World J Urol (2007) 25:95–103
DOI 10.1007/s00345-006-0132-9

ORIGINAL ARTICLE

Assay-specific artificial neural networks for five different PSA assays and populations with PSA 2–10 ng/ml in 4,480 men

Carsten Stephan · Chuanliang Xu · Henning Cammann · Markus Graefen · Alexander Haese · Hartwig Huland · Axel Semjonow · Eleftherios P. Diamandis · Mesut Remzi · Bob Djavan · Mark F. Wildhagen · Bert G. Blijenberg · Patrik Finne · Ulf-Hakan Stenman · Klaus Jung · Hellmuth-Alexander Meyer

Received: 22 September 2006 / Accepted: 26 October 2006 / Published online: 28 February 2007

Abstract Use of percent free PSA (%fPSA) and artificial neural networks (ANNs) can eliminate unnecessary prostate biopsies. In a total of 4,480 patients from five centers with PSA concentrations in the range of 2–10 ng/ml an IMMULITE PSA-based ANN (iANN) was compared with other PSA assay-adapted ANNs (nANNs) to investigate the impact of different PSA assays. ANN data were generated with PSA, fPSA (assays from Abbott, Beckman, DPC, Roche or Wallac), age, prostate volume, and DRE status. In 15 different ROC analyses, the area under the curve (AUC) in the PSA ranges 2–4, 2–10, and 4–10 ng/ml for the nANN was always significantly larger than the AUC for %fPSA or PSA. The nANN and logistic regression models mostly also performed better than the iANN. Therefore, for each patient population, PSA assay-specific ANNs should be used to optimize the ANN outcome in order to reduce the number of unnecessary biopsies.

Keywords Prostate cancer · Prostate-specific antigen · Receiver operating characteristic curve · Artificial neural network · Prostate biopsy

Introduction

Early prostate cancer (PCa) detection is considerably enhanced by measurement of prostate specific antigen (PSA) [1]. However, PSA lacks specificity, since elevated

C. Stephan (✉) · C. Xu · K. Jung · H.-A. Meyer
Department of Urology, Charité—Universitätsmedizin Berlin, CCM, Charitéplatz 1, 10098 Berlin, Germany
e-mail: carsten.stephan@charite.de

H. Cammann
Institute of Medical Informatics, Charité—Universitätsmedizin Berlin, Berlin, Germany

C. Xu
Department of Urology, Changhai Hospital, Second Military Medical University, Shanghai, China

M. Graefen · A. Haese · H. Huland
Department of Urology, University Hospital Hamburg-Eppendorf, University Hamburg, Hamburg, Germany

A. Semjonow
Prostate Center, University Hospital, University Münster, Münster, Germany

E. P. Diamandis
Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Canada

M. Remzi · B. Djavan
Department of Urology, University Hospital of Vienna, Vienna, Austria

M. F. Wildhagen
Department of Urology, University Hospital Rotterdam, Rotterdam, The Netherlands

B. G. Blijenberg
Department of Clinical Chemistry, University Hospital Rotterdam, Rotterdam, The Netherlands

P. Finne · U.-H. Stenman
Department of Clinical Chemistry, Helsinki University Central Hospital, Helsinki, Finland

PSA concentrations are also found in patients with benign prostatic diseases. Especially in the 4–10 ng/ml PSA "gray zone", this serum test alone cannot distinguish between PCa and benign prostatic hyperplasia (BPH). In addition, PSA values <4 ng/ml do not indicate the absence of PCa, because the PCa detection rate is similar to the 4–10 ng/ml PSA range [2]. Measurements of the two major molecular forms of PSA have been shown to improve specificity over total PSA (tPSA) alone [3, 4]. Using the ratio of free PSA (fPSA) to tPSA (%fPSA) in the tPSA range of 4–10 ng/ml, approximately 20–25% of unnecessary biopsies can be avoided [5, 6]. For tPSA values <4 ng/ml the use of the ratio PSA to prostate volume [7] and %fPSA have also been reported to increase specificity [8, 9].

