[1] The free PSA fraction (fPSA) and the percentage of fPSA (% fPSA) significantly improve the discrimination between PCa and other benign situations, especially in patients with tPSA levels between 4 and 10 ng/ml.[2] The p2PSA molecule, an isoform of fPSA, and the prostate health index (PHI), which incorporates tPSA, fPSA and p2PSA in its calculation, improve the diagnostic accuracy of PCa and of clinically significant or aggressive disease.[3,4]

However, the calculation of the diagnostic accuracy in terms of reliability indexes, or the calculation of the area under the curve (AUC) are not enough to answer the following questions: what is the probability that a patient has of suffering from PCa? Or how can I stratify a patient's risk of having PCa? And, therefore, to whom should I biopsy and who benefits from the prostate biopsy?

The development of predictive models that make it possible to stratify and discriminate the risk, its calibration, and the development of decision curves could be more useful in clinical practice.[5,6] The purpose of this study is to analyze and validate the clinical usefulness of a predictive model of PCa that incorporates the biomarker "[-2] prostate-specific proantigen" through the PHI in the decision-making process to perform a prostate biopsy.

Material and methods

Prospectively and through an observational cohort design, between January 2015 and December 2016, 197 men were included with indication of first or successive prostate biopsy, who signed an informed consent for the inclusion in the study and for the extraction of a blood sample. The study was approved by the Hospital Ethics Committee. A tPSA 2-20 ng/ml or the

presence of a suspicious digital rectal examination in men over 45 years of age was fixed as inclusion criterion. Patients with a history of taking 5-α-reductase inhibitors in the last 6 months and patients with a history of urinary tract infection or lower urinary tract manipulation in the 3 months prior to the biopsy indication were excluded from the study.

The clinical variables of interest (first or successive biopsy, age of the patients, and digital rectal examination), the laboratory variables (tPSA, fPSA, and p2PSA) and the findings in the analyzed biopsy samples (presence of cancer or not in biopsy and presence of at least one Gleason grade ≥ 7 in the biopsy) were collected. Two models were defined: PSA model (age, digital rectal examination, previous biopsy, and tPSA) and PHI model (age, digital rectal examination, previous biopsy, and PHI).

Blood samples were taken immediately before the biopsy, they were ultracentrifuged and frozen at −80° in the first 2 h after extraction to minimize the low stability of p2PSA at room temperature in serum.[7] Blindness was maintained for the study investigators regarding the recognition of blood samples.

All patients underwent prostate biopsy through transrectal ultrasound and at least 12 cylinders were obtained from the peripheral zone (apical, medial and cranial) of the prostate gland in all cases. When necessary, additional cylinders were taken. The histological analysis of the samples was performed and their Gleason grade was obtained according to the criteria agreed at the Consensus Conference of the International Association of Uro-pathologists in 2005 (ISUP 2005).[8] Aggressive PCa was defined as the presence in the biopsy of at least one Gleason grade ≥ 7. Blindness was maintained for the study investigators as to the result of the biopsy.

The analysis of tPSA, fPSA and p2PSA was carried out using the Access Hybritech immunoassay by Beckman Coulter and the PHI was calculated using the formula

p2PSA/fPSA×. The percentage of fPSA and the percentage of p2PSA (% p2PSA) were also reported.

An estimation of the necessary sample size was made in order to, with a PCa detection rate in our population of 43%, find differences in the AUC with a statistical power of 80% and an alpha error of 5%. With these premises, at least 195 patients would be needed.

The variables of interest of the study population were compared using the Student's *t* test for independent data and the Mann-Whitney *U* test in the case of quantitative variables and Pearson's chi-square test in the case of qualitative variables. The receiver operating characteristics (ROC) curves were used to determine the diagnostic accuracy of each of the variables and models (PSA and PHI), to determine the cut-off point in which the highest sensitivity and specificity is reached and to evaluate the discriminative capacity of the variables and models, that is, their ability to differentiate subjects without cancer versus subjects with cancer, and subjects with aggressive cancer versus sub- jects with non-aggressive cancer.[9] The comparison of the ROC curves was done non-parametrically by means of the DeLong test.[10]

The calibration of tPSA, fPSA and p2PSA and of the different models was performed by comparing the predicted probabilities and the probabilities observed by the Hosmer-Lemeshow test and it was represented graphically. The relationship between the probabilities predicted by both models was also graphically represented both for the diagnosis of cancer and for the diagnosis of aggressive cancer using scatter plots.

