Clinical Chemistry 59:1
000–000 (2013)

Multicenter Evaluation of [−2]Proprostate-Specific Antigen and the Prostate Health Index for Detecting Prostate Cancer

Carsten Stephan,[1,2*] Sébastien Vincendeau,[3] Alain Houlgatte,[4] Henning Cammann,[5] Klaus Jung,[1,2] and Axel Semjonow[6]

BACKGROUND: Total prostate-specific antigen (tPSA) is flawed for prostate cancer (PCa) detection. [−2]proprostate-specific antigen (p2PSA), a molecular isoform of free PSA (fPSA), shows higher specificity compared with tPSA or percentage of free PSA (%fPSA). The prostate health index (Phi), a measure based on p2PSA and calculated as p2PSA/fPSA × √tPSA, was evaluated in a multicenter study for detecting PCa.

METHODS: A total of 1362 patients from 4 different study sites who had tPSA values of 1.6–8.0 μg/L (668 patients with PCa, 694 without PCa) underwent ≥10 core biopsies. Serum concentrations of tPSA, fPSA (both calibrated against a WHO reference material), and p2PSA were measured on Access2 or DxI800 analyzers (Beckman Coulter).

RESULTS: The percentage ratio of p2PSA to fPSA (%p2PSA) and Phi were significantly higher in all PCa subcohorts (positive initial or repeat biopsy result or negative digital rectal examination) ($P < 0.0001$) compared with patients without PCa. Phi had the largest area under the ROC curve (AUC) (AUC = 0.74) and provided significantly better clinical performance for predicting PCa compared with %p2PSA (AUC = 0.72, $P = 0.018$), p2PSA (AUC = 0.63, $P < 0.0001$), %fPSA (AUC = 0.61) or tPSA (AUC = 0.56). Significantly higher median values of Phi were observed for patients with a Gleason score ≥7 (Phi = 60) compared with a Gleason score <7 (Phi = 53; $P = 0.0018$). The proportion of aggressive PCa (Gleason score ≥7) increased with the Phi score.

CONCLUSIONS: The results of this multicenter study show that Phi, compared with tPSA or %fPSA, demonstrated superior clinical performance in detecting PCa at tPSA 1.6–8.0 μg/L (i.e., approximately 2–10 μg/L in traditional calibration) and is better able to detect aggressive PCa.

Prostate-specific antigen (PSA)[7] has certainly revolutionized the clinical practice for detection of prostate cancer (PCa), and large randomized clinical trials have shown reductions in PCa-specific mortality with the use of PSA *(1, 2)*. However, PSA has severe limitations as a screening test. First, PSA has a low specificity, and its positive predictive value is only approximately 25% in a pooled metaanalysis *(3)*, leading to a large number of false-positive results and up to 75% of unnecessary prostate biopsies *(4)*. Second, PSA lacks sensitivity, because up to 30% of PCa cases and among these 10% of aggressive PCa cases can be identified in patients with a PSA below 4 μg/L *(5)*. Finally, overdetection and subsequent overtreatment of indolent PCa was estimated to reach levels of >50% in the European screening program because PSA alone cannot identify aggressive PCa *(6)*. New biomarkers are clearly needed, particularly to improve the detection of aggressive PCa.

Molecular forms of total PSA (tPSA), such as free PSA (fPSA) and its ratio to tPSA (f/tPSA or %fPSA) have been proposed to enhance the clinical performance of PSA but with limited success for the detection of aggressive PCa *(7–9)*. Another form of fPSA, the benign prostate hyperplasia-associated benign PSA, has limited value to detect PCa *(8)* but can improve

[1] Department of Urology, Charité-Universitätsmedizin Berlin, Berlin, Germany; [2] Berlin Institute for Urologic Research, Berlin, Germany; [3] Department of Urology, Hospital Pontchaillou, Rennes, France; [4] Department of Urology, HIA du Val de Grâce, Paris, France; [5] Institute of Medical Informatics, Charité-Universitätsmedizin Berlin, Germany; [6] Prostate Center, Department of Urology, University Hospital Münster, Germany.
* Address correspondence to this author at: Department of Urology, Charité - Universitätsmedizin Berlin, CCM, Charitéplatz 1, D-10117 Berlin, Germany. Fax +49-30 450 515904; e-mail carsten.stephan@charite.de.

Received September 3, 2012; accepted November 2, 2012.
Previously published online at DOI: 10.1373/clinchem.2012.195784

[7] Nonstandard abbreviations: PSA, prostate specific antigen; PCa, prostate cancer; tPSA, total PSA; fPSA, free PSA; proPSA, proprostate-specific antigen; p2PSA, [−2]proprostate-specific antigen; Phi, prostate health index; AUC, area under the ROC curve; ANN, artificial neural network; TRUS, transrectal ultrasound; DRE, digital rectal examination; DCA, decision curve analysis; r_s, Spearman rank correlation coefficient.

specificity within a multivariate model *(10)*. Intact PSA, a further subfraction of fPSA, when combined with tPSA and fPSA, has demonstrated a discriminative ability between patients with and without PCa *(11)* but further studies are lacking. Further PSA complexes and their limitations have been reviewed elsewhere *(12)*.