Although prostate biopsy is required for the diagnosis of PCa, this invasive and expensive procedure should be avoided in men with a low probability of harboring PCa. Therefore, %fPSA has been proposed as a primary decision tool for first time biopsy in men with unsuspicious digital rectal examination (DRE) [10, 11]. As %fPSA and tPSA are also influenced by factors like prostate volume [6, 12–14] and age [6, 13, 15] different multivariate logistic regression models [16, 17] and artificial neural networks (ANN) [18–22] including these or similar [23] parameters have been introduced to improve cancer specificity. ANNs are computational methods that perform multifactorial analyses based on weighing different signals (input factors) for disease classification. The ANN consists of a group of nodes, termed neurons, in layers that communicate to each other. The categorization of data by an ANN requires prior knowledge of the relevant key data. After training on a database the ANN can classify data, which have not been shown previously. In general, ANNs are able to model complex biological systems by revealing relationships among the input data that cannot always be recognized by conventional analyses [24]. However, even if the applied ANN and logistic regression models show a further improvement in specificity between 11 and 49% compared to %fPSA, they use partially different input data as well as different PSA ranges and, more importantly, results from different PSA assays [16, 18–20, 22]. It has been recommended, that %fPSA should not be used when tPSA and fPSA assays are obtained from different manufacturers [25–28]. Different tPSA assays and subsequently different %fPSA values may also have consequences for the number of recommended biopsies, especially with values near the cutoffs [29].

To our knowledge, the applicability of one trained ANN to patient populations with values from different PSA assays has not been tested so far. Furthermore, a comparison of PSA assay-specific new ANNs (nANN) to original ANNs has not been performed. Therefore, we aimed to use our multicenter-evaluated IMMULITE PSA-based ANN (iANN) [30] on five different patient populations with tPSA and fPSA assays from different manufacturers to answer the following questions:

1. Is the iANN applicable with different PSA assays?
2. Is it necessary to create an ANN for each PSA assay, or is it possible to use the iANN with limitations but better results than conventional tests without ANN?
3. Do population characteristics (referred vs. screened) have an impact on the performance of the respective ANNs?

Materials and methods

A total of 4,480 patients within the tPSA range 2–10 ng/ml (1,505 patients with tPSA concentrations of 2–4 ng/ml) were evaluated in five different centers with five different PSA assays. The distribution of data for tPSA, %fPSA, and age in PCa patients and controls are shown in Table 1. The following centers (including number of patients and manufacturers) participated in this study (sorted by manufacturer):

1. University Hospital Hamburg-Eppendorf, $n = 1841$, Abbott
2. University Hospital of Vienna, $n = 818$, Abbott
3. University Hospital Rotterdam, $n = 955$, Beckman Coulter
4. Westfälische Wilhelms-University Münster, $n = 210$ (201), DPC and Roche
5. Helsinki University Central Hospital, $n = 656$, Wallac

Two of these patient groups comprised screening populations, i.e., the 955 patients from the University Hospital Rotterdam (center 3) and 656 patients from the Finnish prostate cancer screening trial (center 5). These are part of the European randomized study for prostate cancer (ERSPC) and included 25.6 and 22.6% PCa patients, respectively. All other patients (41.5% of them PCa patients) were urologically referred. Indications for biopsy in these referred patients were either a tPSA >2.5 or >4 ng/ml, a suspicious low %fPSA, a suspicious DRE or TRUS result. All patients had a complete data set on tPSA, %fPSA, age, prostate volume, and DRE status.

In all participating centers prostate volume was determined by transrectal ultrasound (TRUS) using the prolate ellipse formula [p/6(transverse diameter ×

Table 1 Patients investigated in the total PSA (tPSA) ranges 2–10, 4–10, and 2–4 ng/ml in five centers