The decision curves were used to examine the net benefit (cancer detection) obtained with each model (PSA and PHI) and the number of interventions that can be avoided (number of unnecessary biopsies).

Results

85 cancers (43%) and 44 aggressive cancers (52%) were diagnosed. Twenty-five (12.6%)

males had a tPSA >P10 and the median PSA (ng/ml) among these males was 14 (10.24-19.56). Table 1 shows the characteristics of the study cohort. The analysis by means of ROC curves of the available variables and their combinations (Table 2) showed that PHI had better discrimination capacity for the diagnosis of PCa (AUC: 0.82), but not for the diagnosis of aggressive cancer (AUC: 0.66). The best combination to discriminate aggressive cancer was tPSA+% fPSA+% p2PSA (AUC: 0.71).

Fig. 1 shows how PHI is the tool that presents the best calibration as a predictor of PCa diagnosis. Fig. 2 shows the superior net benefit of PHI in the diagnosis of PCa compared to the rest of the variables along the different cut-off points or probability thresholds. However, there is no superior net benefit of PHI in the diagnosis of aggressive cancer.

The results of the clinical usefulness of the 2 proposed models (PSA and PHI) in the diagnosis of PCa and aggressive disease are shown by ROC curves (Fig. 3), by means of scatter plots (Fig. 4), by means of calibration (Fig. 5), and by decision curves: the net benefit (Fig. 6A and B) and the number of interventions (biopsies) that we can save by incorporating PHI to the predictive model of PCa is observed (Fig. 6C).

Discussion

In this study, as it happens in others, both belonging to an institution alone and in multicenter studies, PHI is much more accurate and has greater discriminatory power than the rest of the markers in the prediction of PCa in patients with indication of first biopsy and in successive biopsies, although our cohort only includes 13% of patients with indication of 2nd or 3rd biopsy.[11---13]

However, is this last by itself enough to be able to assert that p2PSA and the calculated index - PHI - is a better marker to predict PCa than other markers? Or, in other words, what criteria should a marker gather to justify its inclusion in the clinical decision-making processes?[14] On

the one hand, we see that the AUC of PHI as an isolated variable when compared to other variables alone or combined and the mixed clinical-analytical PHI model when compared to the mixed clinical-analytical model, PSA is better than the AUC of tPSA (0.82-0.85 versus 0.56-0.70). Therefore, the power of discrimination of PHI is better. Even if we incorporated a multivariate logistic regression model, previously shown in our patients,[15] although not shown in the current results of this study, the odds ratio of PHI in a multivariate model is only slightly higher than 1, 1.03 (1.02-1.04, $p < 0.001$). In this model, although the statistical significance is really small and is adjusted by other markers, the model would be subject to great variability due to the way in which the variable is analyzed (in this case in a continuous quantitative way), and it would probably differ if quartiles were used in its analysis. Even the model would be influenced by the type of variables that we include (clinical variables, other biomarkers . . .). That is to say, the discrimination provided by the ROC curves and the implementation using a multivariate logistic regression model would not be sufficient to assert that PHI is really useful in clinical practice.

On the other hand, it is clearly established that the evaluation of a new biomarker cannot be performed independently of a model, because the biomarker will be more valuable as soon as the predictive accuracy increases for a given event - in our case the diagnosis of PCa -, compared to a model that does not include the new biomarker-in our case PHI.[16] In this sense, PHI has also been evaluated in combination with other markers in different studies and different predictive models of PCa have been created that have also resulted with greater predictive power,[17] although it is more interesting to use predictive models that incorporate clinical variables with biomarkers, as we have done in this study, because they offer a more accurate approach to the reality of our patients.

An internal validation of the behavior of the different markers in the form of calibration has

been carried out, since the calibration informs us of how close the risks predicted by our isolated variables or included in our models are to the rates observed in the reality of our patients. A perfect calibration should show a 1:1 ratio between predictions and observations, which would result in an inclination slope of about 45°.[18] In this sense, PHI offers the smallest jump between the observed and expected values for the diagnosis of PCa along a linear model with a 45° inclination slope. Neither tPSA nor the rest of the variables in isolation offer it in our patients, in whom the fact that the prediction capacity of tPSA, fPSA, and % fPSA is lower than 50% stands out. In addition, the model that incorporates PHI shows better calibration at the different points of the 45° slope.

The dispersion graphs also help to understand that if the predictions were very similar between the 2 models, the points would be very close to 45°, without much dispersion; therefore, the model that includes PHI would not be of clinical usefulness.[19] The dispersion seen in the predictive model of PCa indicates that the incorporation of PHI into the model has clinical usefulness in our patients.