Recently, several studies demonstrated significant improvement in PCa detection using a PSA isoform [−2]proprostate-specific antigen (proPSA) (p2PSA) and its percentage derivative %p2PSA [p2PSA/(fPSA × 1000) × 100], especially within multivariate models *(13, 14)*. A further improvement in PCa diagnosis was seen with the additional inclusion of tPSA to %p2PSA by developing the prostate health index (Phi), calculated as p2PSA/fPSA × $\sqrt{}$tPSA *(8, 15–17)*. In screening populations *(8, 17)* and in referred populations *(15, 16)*, %p2PSA (based on 2 markers) and even more Phi (based on 3 markers) demonstrated better diagnostic performance than tPSA and %fPSA for PCa detection. This was indicated by larger areas under the ROC curve (AUC) and better specificities at high sensitivities, results that indicate a potential reduction of unnecessary biopsies. A relationship between %p2PSA or Phi with biopsy or pathological Gleason score also suggested that these biomarkers may more accurately detect aggressive PCa *(8, 15, 16)*.

In the multicenter study presented in this report, the clinical performance of p2PSA and its derivatives, such as %p2PSA and Phi, was assessed for the detection of PCa at initial and repeat biopsies. The data were obtained from a large cohort of patients with tPSA of 1.6–8 μg/L (calibration against a WHO PSA reference material) corresponding to the former 2–10 μg/L Hybritech calibration. The potential reduction in unnecessary biopsy with %p2PSA and Phi and the preferential detection of aggressive PCa of these biomarkers were evaluated.

Materials and Methods

STUDY POPULATION

At each of the 4 participating institutions, patients referred to the urology department were recruited prospectively and retrospectively between 2003 and 2012 on the basis of having tPSA results between 1.6 and 8.0 μg/L (calibration against a WHO PSA reference material). The numbers of patients in each center were as follows: Münster, n = 544 with 266 PCa; Berlin, n = 393 with 167 PCa; Rennes, n = 227 with 132 PCa; and Paris, n = 198 with 103 PCa. Patients scheduled for initial or repeat prostate biopsy were included in the study. All patients enrolled underwent a transrectal ultrasound (TRUS)-guided needle biopsy (10 or more cores). Several exclusion criteria were defined, including clinical acute/chronic prostatitis or infection of the urinary tract, use of medications that interfere with tPSA serum concentration (e.g., 5-α reductase inhibitors, androgen therapy), previous history of PCa, and previous transurethral resection of the prostate for benign prostatic hyperplasia. The final study population included 1362 men; 681 patients (50%) were included for initial biopsy and 280 patients (21%) were scheduled for a repeated biopsy, and for the remaining 401 patients (29%) this information was missing.

STUDY DESIGN

The study was a multicenter, nonrandomized case control trail to evaluate the diagnostic performance of p2PSA, %p2PSA, and Phi in comparison with the current gold standard markers (tPSA, %fPSA) and other parameters. Participants and investigators were blinded to p2PSA results and the personnel involved in testing were blinded to patients' clinical information. The study was approved by the local hospital ethics committees and reported in accordance with the Standards for the Reporting of Diagnostic Accuracy *(18)*. All patients gave informed consent to participate in the study.

METHODS

Blood was drawn at each participating site and serum samples were prepared and frozen at −20 to −80 °C within 3 h of blood collection according to recommendations for preanalytic tPSA and fPSA *(19)* and p2PSA as previously published *(20)*. All blood samples were obtained before any manipulations involving the prostate and at least 3 weeks after a digital rectal examination (DRE). Serum samples were analyzed on Access2 (Münster and Berlin) or DxI800 instruments (Beckman Coulter) using WHO standard calibrated Access tPSA and fPSA immunoassays or the automated p2PSA assay. Measurements were performed within 1–2 weeks after blood draw or in series. The analytical performance of the measurements assessed with control materials (Beckman Coulter) showed values within the recommended limits.

Prostate volume was determined using TRUS and calculated with the prostate ellipse formula. TRUS-guided biopsies of 10–22 cores were performed on all included patients according to the standard clinical practice routinely used at each participating site. Pathological analysis and Gleason grading were performed at each institution separately according to the updated Gleason grading from the consensus conference in 2005 *(21)*. Samples from PCa patients without the updated Gleason grading or only WHO grading information were not included in analyses regarding Gleason grade (n = 157).

STATISTICAL ANALYSIS

Statistical analyses were performed with MedCalc version 12.2.1 (MedCalc Software). Differences between 2 groups were assessed with the nonparametric Mann–Whitney *U*-test. Multiple groups (4 centers) were compared by use of the nonparametric Kruskal–Wallis test. In cases of nominal data (differences in specificity at 90% and 95% sensitivity in ROC analyses) we used the nonparametric McNemar test. AUCs were estimated according to DeLong et al. *(22)*. ROC curves were used to compare specificities at given sensitivities. Two-sided *P* values <0.05 were considered statistically significant.

Phi, %p2PSA, and p2PSA were added to an artificial neural network (ANN) and to binary logistic regression models with the classic variables age, tPSA, %fPSA, prostate volume, and DRE status *(14, 23)* to evaluate their ability to improve specificity. The models were constructed with the MATLAB Neural Network Toolbox (Mathworks). Each of the ANN models had 3 layers: 1 input layer with 2–8 neurons and 1 hidden layer with 3 or 2 neurons and 1 output neuron, ranging from 0 (low) to 1 (high) PCa risk. To avoid overfitting during training we used the Bayesian regularization algorithm *(24)*. With this method there is no necessity for a special validation data set, and the whole training data set can be used for training. For the respective best models and the single parameters, the calibration plots and decision curve analysis (DCA) *(25)* were used.