		All patients				Prostate cancer patients				Patients with benign prostates			
Center (number)	PSA range (ng/ml)	Number	tPSA (ng/ml)	fPSA (%)	Age (years)	Number	tPSA (ng/ml)	fPSA (%)	Age (years)	Number	tPSA (ng/ml)	fPSA (%)	Age (years)
Hamburg (1)	2–10	1,841	5.71 (2–10)	17.3 (0.8–91.2)	64.7 (33–91)	837	5.98*	13.8*	65	1,004	5.42	20	64.6
	4–10	1,471	6.31	16.6	64.9	715	6.35	13.4*	65	756	6.3	19.55	64.7
	2–4	370	3.1	20.4	64.4	122	3.28	16.9*	64.8	248	3.07	21.1	64.2
Vienna (2)	2–10	818	5.35 (2.2–10)	27.0 (2.1–95.0)	68 (31–89)	289	6.5*	12.6*	68	529	4.82	33	68
	4–10	559	6.9	22.0	68	226	7.5*	12.1*	68	333	6.4	31	69
	2–4	259	3.22	31.0	66	63	3.3	19.1*	66	196	3.2	34.6	65.5
Rotterdam (3)	2–10	955	3 (2–10)	17.3 (2.6–59.6)	66 (54–75)	245	4.2*	11.6*	66.5	710	2.78	18.6	65.9
	2–4	775	2.8	18.3	66	118	3.2*	15.3*	66.5	657	2.7	18.9	66
Münster (4)	2–10	210	4.2 (2–9.9)	15.0 (3.1–52.9)	64.7 (43–89)	65	5.9*	11.6*	65.5	145	3.5	15.8	64.3
	4–10	109	6.6	12.7	64.8	51	7.2	11.0	66.3	58	5.9	14.7	64.6
	2–4	101	2.7	16.3	64.6	14	3	15.2	64.9	87	2.6	16.5	63.9
Helsinki (5)	4–10	656	5.41 (4–9.96)	18.2 (5.2–55.7)	63.1 (54–68)	148	5.73	14.1*	62.9	508	5.34	19.7	63.2
All center	2–10	4,480	4.9	18.0	65	1,584	5.8*	13.3*	65.5	2,896	4.5	20.2	64.9

Median age, tPSA, and %fPSA and respective ranges (all patients)

*Significances ($P < 0.001$) when compared to patients with benign prostates

anteriorposterior diameter × cephalocaudal diameter)]. A DRE finding nonsuspicious for cancer was defined as negative and a finding suspicious for cancer as positive.

The 1,584 PCa patients (31–91 years of age, prostate volume range: 8.5–165 cm^3) were diagnosed histopathologically by ultrasound-guided prostate biopsy (6–18 cores) as described before [31] and modified with additional lateral cores (reviewed in [32]). The respective median values of each center for tPSA, %fPSA, and age are given in Table 1. Median prostate volume in the different centers ranged from 29.3 to 45 cm^3. Of the PCa patients 49.1% had positive DRE-findings (Table 2). No patient received antiandrogen treatment prior to blood sampling.

The diagnosis of the 2,896 patients with benign prostates was also histopathologically confirmed on the basis of at least one transrectal ultrasound-guided biopsy (6–18 cores).

Serum samples were collected before any prostate manipulation or at least 3–4 weeks after an earlier prostate manipulation. PSA was measured within 2 years after storage between −70 and −80°C. This eliminates potential problems with fPSA stability [33].

Center 1 (University Hospital Hamburg-Eppendorf) and center 2 (University Hospital of Vienna) used the AxSYM Total PSA and AxSYM Free PSA assay (Abbott Laboratories, Abbott Park, IL, USA). Data for center 2 were already partially published [19]. Center 3

Table 2 Prostate volume and digital rectal examination (DRE) status in the five centers

	All patients				Prostate cancer patients				Patients with benign prostates			
Center (number)	Number	Median volume (cm^3)	Range volume (cm^3)	Positive DRE (%)	Number	Median volume (cm^3)	Range volume (cm^3)	Positive DRE (%)	Number	Median volume (cm^3)	Range volume (cm^3)	Positive DRE (%)
1	1,841	53	17–214	43.1	837	45*	17–165	58.9*	1,004	61	18–214	29.9
2	818	37	13–119	39.5	289	33*	13–79	41.5	529	39.4	18–119	38.4
3	955	39.9	15–148	17.8	245	34.9*	15.5–122	36.3*	710	41.2	15–148	11.4
4	210	34.9	7–115	19.1	65	30.8*	13.8–86.5	50.8*	145	35.1	7–115	4.8
5	656	37	8.5–154	15.2	148	29.3*	8.5–91.5	28.4*	508	38.5	9.3–154	11.4
All centers	4,480	42.2	7–214	31.8	1,584	39*	8.5–165	49.1*	2,896	45	7–214	22.4