To understand if our predictive model will modify our clinical practice, that is, it will have an impact on the clinical decisions that can be made by the doctor and the patient, the decision curves are fundamental, since they show us the net benefit of our marker or of our models in different cut-off points of probability of suffering from PCa.[20] The net benefit of PHI as an isolated marker is clearly higher than that of the rest of the variables when the probability of suffering from PCa is between 20 and 40%. In the model that includes PHI compared to the model that includes tPSA, the net benefit is appreciated from very early with a clear separation of the curve, so when the probability is between 15 and 35%, the net benefit is clearly higher. In recent years, different decision curves have been published that try to improve clinical decision-making in patients. In the study by Lazzeri et al.,[17] a greater net benefit is reported of

the model that includes PHI between 30 and 60%; therefore, in this interval, the use of the model would be supported to improve the diagnostic performance; however, with a probability of 60%, the net benefit is only 2.5%.

PHI has also been evaluated from the perspective of whether it is capable of predicting a more aggressive disease in patients diagnosed with PCa. Several studies have shown this, although in some only the discrimination capacity of PHI has been evaluated.[11,17,21,22] In this sense, in our study, PHI does not provide greater usefulness than tPSA in the aggressiveness of the diagnosed PCa. Probably, the known panel of the 4 kallikreins together with the clinical variables age, digital rectal examination, and previous biopsy history through the 4 K score test better fulfills the role of high-grade cancer prediction when compared to other available multivariate prognostic models as the Prostate Cancer Prevention Trial-Risk Calculator 2.0 and the European Research Screening Prostate Cancer-Risk Calculator 4.[23]

The number of biopsies that we would save by incorporating PHI to the model versus the model that incorporates tPSA is clearly superior: it reaches almost 20% in some of the cut-off points.

We can conclude that PHI has greater discrimination power, a higher calibration, and a higher net benefit, both as an individual variable and when introducing it in predictive models of PCa diagnosis and, therefore, it has a greater clinical utility in our decision-making before patients, which also enables us a greater percentage of savings in the number of biopsies.

However, we must recognize several limitations. The first one concerns the small sample size, which can explain the amplitude of some confidence intervals. Secondly, not all candidates for biopsy gave their consent to participate in the study, so we cannot exclude a clear selection bias. The possible bias in the tPSA variable as indicator of biopsy and performed in the routine laboratory and the tPSA analyzed by the Beckman technique was

resolved by contrasting that there were no statistically significant differences between both tPSAs in the sample and, therefore, the tPSA obtained by Beckman's technique was valid for the study. There could also be a potential bias of the false negative in the patients who represented the group of successive biopsies; even so, the bias would affect less than 13% of the sample. Also, the fact of performing transrectal biospy has a potential false negative rate. The limited use- fulness that PHI shows for predicting aggressive PCa could be due to several circumstances, one of them the sample size, and another one that has not been used the Gleason grade of the radical prostatectomy specimen. Finally, this internal validation of the model does not allow us to generalize the results to other patients, since, although it would have been desirable, we did not validate our sample externally.

Funding

This work has been developed thanks to a research grant from the Foundation for Biomedical Research of the Príncipe de Asturias University Hospital.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Schröder FH, Hugosson J, Roobol MJ, Tammela TLJ, Ciatto S, Nelen V, et al. Screening and prostate-cancer mortality in a randomized European study. N Engl J Med. 2009;360:1320-8.
2. Catalona WJ, Smith DS, Wolfert RL, Wang TJ, Rittenhouse HG, Ratliff TL, et al. Evaluation of percentage of free serum prostate-specific antigen to improve specificity of prostate cancer screening. JAMA. 1995;274:1214-20.
3. Filella X, Gimenez N. Evaluation of [-2] proPSA and Prostate Health Index (phi) for the detection of prostate cancer: a systematic review and meta-analysis. Clin Chem Lab

Med. 2013;51:729-39.