Results

For the whole population of 1362 men (any biopsy), Phi, %p2PSA, p2PSA, age, and tPSA were significantly higher in patients with PCa, whereas prostate volume, fPSA, and %fPSA were significantly higher in patients without PCa (Table 1). Only the p2PSA-based parameters %p2PSA and Phi maintained the highest significance level ($P < 0.0001$) when the patient data were subdivided into those with negative DRE or with initial or repeat biopsy.

ROC ANALYSIS

The 4 centers showed similar AUCs for each parameter, with AUCs ranging from 0.55 to 0.56 for tPSA, 0.59 to 0.64 for %fPSA, 0.70 to 0.74 for %p2PSA, and 0.72 to 0.74 for Phi. As evident in Table 2, in the whole population of 1362 men the overall performance of Phi was superior for all parameters and significantly better than that of %p2PSA, the nearest competing measure (AUC 0.74 vs 0.725; $P = 0.022$). Both Phi and %p2PSA significantly outperformed (*P* always <0.0001) all other laboratory (p2PSA, %fPSA, tPSA, fPSA) and clinical parameters (age, prostate volume, DRE status) as judged by AUC comparison (See Fig. 1 in the Data

Table 1. Median values for all, initial biopsy, repeat biopsy, and DRE-negative patients.[a]

		Any biopsy (all)			DRE-negative patients			Initial biopsy			Repeat biopsy		
	All	No PCa	PCa	*P*	No PCa	PCa	*P*	No PCa	PCa	*P*	No PCa	PCa	*P*
Patients, n (%)	1362	694 (51)	668 (49)		593 (56)	466 (44)		297 (44)	384 (56)		165 (59)	115 (41)	
Age, years	64 (63–65)	63 (62–64)	64.5 (64–65)	0.0002	63 (62–64)	64 (63–65)	0.0135	62 (61–63)	65 (64–66)	<0.0001	61 (59–63)	64 (62.6–66)	0.0011
DRE negative, n (%)	1059 (78)	593 (85)	466 (70)		593 (100)	466 (100)		223 (75)	225 (59)		154 (93)	92 (80)	
DRE positive, n (%)	303 (22)	101 (15)	202 (30)		—	—		74 (25)	159 (41)		11 (7)	23 (20)	
Prostate volume, cm^3	39 (38–40)	43 (40–45)	35 (34–37)	<0.0001	43 (40–45)	36 (34–38)	<0.0001	42 (40–45)	35 (32–37)	<0.0001	46 (42–50)	39 (37–43)	0.003
Gleason <7, %		NA[b]	NA	283 (55)		NA	207 (57)		NA	150 (51)		NA	73 (71)
Gleason ≥7, %		NA	NA	228 (45)		NA	156 (43)		NA	143 (49)		NA	30 (29)
tPSA, μg/L	4.68 (4.57–4.83)	4.54 (4.27–4.69)	4.85 (4.67–5.36)	0.0002	4.59 (4.34–4.79)	4.92 (4.64–5.15)	0.002	4.67 (4.49–4.88)	4.92 (4.67–5.14)	0.0097	5.15 (4.85–5.55)	5.54 (5.14–5.95)	0.208
fPSA, μg/L	0.61 (0.58–0.64)	0.64 (0.61–0.68)	0.57 (0.55–0.60)	0.0016	0.64 (0.61–0.68)	0.57 (0.54–0.61)	0.0009	0.65 (0.60–0.70)	0.60 (0.54–0.65)	0.0221	0.70 (0.63–0.80)	0.61 (0.57–0.72)	0.216
%fPSA	14.2 (13.7–14.7)	15.4 (15.0–16.0)	12.6 (12.3–13.2)	<0.0001	15.3 (14.8–15.8)	12.6 (11.8–13.0)	<0.0001	15.1 (14.2–15.6)	12.5 (11.6–13.2)	<0.0001	15.6 (13.9–16.7)	12.8 (12.1–14.2)	0.014
p2PSA, ng/L	12.4 (11.9–12.9)	10.9 (10.5–11.4)	14.1 (13.5–14.8)	<0.0001	10.8 (10.4–11.3)	13.5 (12.8–14.1)	<0.0001	11.4 (10.9–12.2)	15.2 (14.4–17.1)	<0.0001	11.7 (10.5–13.9)	14.1 (13.2–16.7)	0.0002
%p2PSA	2.11 (2.03–2.18)	1.73 (1.66–1.80)	2.49 (2.39–2.57)	<0.0001	1.70 (1.63–1.78)	2.39 (2.28–2.50)	<0.0001	1.78 (1.69–1.91)	2.53 (2.42–2.75)	<0.0001	1.64 (1.53–1.72)	2.43 (2.19–2.67)	<0.0001
Phi	43.2 (41.8–44.6)	35.5 (34.2–36.8)	53.8 (51.7–56.6)	<0.0001	34.7 (33.6–36.5)	52.1 (49.8–55.6)	<0.0001	37.4 (35.4–39.7)	56.4 (52.9–59.9)	<0.0001	36.6 (34.0–39.3)	55.0 (50.6–60.0)	<0.0001

[a] All values given as median (95% CI).
[b] NA, not applicable.