Median prostate volume, ranges of prostate volume, and percentage of positive (suspicious for PCa) DRE

*Significances ($P < 0.001$), when compared to patients with benign prostates

used the Tandem R Total and Free PSA assay (Beckman Coulter, San Diego, CA, USA). This cohort's data and the assay have been described [34]. Center 4 used two assays. A total of 210 patients had tPSA values between 2 and 10 ng/ml and were analyzed with the IMMULITE Total and IMMULITE Free PSA assays [Diagnostic Products Corporation (DPC), Los Angeles, CA, USA]. Using the Elecsys Total and Free PSA (Roche Diagnostic GmbH, Mannheim, Germany) for the same patients, only 201 patients had values between 2 and 10 ng/ml since nine patients had tPSA level below 2 ng/ml. Data on these patients have been published [26]. Center 5 measured all patients with the Total and Free PSA simultaneous dual label Prostatus immunofluorometric assay (EG&G-Wallac, Turku, Finland). Data on these 656 patients have been published previously [20].

All data were analyzed with the recently introduced multicentric IMMULITE PSA-based iANN, based on data from 1,188 patients [30]. The iANN can be used online at http://www.charite.de/ch/uro/de/html/prostatabiopsie/prostata_en.html with the computer program "ProstataClass".

For this study, all different patient data were normalized. All six (two for center 4) nANN models were constructed with the SPSS module Neural Connection 2.0 (SPSS, Chicago, IL) using feed-forward backpropagation with five input variables: tPSA, %fPSA, patient age, prostate volume, and DRE status. The three hidden layer neurons and the single output neuron used the hyperbolic tangent (tanh) as the transfer function, producing output values between −1 and 1 which can be transformed to represent the probability of PCa. For evaluation of the nANN models the method of tenfold cross validation was used [30]. In brief, each data set was divided randomly into ten subsets. The training-test-run was repeated ten times and each time one of the ten subsets was used as the test set while the other nine subsets were the training set. Results from the ten training-test-runs were taken together and used for receiver operating characteristic (ROC) analysis and estimation of the diagnostic performance for each ANN model for the respective PSA assays. The nANNs were separately generated for the tPSA ranges 2–10, 2–4, and 4–10 ng/ml (where 4–10 means 4.01–10).

We used the statistical software SPSS 10.0 for Windows (SPSS, Chicago, IL). The nonparametric Kruskal–Wallis test of variance, the Mann–Whitney *U* test, and logistic regression analysis (LR) were employed. In LR the explanatory variables were not transformed and no interaction terms were evaluated. The diagnostic validity of LR with its output values, tPSA, %fPSA, and the different ANNs by using the training data was evaluated by ROC curve analysis. Areas under the ROC curve (AUC) were compared using the software GraphROC 2.1 for Windows. Significance was defined as $P < 0.05$.

Results

The distributions of tPSA, %fPSA, and age in PCa and non-cancer patients are given in Table 1. Whereas %fPSA was always (except for the two relatively small subgroups in center 4) lower in the cancer population, age did not differ between the groups. Prostate volume was always significantly larger in the patients with benign prostates. The number of DRE-positive cases was significantly higher in PCa patients, except for center 2 (Table 2). When comparing the AUC values for %fPSA and tPSA in each center, only centers 1, 2, and 5 showed larger values for %fPSA (Table 3). The screening population of center 3 with values mostly in the low tPSA range did not show a difference ($P = 0.43$) between tPSA (AUC: 0.65) and %fPSA (AUC: 0.64) at tPSA 2–4 ng/ml. However, if using the recommended matching procedure [35] for the data at tPSA 2–4 ng/ml to compensate for unequal tPSA distribution before comparing to %fPSA, the AUCs for tPSA and %fPSA are 0.47 and 0.61 showing a significant advantage for %fPSA. Data from all centers show that the larger the PSA range the better tPSA performs. Data from urologically referred patients at center 4 showed for both assays systems (IMMULITE, Elecsys) no significant improvement in AUCs for %fPSA compared to tPSA. The large range of AUCs for tPSA in the different centers, from 0.51 to 0.8, demonstrates large variability and the difficulty to compare results for different populations properly.