4. Wang W, Wang M, Wang L, Adams TS, Tian Y, Xu J. Diagnostic ability of %p2PAS and prostate health index for aggressive prostate cancer: a meta-analysis. Sci Rep. 2014;4:5012.
5. Steyerberg EW, Vickers AJ. Decision curve analysis: a discussion. Med Decis Mak. 2008;28:146-9.
6. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. Med Decis Mak. 2015;35:162-9.
7. Semjonow A, Köpke T, Eltze E, Pepping-Schefers B, Burgel H, Darte C. Pre-analytical in-vitro stability of [-2]proPSA in blood and serum. Clin Biochem. 2010;43:926-8.
8. Epstein JI, Allsbrook WC, Amin MB, Egevad LL, ISUP Grading Committee. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason grading of prostate carcinoma. Am J Surg Pathol. 2005;29:1228-42.
9. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29-36.
10. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristics curves: a nonparametric approach. Biometrics. 1988;44:837-45.
11. Lughezzani G, Lazzeri M, Haese A, McNicholas T, de la Taille A, Buffi NM, et al. Multicenter European external validation of a prostate health index-based nomogram for predicting prostate cancer at extended biopsy. Eur Urol. 2014;66:906-12.
12. Catalona WJ, Partin AW, Sanda MG, Wei JT, Klee GG, Bangma CH, et al. A multicenter study of [-2]pro-prostate specific antigen combined with prostate specific antigen and free prostate specific antigen for prostate cancer detection in the 2.0 to 10.0

ng/ml prostate specific antigen range. J Urol. 2011;185:1650-5.

13. Lazzeri M, Briganti A, Scattoni V, Lughezzani G, Larcher A, Gadda GM, et al. Serum index test %[-2]proPSA and Prostate Health Index are more accurate than prostate specific antigen and %fPSA in predicting a positive repeat prostate biopsy. J Urol. 2012;188:1137-43.

14. Nguyen CT, Kattan MW. How to tell if a new marker improves prediction. Eur Urol. 2011;60:226-8.

15. Sanchis-Bonet A, Morales-Palacios M, Barrionuevo-Gonzalez M, Tamayo-Ruiz JC, Sanchez-Chapado M. Clinical performance of [-2] pro-prostate-specific antigen and prostate health index for prediction of prostate cancer in a cohort of Spanish men. Eur Urol Suppl. 2016;15:e1562.

16. Katttan MW. Judging new markers by their ability to improve predictive accuracy. J Natl Cancer Inst. 2003;95:634-5.

17. Lazzeri M, Haese A, de la Taille A, Palou J, McNicholas T, Lughez- zani G, et al. Serum isoform [-2] proPSA derivates significantly improve prediction of prostate cancer at initial biopsy in a total PSA range of 2-10 ng/ml: a multicentric European study. Eur Urol. 2013;63:986-94.

18. Vickers AJ, Cronin AM. Traditional statistical methods for evaluating prediction models are uninformative as to clinical value: towards a decision analytic framework. Semin Oncol. 2010;37:31-8.

19. Steyererg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21:128-38.

20. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak. 2006;26: 565-74.

21. Morote J, Celma A, Planas J, Placer J, Ferrer R, de Torres I, et al. Diagnostic accuracy of prostate health index to identify aggressive prostate cancer. An institutional validation study. Actas Urol Esp. 2016;40:378-85.

22. Foley RW, Gorman L, Sharifi N, Murphy K, Moore H, Tuzova AV, et al. Improving multivariable prostate cancer risk assessment using the prostate health index. BJU Int. 2016;117:409-17.

23. Borque-Fernando A, Esteban-Escaño LM, Rubio-Briones J, Lou- Mercadé AC, Garcia Ruiz R, Tejero-Sánchez A, et al. 4 K score test, Prostate Cancer Prevention Trial-risk calculator y European Research Screening Prostate-risk calculator en la predicción del cáncer de próstata de alto grado; estudio preliminar. Actas Urol Esp. 2016;40:155-63.

Table 1. Descriptive characteristics of patients.

	All	Without PCa	With PCa	*p*
Patients, number (%)	197 (100)	112 (57)	85 (43)	
Age, years[a]	68 (62—71)	67 (61—70)	70 (65—74)	<0.001[b]
DRE, number (%)				
Not suspicious	112 (57)	88 (66)	46 (34)	<0.001[c]
Suspicious	85 (43)	24 (38)	39 (61)	
Previous biopsy, number (%)				
No	171 (87)	96 (86)	75 (88)	0.6[c]
Yes	26 (13)	16 (14)	10 (12)	
Prostate volume in cc[a]	39 (27—54)	44 (28—58)	37 (26—46)	0.005[d]
Gleason grade, number (%)				
≤6	N/A	N/A	41 (48)	
7	N/A	N/A	25 (29)	
≥8	N/A	N/A	19 (23)	
tPSA, ng/ml[a]	5.8 (4.4---7.8)	5.7 (4.1---7.7)	6.2 (4.6—7.9)	0.10[d]
fPSA, ng/ml[a]	0.67 (0.45---1.03)	0.67 (0.40---1.10)	0.69 (0.47—0.96)	0.9[d]
Percentage fPSA (%)	0.11 (0.08---0.16)	0.13 (0.09---0.18)	0.10 (0.07—0.13)	<0.001[d]
p2PSA, pg/ml[a]	16.9 (11.4---26.8)	13.8 (9.3---23.6)	20.3 (13.6—31.6)	<0.001[d]
Percentage p2PSA (%)	0.25 (0.17---0.39)	0.20 (0.14---0.32)	0.32 (0.24—0.43)	<0.001[d]
PHI[a]	57.5 (36.5---90.0)	41.3 (32.5---57.5)	85.8 (64.7—114.0)	<0.001[d]