Table 2. Univariate and multivariate analyses for parameters and models.[a]

Predictors	Any biopsy		Initial biopsy		Repeat biopsy	
	Univariate AUC (95% CI); *P*	Multivariate AUC (95% CI)	Univariate AUC (95% CI); *P*	Multivariate AUC (95% CI)	Univariate AUC (95% CI); *P*	Multivariate AUC (95% CI)
Age	0.56 (0.53–0.59); <0.0001		0.60 (0.56–0.64); <0.0001		0.62 (0.55–0.68); 0.00015	
Prostate volume, cm^3	0.62 (0.59–0.64); <0.0001		0.62 (0.58–0.67); <0.0001		0.60 (0.54–0.67); 0.0019	
DRE	0.58 (0.56–0.60); <0.0001		0.58 (0.55–0.62); <0.0001		0.56 (0.52–0.61); 0.0008	
tPSA, μg/L	0.56 (0.53–0.59); <0.0001		0.56 (0.51–0.61); <0.0001		0.54 (0.47–0.61); 0.21	
%fPSA	0.61 (0.59–0.64); <0.0001		0.60 (0.56–0.64); <0.0001		0.59 (0.52–0.65); 0.011	
%p2PSA	0.72 (0.70–0.75); <0.0001		0.72 (0.69–0.76); <0.0001		0.74 (0.68–0.80); <0.0001	
Phi	0.74 (0.71–0.76); <0.0001		0.73 (0.69–0.77); <0.0001		0.74 (0.68–0.80); <0.0001	
Base model[a]		0.69 (0.66–0.72)		0.69 (0.65–0.73)		0.74 (0.67–0.80)
Base model + %p2PSA		0.75 (0.72–0.77)		0.73 (0.69–0.77)		0.79 (0.74–0.84)
Base model + Phi		0.75 (0.72–0.78)		0.73 (0.69–0.77)		0.80 (0.74–0.85)
Gain in predictive accuracy for %p2PSA		0.06		0.04		0.05
Gain in predictive accuracy for Phi		0.06		0.04		0.06

[a] Base model includes: age, prostate volume, DRE, tPSA, %fPSA.

Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol59/issue1). As also shown in online Supplemental Fig. 1, an ANN including %p2PSA or Phi did not significantly enhance the AUC over Phi alone (0.75 vs 0.74; $P = 0.41$). However, the gain in predictive accuracy of Phi compared with a base model (based on tPSA, %fPSA, age, prostate volume, and DRE status) was visible in all subcohorts with observed increases in AUC of 0.04–0.06 (Table 2). The specificities at 90% and 95% sensitivity also demonstrated an advantage for Phi (35.4% and 15%) and %p2PSA (33.6% and 13.3%) compared with p2PSA (20.8% and 10.8%), %fPSA (15% and 7.5%), and tPSA (16.1% and 8.5%).

The calibration curves for both models with (model 1) and without Phi (model 2) were quite similar and excellent (see online Supplemental Fig. 2).

ASSOCIATION OF Phi WITH POSITIVE BIOPSY RATE AND AGGRESSIVE PCa

Increasing rates of a PCa-positive biopsy sample with increasing Phi values were seen in all (Fig. 1A), initial (Fig. 1B), and repeat biopsy (Fig. 1C) patients. Differences between tumors with Gleason scores <7 vs scores ≥7 were seen for all laboratory values (Table 3). Phi provided the best discrimination between aggressive and nonaggressive PCa. The number of tumors with Gleason scores ≥7 increased with increasing Phi values, reaching approximately 50% of all detected cancers when Phi was >50 (Fig. 2). There was a positive relationship between Phi and biopsy Gleason score with a Spearman rank correlation coefficient (r_s) of 0.154 (CI, 0.06–0.24; $P = 0.0005$). Classification of all 1362 patients into quartiles based on Phi (5–31.7; 31.8–43.1; 43.2–60.6; 60.7–260) revealed an increasing likelihood of harboring PCa, with PCa rates of 21.5% in the lowest Phi quartile, 40.8% in the 2nd quartile, 57.1% in the 3rd quartile, and 76.8% of patients in the highest Phi quartile. When we analyzed the 511 PCa patients with an available Gleason score (Phi median, 56), the tumors with Gleason scores ≥7 accounted for 39.5% of all cancers when Phi was below the median, whereas above the median the percentage of aggressive PCa cases (Gleason score ≥7) increased to 49.8%. It was difficult to obtain a practicable Phi cutoff for detection of aggressive PCa. With the use of data only from tumors with Gleason scores <7 (n = 283) and ≥7 (n = 51) the Phi value at 90% sensitivity was 40 and at the highest efficacy Phi was 49.5. However, these cutoffs did not allow significant improvement in the differentiation between the Gleason 4 + 3 = 7 and 3 + 4 = 7 tumors. With 40 as a cutoff, 84% of the more aggressive Gleason 4 + 3 tumors were above this value but only 18% of the Gleason 3 + 4 tumors were below this value. With the use of the 49.5 cutoff for Phi, 76% of Gleason 4 + 3 tumors were above this value and 47% of the Gleason 3 + 4 tumors were below this value.