Comparison of iANN, LR, and nANN

Table 3 shows 15 different calculations of AUCs (%fPSA, iANN, and nANN) for each center and tPSA range. From centers 1, 2, and 4 (both assays) there were three different AUCs calculated for the respective tPSA ranges and assays, whereas for center 3, data for two tPSA ranges and for center 5, data for one tPSA range were available.

There are four main results:

1. The iANN was significantly better than %fPSA or tPSA in 11 of 15 AUC comparisons (Table 3), demonstrating general advantage of iANN compared to %fPSA regardless of the assay used.
2. The AUC of the nANN was always significantly larger than that of %fPSA or tPSA, if tPSA was

Table 3 The areas under receiver operating characteristic (ROC) curves and confidence intervals (in parentheses), for tPSA, %fPSA, the iANN, and nANNs in five centers

Center	PSA (ng/ml)	Areas under the ROC curves			
		tPSA	%fPSA	iANN	nANN
1	2–10	0.562 (0.537–0.587)[c]	0.681 (0.656–0.706)[c]	0.731 (0.694–0.768)[c]	0.765 (0.743–0.787)
	4–10	0.514 (0.485–0.543)[c]	0.685 (0.658–0.712)[c]	0.742 (0.717–0.767)[b]	0.766 (0.742–0.8)
	2–4	0.569 (0.506–0.632)[c]	0.618 (0.553–0.683)[c]	0.697 (0.642–0.752)[b]	0.744 (0.691–0.797)
2	2–10	0.642 (0.601–0.683)[c]	0.806 (0.777–0.835)[c]	0.772 (0.745–0.799)[c]	0.858 (0.831–0.885)
	4–10	0.629 (0.582–0.676)[c]	0.806 (0.771–0.841)[c]	0.797 (0.76–0.834)[c]	0.879 (0.852–0.906)
	2–4	0.577 (0.532–0.622)[c]	0.77 (0.705–0.835)[b]	0.65 (0.568–0.732)[c]	0.83 (0.769–0.981)
3	2–10	0.796 (0.769–0.833)[c]	0.754 (0.715–0.793)[c]	0.787 (0.754–0.82)[c]	0.854 (0.825–0.883)
	2–4	0.652 (0.591–0.713)[c]	0.64 (0.581–0.699)[c]	0.734 (0.685–0.783)[a]	0.76 (0.709–0.811)
4a DPC	2–10	0.742 (0.669–0.814)[c]	0.661 (0.577–0.745)[c]	0.829 (0.766–0.892)[b]	0.882 (0.833–0.931)
	4–10	0.62 (0.514–0.726)[c]	0.643 (0.537–0.749)[c]	0.802 (0.718–0.886)	0.827 (0.751–0.903)
	2–4	0.604 (0.439–0.769)[c]	0.582 (0.4–0.764)[c]	0.815 (0.674–0.956)	0.845 (0.731–0.957)
4b Roche	2–10	0.774 (0.707–0.841)[c]	0.718 (0.644–0.792)[c]	0.841 (0.782–0.9)[b]	0.897 (0.854–0.94)
	4–10	0.656 (0.552–0.76)[c]	0.684 (0.586–0.782)[c]	0.79 (0.706–0.874)[a]	0.833 (0.759–0.907)
	2–4	0.668 (0.519–0.817)[c]	0.691 (0.532–0.85)[c]	0.884 (0.774–0.994)	0.926 (0.855–0.997)
5	4–10	0.557 (0.502–0.612)[c]	0.725 (0.680–0.770)[c]	0.746 (0.705–0.787)[a]	0.77 (0.729–0.811)

[a] $P < 0.05$, [b] $P < 0.01$, [c] $P < 0.0001$, when compared with the nANN

better than %fPSA. The increase in AUC values ranged from 0.045 to 0.14.