[a] Quantitative data expressed as median (range).

[b] Student's *t*.

[c] Pearson's chi square.

[d] Mann---Whitney *U*.

Table 2. Discrimination capacity of the markers and the combination of markers.

	AUC of the markers (95% CI)			AUC of the combination of markers (95% CI)		
	PCa	Aggressive PCa		PCa	Aggressive PCa	
tPSA	0.56 (0.48–0.64)	0.61 (0.49–0.73)	tPSA+fPSA	0.66 (0.58–0.73)	0.59 (0.46–0.71)	
fPSA	0.50 (0.42–0.58)	0.46 (0.34–0.59)	tPSA+%fPSA	0.65 (0.58–0.73)	0.59 (0.47–0.71)	
%fPSA	0.65 (0.57–0.73)	0.56 (0.44–0.69)	fPSA+p2PSA	0.76 (0.69–0.83)	0.67 (0.55–0.78)	
p2PSA	0.67 (0.60–0.75)	0.67 (0.56–0.79)	%fPSA+%p2PSA	0.72 (0.65–0.79)	0.65 (0.53–0.77)	
%p2PSA	0.71 (0.63–0.78)	0.65 (0.53–0.77)	tPSA+%fPSA+%p2PSA	0.73 (0.66–0.80)	0.71 (0.60–0.82)	
PHI	0.82 (0.75–0.88)	0.66 (0.54–0.77)	PHI	0.82 (0.75–0.88)	0.66 (0.54–0.77)	

AUC: area under the curve; 95% CI: 95% confidence interval.

Figure legends

Figure 1. Calibration of the different markers for the diagnosis of PCa: (A) PSA, (B) fPSA, C)% fPSA, (D) p2PSA, (E)% p2PSA and (F) PHI.

Figure 2. Decision curves showing the net benefit of each of the markers: (A) for the diagnosis of PCa; (B) for the diagnosis of aggressive PCa.

Figure 3. ROC curves comparing PSA model and PHI model: (A) for the diagnosis of PCa and (B) of aggressive PCa.

Figure 4. Relationship between the probability prediction (degree of dispersion) of the PSA model and the PHI model: (A) for the detection of PCa and (B) for the detection of aggressive PCas.

Figure 5. Calibration for the diagnosis of PCa of the model: (A) PSA, (B) of the PHI, and for the diagnosis of aggressive PCa of the models: (C) PSA and (D) PHI.

Figure 6. Decision curves showing the net benefit of the PSA and PHI models (A) for the detection of PCa and (B) for the detection of aggressive PCa, and (C) the number of interventions avoided.