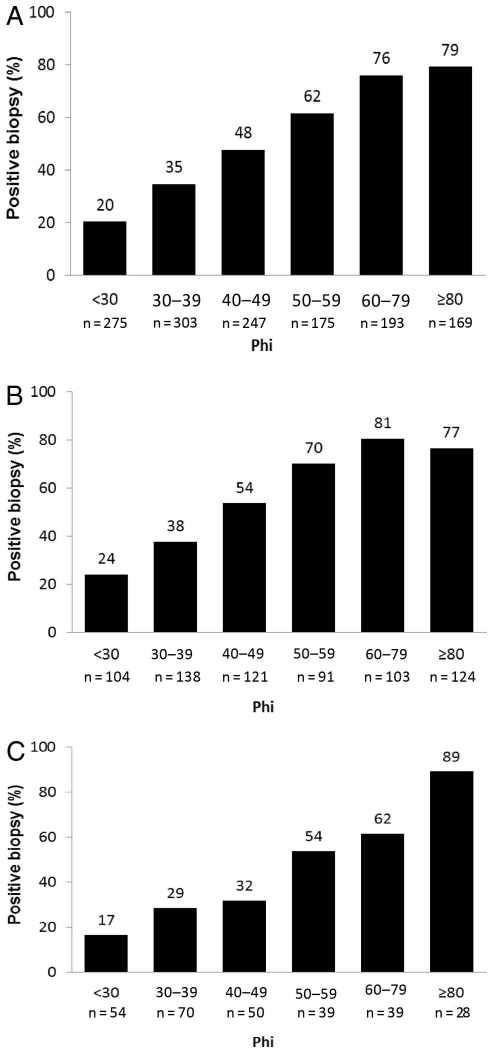

Fig. 1. Percentage of positive biopsies for prostate cancer at various Phi ranges.

(A), All patients; (B), initial biopsy; (C), repeated biopsy.

DECISION CURVE ANALYSIS

The DCA was performed for the models with (model 1) and without p2PSA-based parameters (model 2, Fig. 3A) as well as for the parameters tPSA, %fPSA, %p2PSA, and Phi separately (Fig. 3B). There appeared to be a higher net benefit when we used model 1 with Phi compared to the base model (model 2) based on tPSA, %fPSA, age, prostate volume, and DRE status (Fig. 3A). Phi considered separately also provided the highest net benefit compared with %p2PSA and especially with %fPSA and tPSA (Fig. 3B).

Table 3. Relationship with tumor aggressiveness (biopsy Gleason score).[a]

Group/ parameter	Gleason <7 PCa (n = 283)	Gleason ≥7 PCa (n = 228)	*P*
Age, years	63 (62–65)	66 (64–67)	0.02
Volume, cm^3	38 (35–40)	33 (31–37)	0.031
tPSA, μg/L	4.83 (4.57–5.15)	5.33 (5.03–5.77)	0.003
%fPSA, %	13.3 (12.5–14.3)	11.9 (11–12.8)	0.014
p2PSA, ng/L	14.6 (13.5–15.4)	15 (13.7–17.3)	0.28
%p2PSA	2.34 (2.21–2.49)	2.68 (2.51–2.85)	0.011
Phi	53.1 (47.5–56.6)	59.7 (55.3–62.1)	0.002

[a] All values are given as median (95% CI).

Discussion

The introduction of Phi as a p2PSA-based formula clearly increased the specificity of tPSA and %fPSA for the detection of PCa in cohorts from large multicenter trials *(8, 15, 26)* and prospective trials *(16, 26)* and cohorts with only initial biopsies *(16)* and only repeat biopsies *(27)*. Earlier studies using only %p2PSA but not Phi had indicated that proPSA enhances PCa detection in comparison with tPSA and %fPSA *(13, 14, 17)*. Such studies confirmed the first predictions that p2PSA seemed to be the most promising proPSA subform for PCa detection and especially for detection of aggressive PCa *(28–31)*.

The current study assessed the largest population cohort so far, consisting of 1362 patients in the 1.6–8 μg/L tPSA range. The study findings confirmed %p2PSA and Phi as the strongest discriminating parameters between patients with and without PCa (Table 1). The AUC analysis across all 1362 patients showed similar results for Phi (0.74) as were observed in 2 earlier studies *(15, 16)* that used the same tPSA range of 2–10 μg/L (Hybritech calibration), which translates into the WHO standard calibrated tPSA range of 1.6–8 μg/L. Although Catalona et al. *(15)* found an AUC of 0.65 for %fPSA vs 0.7 for Phi in 892 patients, this difference was much larger in the other prospective study involving 268 patients with an AUC of 0.58 for %fPSA and 0.76 for Phi *(16)*. However, both studies included only DRE-negative patients. When we compared these data to those for our 1059 patients with negative DRE, the AUCs for tPSA (0.56), %fPSA (0.62), p2PSA (0.62), %p2PSA (0.73), and Phi (0.74) were almost identical with those obtained for the whole population. This finding might suggest that the future use of Phi could be recommended without considering the DRE status. ROC comparisons at 95% sensitivity revealed specificities of 15% for Phi and 7.5% for %fPSA in this study, nearly identical to the results of Catalona et al. *(15)*, who found specificities of 16% for Phi and 8.4% for %fPSA. By using absolute Phi thresh-

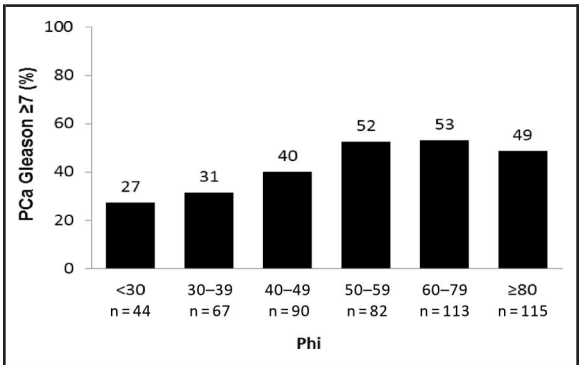

Fig. 2. Proportion of all prostate cancer detected with Gleason score ≥7 in relation to Phi intervals.