3. Compared to LR, the nANNs had greater AUC values in 13 of 15 comparisons while reaching significance only in 3 of 15 calculations. Only in one case did the LR significantly outperform the nANN. For simplification LR results are described but not shown in the Tables 3 and 4.
4. The nANN (12 of 15) and LR performed significantly better than iANN in 9 of 15 AUC comparisons.

For all centers and all tPSA ranges, comparisons between tPSA, %fPSA, iANN, LR, and nANN were performed for cutoffs of 90% (not shown) and 95% sensitivity. At 95% sensitivity, data (Table 4, data for tPSA 2–10 ng/ml) for all calculated tPSA ranges show a significantly better performance of the nANNs compared to %fPSA or tPSA in 12 of 15 comparisons with specificity increases of 1.5–39.4% (data for tPSA 2–4 and 4–10 ng/ml not shown separately). Only the results from center 2 failed to reach significance, although specificity for nANN was higher than for %fPSA. LR was 12 times and iANN 6 times (out of 15) significantly better than %fPSA. In 14 of 15 comparisons, the nANN performed significantly better than the iANN whereas LR did so in 11 calculations at 95% sensitivity.

Comparison between centers

Centers 1 and 2 measured both tPSA and fPSA with the AxSYM test system (Abbott). %fPSA was significantly better than tPSA in all three tPSA ranges for each center. LR was better than %fPSA with the exception of center 2 at tPSA 2–4 ng/ml ($P = 0.45$) and the nANNs performed significantly better than %fPSA in both centers; whereas the iANNs did so only in center 1. In center 2 the AUCs for the iANNs were lower than the AUCs for %fPSA. All nANNs based on the AxSYM PSA tests were significantly better than the iANNs.

AUC values were also calculated for 240 BPH and 334 PCa patients from center 1 with tPSA in the range 10–20 ng/ml. The AUC values were: 0.53 for tPSA, 0.75 for %fPSA, 0.83 for iANN, 0.84 for nANN, and 0.85 for LR. Both ANNs and LR were significantly better than %fPSA but did not differ from one another.

In the screening population of 955 patients in center 3 with tPSA in the range 2–10 ng/ml only 180 patients (127 PCa and 53 BPH patients, only 19% of all patients) had tPSA concentrations above 4 ng/ml and a separate nANN was not calculated for these. At tPSA 4–10 ng/ml the iANN (AUC: 0.82) was not significantly better than %fPSA (AUC: 0.78, $P = 0.058$) while the AUC of LR (0.87) was significantly better than iANN and %fPSA. As already mentioned, %fPSA did not enhance the diagnostic power of tPSA in this population analyzed by the Tandem assays.

Center 4 used the IMMULITE PSA assays, which is the same assay used in the original iANN, and in addition the Elecsys assay. At tPSA 2–4 and 4–10 ng/ml, %fPSA was not significantly better than tPSA; and at tPSA 2–10 ng/ml, tPSA even outperformed %fPSA determined with each assay. However, iANN, LR, and nANN were always significantly better than tPSA and %fPSA and equivalent to one another ($P = 0.17$–0.98) regardless of the assay used. The iANN and nANN

Table 4 Comparison of specificities with 95% confidence intervals (in parentheses), PPV, and NPV at 95% sensitivity for tPSA, %fPSA, the iANN, and nANNs in five centers at the tPSA range 2–10 ng/ml