Figure 1

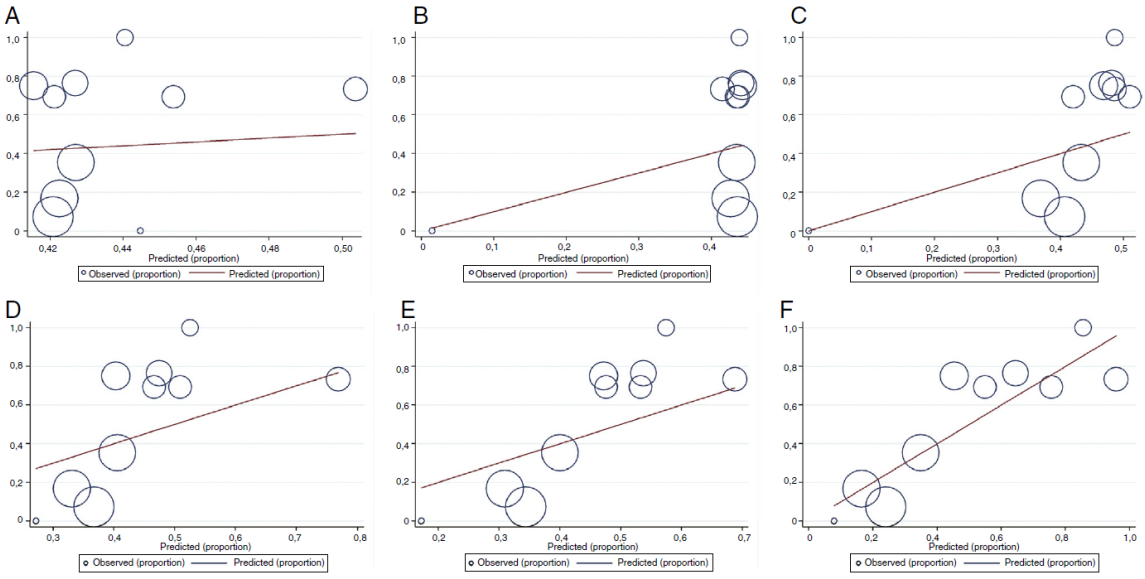

A
B
C
D
E
F
Predicted (proportion)
Observed (proportion)
Predicted (proportion)

Figure 2

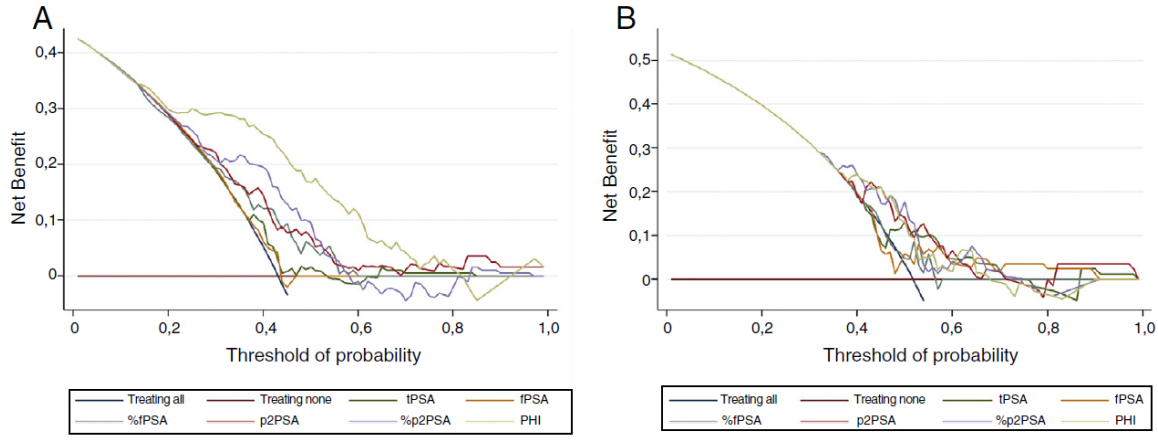

A
Net Benefit
0,4
0,3
0,2
0,1
0
0
0,2
0,4
0,6
0,8
1,0
Threshold of probability
Treating all
Treating none
tPSA
fPSA
%fPSA
p2PSA
%p2PSA
PHI
B
Net Benefit
0,5
0,4
0,3
0,2
0,1
0
0
0,2
0,4
0,6
0,8
1,0
Threshold of probability
Treating all
Treating none
tPSA
fPSA
%fPSA
p2PSA
%p2PSA
PHI