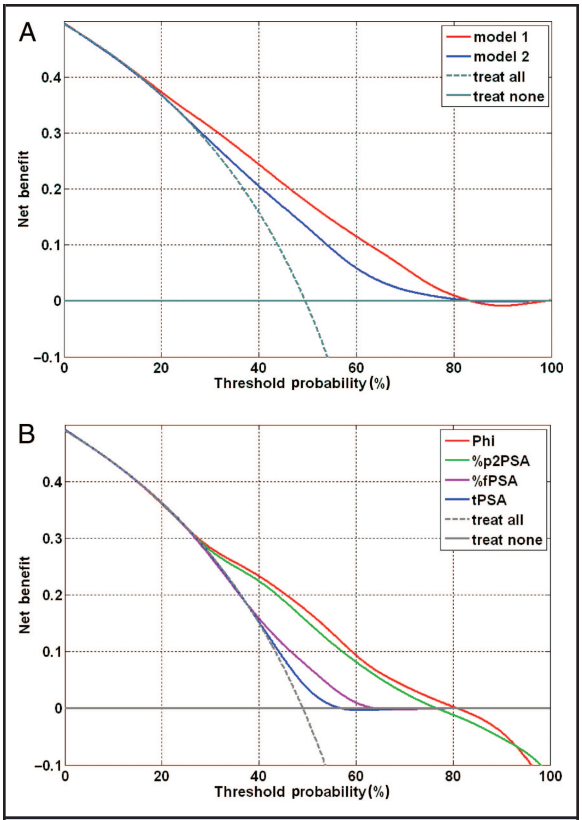

Fig. 3. Decision curve analysis (DCA).
(A), DCA for the models with Phi (model 1) and without Phi (model 2); (B), DCA for tPSA, %fPSA, %p2PSA, and Phi separately.

olds of 24 (Hybritech calibration for tPSA and fPSA) or 31 when using the WHO standard calibrated tPSA and fPSA instead of %fPSA or tPSA, a similar specificity gain of 8% *(15)* and approximately 20% (current study) was evident at 90% sensitivity in these 2 large studies that involved more than 2000 patients. These results reflect a potential economic benefit and reduced risks of biopsy side effects from saving 8%–20% of unnecessary prostate biopsies. Aspects of cost savings have been recently published by Nichol et al. *(32)*, for which the expected 1-year costs for PCa detection were $356 647 lower when using Phi rather than a PSA cutoff of 2 μg/L or $94 219 lower in a model using Phi compared with the traditional tPSA cutoff of 4 μg/L. Unfortunately Guazzoni et al. *(16)* did not provide data at 90% or 95% sensitivity. However, at 90% specificity Phi and %p2PSA had significantly higher sensitivities of 43% and 39% compared with 20% for %fPSA *(16)*.

The detection of aggressive tumors is one of the main concerns in current discussions on the usefulness of PCa biomarkers. Our data in more than 1300 men confirm that with higher Phi values the likelihood of PCa increases (Fig. 1). Although fewer than one-fourth of all patients in the lowest Phi quartile harbor PCa, the fraction with PCa increases to more than three-fourths in those patients in the highest Phi quartile. More importantly, the likelihood of Gleason ≥7 tumors increases with higher Phi scores (Fig. 2). Most parameters show significant differences when we differentiated between Gleason <7 and Gleason ≥7 tumors, but Phi had the largest discriminative power of all parameters (Table 3). Below the Phi median the tumors with Gleason scores ≥7 accounted for 39.5% of all cancers. Above the Phi median, the percentage of tumors with Gleason scores ≥7 increased to almost 50%. In another large multicenter study there was an almost 5-fold increased PCa risk in patients with a Phi value above 55 *(15)*. The risk of having PCa with a Gleason score ≥7 was 1.6-fold higher in the same patient group *(15)*. In another study, even stronger correlations were observed with the Gleason score for Phi (r_s = 0.39) and %p2PSA (r_s = 0.3) *(16)*, but both measures failed to predict Gleason ≥7 tumors. However, the same group of investigators, in a study of 350 men who underwent radical prostatectomy, found that %p2PSA and Phi both could predict pathological Gleason ≥7 PCa, pT3 tumor stage, and a tumor volume of >0.5 mL, and that p2PSA-based parameters were the most accurate predictors for final pathology results *(33)*. These important data suggest a role for p2PSA-based markers not only for PCa diagnosis but also for prediction of aggressiveness and possibly also for prognosis. A further study in a total of 756 men found AUCs for Phi of 0.75 and 0.71 in 2 populations and increased specificities at 90% and 95% sensitivity compared with tPSA and %fPSA *(8)*. In that study the p2PSA-based model also missed the fewest of the Gleason ≥7 tumors at 90% and 95% sensitivity *(8)*.