Center	Parameter (%)	tPSA	%fPSA[d]	iANN	nANN
1	Specificity	10 (8.45–11.7)[c]	10.6 (9.1–12.4)[c]	12.9 (11.2–14.8)[c]	24.6 (22.4–27)
	PPV	46.8	47	47.7	51.2
	NPV	70.4	71.8	76	85.5
2	Specificity	8.7 (6.8–11)[c]	37.1 (33.6–40.7)	19.5 (16.7–22.5)[c]	38.6 (35.1–42.2)
	PPV	36.5	45.1	39.1	45.7
	NPV	79.3	92.5	87.3	93.2
3	Specificity	20.4 (18–23.1)[c]	15.1 (13–17.5)[c]	26.5 (23.8–29.4)[c]	33.2 (30.3–36.3)
	PPV	29	27.9	30.9	33
	NPV	91.2	89.9	94	95.2
4a DPC	Specificity	17.2 (12–23.3)[c]	5.5 (2.8–9.8)[c]	33.8 (27.3–40.8)[c]	56.6 (49.4–63.5)
	PPV	33.7	30.8	39.2	49.6
	NPV	89.3	72.7	94.2	96.5
4b Roche	Specificity	26.8 (20.4–34)[c]	11.8 (7.5–17.7)[c]	31.5 (24.7–39)[c]	51.2 (43.5–58.8)
	PPV	42.9	38.5	44.6	53
	NPV	89.5	79	90.9	94.2
5[e]	Specificity	4.7 (3.3–6.6)[c]	20.1 (17–23.2)[c]	35.6 (32.1–39.3)	34.3 (30.8–37.9)
	PPV	22.4	25.8	30.1	29.7
	NPV	75	93.6	96.3	96.1

[a] $P < 0.05$, [b] $P < 0.001$, [c] $P < 0.0001$, when compared with the nANNs

[d] The absolute %fPSA-values at the 95% sensitivity cutoffs for the center 1–5 are: 33, 39, 26, 31 (4a and 4b), and 26%

[e] For center 5 (Finland) only data for the tPSA range 4–10 ng/ml were available

performed similarly at tPSA 2–4 and 4–10 ng/ml if using the same assay; but at tPSA 2–10 ng/ml, the nANN outperformed the iANN.

The data from center 5, the Finnish prostate cancer screening trial included only patients with tPSA concentrations higher than 4 ng/ml. Thus it differs from the screening study (center 3) with most patients having tPSA values of 2–4 ng/ml. In the Finnish population, %fPSA enhanced the performance of tPSA and both ANNs and LR were significantly better than %fPSA and the nANN further outperformed iANN in the AUC comparison.

Discussion

More accurate indications for prostate biopsy, aimed at the reduction of unnecessary biopsies, have gained new attention following the discovery of molecular forms of PSA in the early 1990s [3, 4]. Generally %fPSA can improve specificity by approximately 20% compared to tPSA [4–6, 36]. However, in the present study only in three of five centers an advantage of %fPSA over tPSA was observed. This may be explained by the patient population studied, by the assay used or by an unequal tPSA distribution [35, 37]. In center 3 the possible reason could be the relatively large proportion of patients (~81%) with tPSA in the low range 2–4 ng/ml. Most of the non-cancer patients in this screening study had low tPSA concentrations around 2 ng/ml and therefore tPSA alone was already a good discriminator. The better performance of %fPSA in the Finnish screening study with tPSA levels of 4–10 ng/ml may be explained by the higher tPSA range but also by the different PSA assays. However, Catalona et al. [5] and Partin et al. [6] used the same tPSA and fPSA assays as center 3 and found an improved performance of %fPSA compared to tPSA. In a further analysis of 965 patients in the narrow tPSA range 2.6–4 ng/ml by Roehl et al. [38], the authors concluded that %fPSA provides risk assessment but does not eliminate many unnecessary biopsies. Data from center 4 with two other PSA assays also showed that %fPSA did not enhance the power of tPSA. This may be partially explained by the fact that accurate measurement of %fPSA becomes increasingly difficult at low PSA concentrations. In the Wallac assay much of the inaccuracy in the determination of %fPSA has been eliminated by simultaneous measurement of free and tPSA by using a dual label assay. This may contribute to the good performance of %fPSA in center 5.

However, regardless of the assay used ANNs based on tPSA, %fPSA, age, prostate volume, and DRE status significantly enhance the performance of tPSA and %fPSA alone. This is a key result from this large study in five different populations with five different tPSA and fPSA assays.