Figure 3

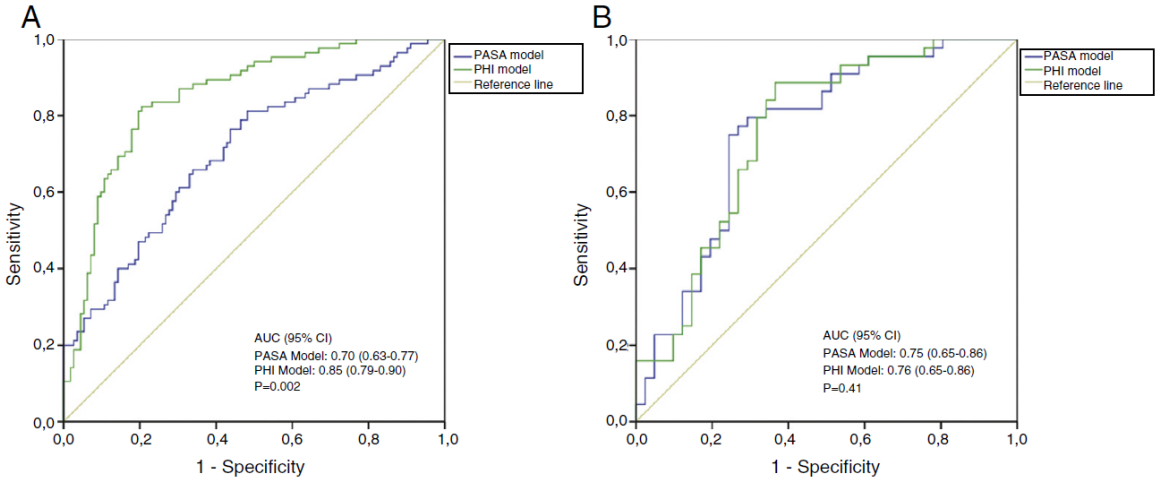

A
Sensitivity
1 - Specificity
PASA model
PHI model
Reference line
AUC (95% CI)
PASA Model: 0.70 (0.63-0.77)
PHI Model: 0.85 (0.79-0.90)
P=0.002
B
Sensitivity
1 - Specificity
PASA model
PHI model
Reference line
AUC (95% CI)
PASA Model: 0.75 (0.65-0.86)
PHI Model: 0.76 (0.65-0.86)
P=0.41

Figure 4

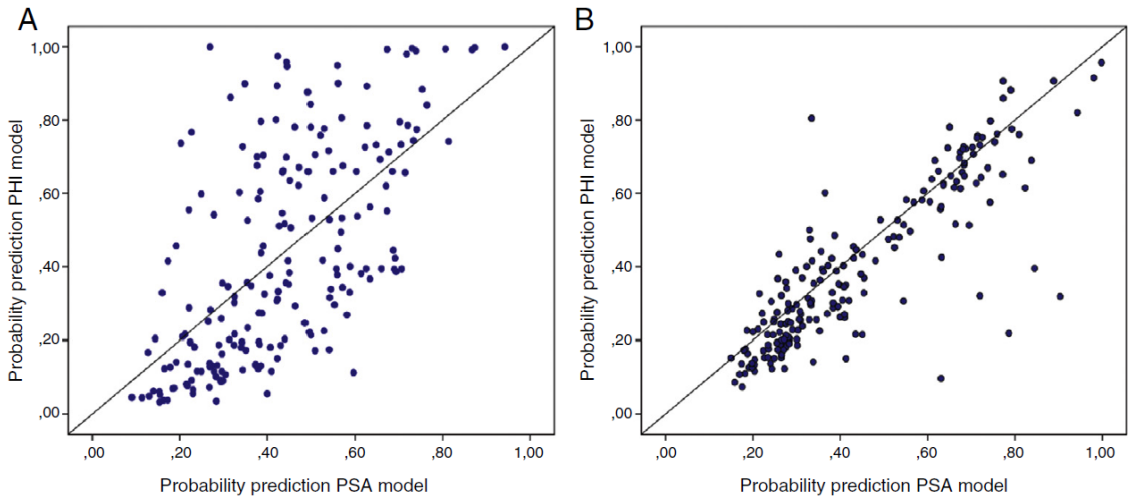

A
B
Probability prediction PHI model
Probability prediction PSA model
1,00
,80
,60
,40
,20
,00