In addition to investigations of the potential role of p2PSA and Phi for PCa diagnosis *(8, 15, 16, 26, 34)* and staging *(33)*, other potential uses have also been explored. Longitudinal changes of p2PSA with increasing age suggest a role for p2PSA and its derivatives as useful predictors of PCa development *(35)*. Furthermore, a baseline p2PSA value in the upper quartile results in an almost 8-fold risk of developing a PCa *(36)*. For active surveillance, p2PSA-based parameters including Phi appear to provide improved prediction of biopsy reclassification during follow-up *(37)*. Lazzeri et al. *(27)* demonstrated in 222 patients with 1–2 repeat biopsies that with the use of Phi more than 50% of biopsies could have been avoided and that %p2PSA and Phi were the most accurate PCa predictors. These findings are in accord with our results in those 280 patients with repeat biopsies, for whom %p2PSA and Phi reached the highest significance levels (Table 1) and showed the largest AUCs with 0.74 (Table 2).

A recent comparison of Phi with another promising PCa biomarker, PCA3, revealed a larger AUC for Phi compared with PCA3, but the difference was not statistically significant *(38)*. In an analytical multicenter study by Sokoll et al. *(39)*, the p2PSA assay showed a clinically acceptable analytical performance with excellent precision and reproducibility and had negligible interference from other PSA isoforms.

It is also important to note that our present study is one of the first to use the WHO standard calibrated tPSA and fPSA assays in conjunction with p2PSA. Other studies on smaller populations still used the traditional Hybritech calibration *(15, 16)* or did not provide this information *(27)*. We believe that researchers should pay attention to this aspect because median values for Phi tend to be higher with the use of the WHO calibration. Whereas Guazzoni et al. *(16)* found median values of 44.3 and 33.1 in PCa and non-PCa patients, Jansen et al. reported medians of 36.7 and 26.1. The present study has higher values, with 53.8 and 35.5 for PCa and non PCa patients.

In summary, in the largest cohort of patients reported so far who had tPSA in the WHO standard calibrated range of 1.6–8 μg/L, we found %p2PSA and Phi were the best parameters to predict PCa. These two parameters provided the largest AUCs of 0.72–0.74 in all investigated cohorts, with a small overall advantage for Phi. With increasing Phi values, the percentage of PCa-positive biopsy results increased. Phi was significantly higher in patients with Gleason score ≥7 tumors compared with Gleason score <7 tumors ($P = 0.0018$).

The proportion of aggressive PCa (Gleason score ≥7) increased with the Phi score.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: None declared.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honoraria: A. Semjonow, Beckman Coulter Speakers and Advisory Board Honoraria, Advisory Meeting, Spüeakers honoraria from Ferring, GenProbe, Glaxo Smith Kline, Ipsen, and Takeda.
Research Funding: A. Semjonow, Beckman Coulter, Marie-Curie project EU, and German Cancer Aid.
Expert Testimony: None declared.
Patents: A. Semjonow, PCT/EP03/04037.
Other Remuneration: A. Semjonow, Beckman Coulter.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

References

1. Hugosson J, Carlsson S, Aus G, Bergdahl S, Khatami A, Lodding P, et al. Mortality results from the Goteborg randomised population-based prostate-cancer screening trial. Lancet Oncol 2010;11:725–32.
2. Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, et al. Prostate-cancer mortality at 11 years of follow-up. N Engl J Med 2012;366:981–90.
3. Mistry K, Cable G. Meta-analysis of prostate-specific antigen and digital rectal examination as screening tests for prostate carcinoma. J Am Board Fam Pract 2003;16:95–101.
4. Postma R, Schroder FH. Screening for prostate cancer. Eur J Cancer 2005;41:825–33.
5. Thompson IM, Ankerst DP, Chi C, Goodman PJ, Tangen CM, Lucia MS, et al. Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. J Natl Cancer Inst 2006;98:529–34.
6. Draisma G, Etzioni R, Tsodikov A, Mariotto A, Wever E, Gulati R, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. J Natl Cancer Inst 2009;101:374–83.
7. 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;279:1542–7.
8. Jansen FH, van Schaik RH, Kurstjens J, Horninger W, Klocker H, Bektic J, et al. Prostate-specific antigen (PSA) isoform p2PSA in combination with total PSA and free PSA improves diagnostic accuracy in prostate cancer detection. Eur Urol 2010;57:921–7.
9. Roddam AW, Duffy MJ, Hamdy FC, Ward AM, Patnick J, Price CP, et al. Use of prostate-specific antigen (PSA) isoforms for the detection of prostate cancer in men with a PSA level of 2–10 ng/ml: systematic review and meta-analysis. Eur Urol 2005;48:386–99.
10. Stephan C, Cammann H, Deger S, Schrader M, Meyer HA, Miller K, et al. Benign prostatic hyperplasia-associated free prostate-specific antigen improves detection of prostate cancer in an artificial neural network. Urology 2009;74:873–7.
11. Steuber T, Nurmikko P, Haese A, Pettersson K, Graefen M, Hammerer P, et al. Discrimination of benign from malignant prostatic disease by selective measurements of single chain, intact free prostate specific antigen. J Urol 2002;168:1917–22.
12. Stephan C, Rittenhouse H, Cammann H, Lein M, Schrader M, Deger S, et al. New markers and multivariate models for prostate cancer detection. Anticancer Res 2009;29:2589–600.
13. Sokoll LJ, Wang Y, Feng Z, Kagan J, Partin AW, Sanda MG, et al. [-2]proenzyme prostate specific antigen for prostate cancer detection: a national cancer institute early detection research network validation study. J Urol 2008;180:539–43.
14. Stephan C, Kahrs AM, Cammann H, Lein M, Schrader M, Deger S, et al. A [-2]proPSA-based artificial neural network significantly improves differentiation between prostate cancer and benign prostatic diseases. Prostate 2009;69:198–207.
15. 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.
16. Guazzoni G, Nava L, Lazzeri M, Scattoni V, Lughezzani G, Maccagnano C, et al. Prostate-specific antigen (PSA) isoform p2PSA significantly improves the prediction of prostate cancer at initial extended prostate biopsies in patients with total PSA between 2.0 and 10 ng/ml: results of a prospective study in a clinical setting. Eur Urol 2011;60:214–22.
17. Le BV, Griffin CR, Loeb S, Carvalhal GF, Kan D, Baumann NA, Catalona WJ. [-2]Proenzyme prostate specific antigen is more accurate than total and free prostate specific antigen in differentiating prostate cancer from benign disease in a prospective prostate cancer screening study. J Urol 2010;183:1355–9.
18. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Clin Chem 2003; 49:1–6.
19. Woodrum D, York L. Two-year stability of free and total PSA in frozen serum samples. Urology 1998;52:247–51.
20. Semjonow A, Kopke 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.
21. Epstein JI, Allsbrook WC Jr, Amin MB, Egevad LL. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. Am J Surg Pathol 2005;29:1228–42.
22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44: 837–45.
23. Stephan C, Jung K, Soosaipillai A, Yousef GM, Cammann H, Meyer H, et al. Clinical utility of human glandular kallikrein 2 within a neural network for prostate cancer detection. BJU Int 2005;96:521–7.
24. Finne P, Finne R, Auvinen A, Juusela H, Aro J, Maattanen L, et al. Predicting the outcome of prostate biopsy in screen-positive men by a multilayer perceptron network. Urology 2000;56: 418–22.
25. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–74.
26. Sokoll LJ, Sanda MG, Feng Z, Kagan J, Mizrahi IA, Broyles DL, et al. A prospective, multicenter, National Cancer Institute Early Detection Research Network study of [-2]proPSA: improving prostate cancer detection and correlating with cancer aggressiveness. Cancer Epidemiol Biomarkers Prev 2010;19:1193–200.
27. 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.
28. Catalona WJ, Bartsch G, Rittenhouse HG, Evans CL, Linton HJ, Amirkhan A, et al. Serum pro prostate specific antigen improves cancer detection compared to free and complexed prostate specific antigen in men with prostate specific antigen 2 to 4 ng/ml. J Urol 2003;170:2181–5.
29. Catalona WJ, Bartsch G, Rittenhouse HG, Evans CL, Linton HJ, Horninger W, et al. Serum pro-prostate specific antigen preferentially detects aggressive prostate cancers in men with 2 to 4 ng/ml prostate specific antigen. J Urol 2004;171: 2239–44.
30. Mikolajczyk SD, Marker KM, Millar LS, Kumar A, Saedi MS, Payne JK, et al. A truncated precursor

form of prostate-specific antigen is a more specific serum marker of prostate cancer. Cancer Res 2001;61:6958–63.
31. Mikolajczyk SD, Catalona WJ, Evans CL, Linton HJ, Millar LS, Marker KM, et al. Proenzyme forms of prostate-specific antigen in serum improve the detection of prostate cancer. Clin Chem 2004;50: 1017–25.
32. Nichol MB, Wu J, An JJ, Huang J, Denham D, Frencher S, Jacobsen SJ. Budget impact analysis of a new prostate cancer risk index for prostate cancer detection. Prostate Cancer Prostatic Dis 2011;14:253–61.
33. Guazzoni G, Lazzeri M, Nava L, Lughezzani G, Larcher A, Scattoni V, et al. Preoperative prostate-specific antigen isoform p2PSA and its derivatives, %p2PSA and prostate health index, predict pathologic outcomes in patients undergoing radical prostatectomy for prostate cancer. Eur Urol 2012;61:455–66.
34. Houlgatte A, Vincendeau S, Desfemmes F, Ramirez J, Benoist N, Bensalah K, Durand X. [Use of [-2] pro PSA and phi index for early detection of prostate cancer: a prospective of 452 patients]. Prog Urol 2012;22:279–83.
35. Rhodes T, Jacobson DJ, McGree ME, St Sauver JL, Girman CJ, Lieber MM, et al. Longitudinal changes of benign prostate-specific antigen and [-2]proprostate-specific antigen in seven years in a community-based sample of men. Urology 2012;79:655–61.
36. Rhodes T, Jacobson DJ, McGree ME, St Sauver JL, Sarma AV, Girman CJ, et al. Distribution and associations of [-2]proenzyme-prostate specific antigen in community dwelling black and white men. J Urol 2012;187:92–6.
37. Tosoian JJ, Loeb S, Feng Z, Isharwal S, Landis P, Elliot DJ, et al. Association of [−2]proPSA with biopsy reclassification during active surveillance for prostate cancer. J Urol 2012;188: 1131–6.
38. Ferro M, Bruzzese D, Perdona S, Mazzarella C, Marino A, Sorrentino A, et al. Predicting prostate biopsy outcome: prostate health index (phi) and prostate cancer antigen 3 (PCA3) are useful biomarkers. Clin Chim Acta 2012;413:1274–8.
39. Sokoll LJ, Chan DW, Klee GG, Roberts WL, van Schaik RH, Arockiasamy DA, et al. Multi-center analytical performance evaluation of the Access Hybritech(R) p2PSA immunoassay. Clin Chim Acta 2012;413:1279–83.