Figure 5

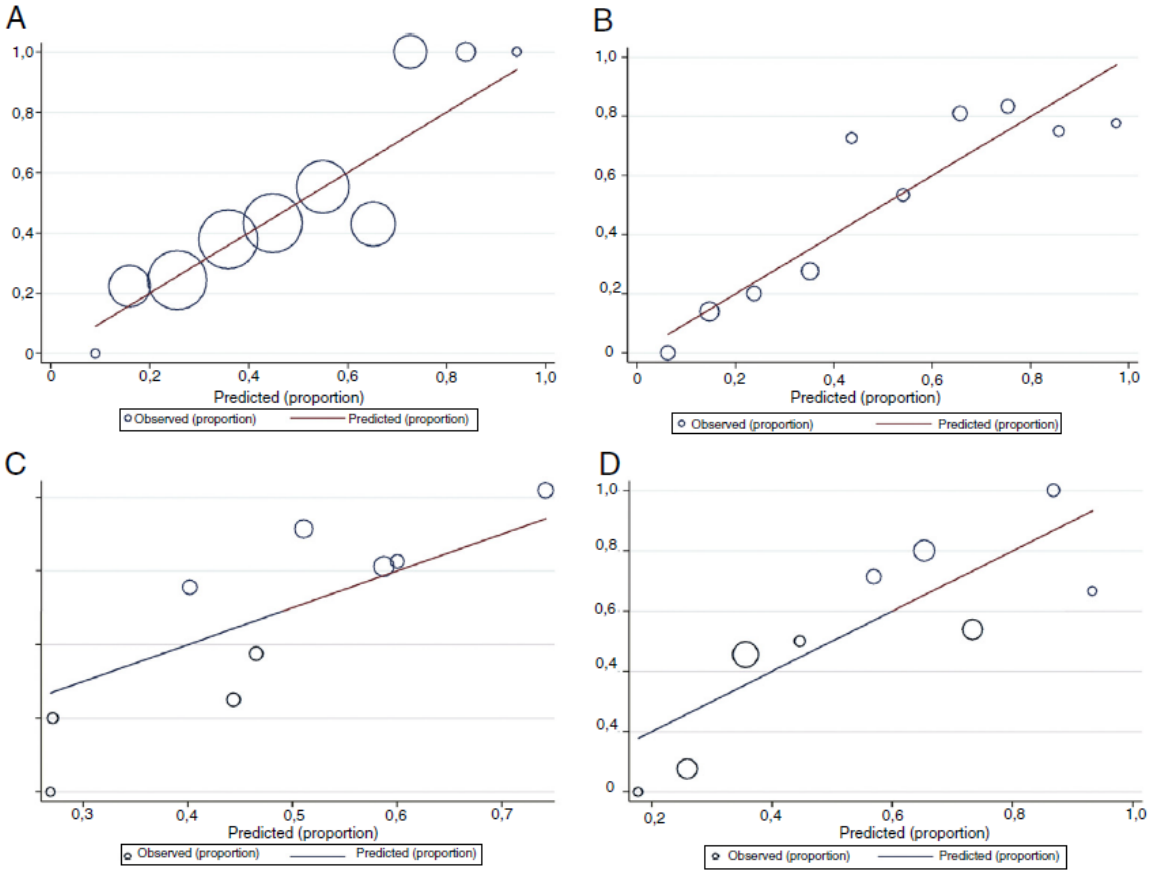
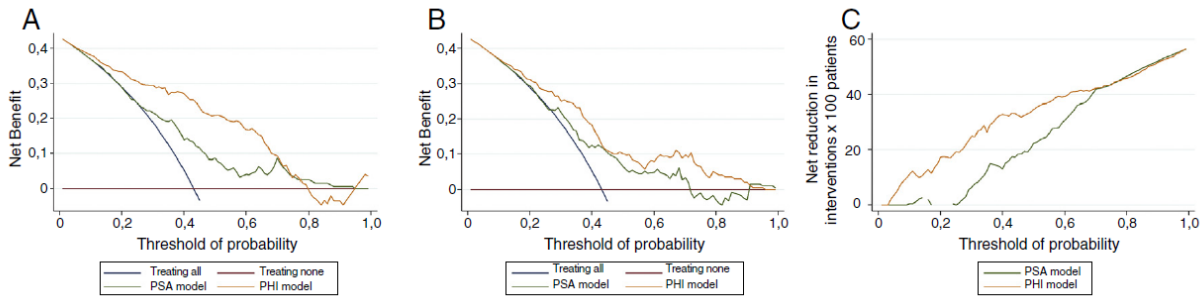

A
1,0
0,8
0,6
0,4
0,2
0
0
0,2
0,4
0,6
0,8
1,0
Predicted (proportion)
Observed (proportion)
Predicted (proportion)
B
1,0
0,8
0,6
0,4
0,2
0
0
0,2
0,4
0,6
0,8
1,0
Predicted (proportion)
Observed (proportion)
Predicted (proportion)
C
0,3
0,4
0,5
0,6
0,7
Predicted (proportion)
Observed (proportion)
Predicted (proportion)
D
1,0
0,8
0,6
0,4
0,4
0
0,2
0,4
0,6
0,8
1,0
Predicted (proportion)
Observed (proportion)
Predicted (proportion)

Figure 6


A
Net Benefit
0,4
0,3
0,2
0,1
0
0
0,2
0,4
0,6
0,8
1,0
Threshold of probability
Treating all
Treating none
PSA model
PHI model
B
Net Benefit
0,4
0,3
0,2
0,1
0
0
0,2
0,4
0,6
0,8
1,0
Threshold of probability
Treating all
Treating none
PSA model
PHI model
C
Net reduction in interventions x 100 patients
60
40
20
0
0
0,2
0,4
0,6
0,8
1,0
Threshold of probability
PSA model
PHI model