In the largest study on 1,841 patients in center 1, the iANNs were better than %fPSA. The relatively large prostate volumes may explain the relatively low

specificity in center 1 (Table 4). Center 2 used the same PSA assay as center 1 and had the largest improvement of %fPSA of all centers compared to tPSA. However, the advantage of the nANN was lower than in center 1. Most of the 818 patients in center 2 are a subset of the 1,246 patients studied by Djavan et al. [19] where their ANN reached an AUC of 0.913 compared to 0.90 for LR and 0.81 for %fPSA. Data from this study also demonstrated a significant increase in AUC for ANN (0.86) and LR (0.85) over %fPSA (0.81). Despite a similar improvement of the nANN compared to tPSA and %fPSA when analyzing the AUC (center 1 and 2) it is visible that there are large differences between both cohorts with median %fPSA differences in the patients with benign prostates of more than 13% (20 vs. 33%). This shows that the factor of cohort specific differences may be partially more important than the assay used.

Within the screening population from center 3 the iANN outperformed %fPSA and tPSA at the low tPSA range 2–4 ng/ml and also performed relatively well in comparison to the nANN. In contrast, at tPSA 2–10 ng/ml, the iANN was inferior to tPSA, whereas LR and nANN were significantly better than all other parameters.

In center 4 there were no differences between the outcome of the iANN and the respective nANN at tPSA levels 2–4 ng/ml (both assays) and 4–10 ng/ml (IMMULITE assay). However, both iANN and nANN were based on IMMULITE PSA data. This indicates that the iANN can be used if the IMMULITE PSA assay is used in other populations. Even if %fPSA alone was not useful in these patients, the ANNs still improved the discriminatory power of tPSA.

In the Finnish patients (center 5), the AUC increase was also significant when comparing nANN to %fPSA. The iANN was also better than %fPSA ($P = 0.049$); but the nANN was significantly better than iANN. Thus, nANN was again the best discriminator. It remains to be established how the differences between the five centers are related to differences in assay calibration and performance [39] and/or to the characteristics of the patient populations.

Whereas the iANN enhanced the diagnostic outcome of %fPSA in 11 of 15 AUC comparisons, LR and nANNs outperformed %fPSA in 14 or all of 15 calculations and also outperformed the iANN in 9 and 12 of all 15 calculations. The disadvantage of the externally tested iANN compared to the nANNs, which were only internally trained and validated may reflect a better performance of internally versus externally tested algorithms.

The fact that nANNs almost always outperformed the iANN suggests that algorithms based on the same population are better than an algorithm based on a different population, which decreases the bias by the ANN itself. However, this may also indicate that there are differences between the assays used. The use of different study populations also limits this study regarding the comparison of different assays and different ANNs. Parallel measurement with different assays in the same patients will clarify the latter issue.

The comparison of nANNs to LR regarding AUC, 95% sensitivity and 95% specificity cutoffs showed in 15 of 45 calculations significant by higher values for nANN than for LR. The advantage of nANNs compared to LR may be related to the limited data size of the individual centers since both methods perform equally in studies with large cohorts [40].

Conclusion

Use of ANN technology is helpful to assess the patient's risk for PCa and to decide whether a biopsy is indicated. In the future, inclusion of new serum markers in ANNs may further improve the PCa detection rate.

In summary, this study indicates that assay-specific ANNs and partially the iANN significantly enhance the performance of %fPSA reducing the number of unnecessary biopsies within the tPSA range 2–10 ng/ml. Parallel measurements and comparisons of the different assays in a large cohort may answer the question if separate ANNs for different tPSA assays are necessary or if one general ANN may be established for use with any PSA assay combination.

Acknowledgments We gratefully acknowledge Paul E. C. Sibley for his helpful corrections. We thank S. Kreuzer, C. Wülfing, and J. Chun for helpful database support. This work was partly supported by the Mildred-Scheel-Foundation (Grant 70-3295-ST1 to C.S., H.C., K.J.), the Berliner Sparkassenstiftung Medizin (to C.S., H.C.), and the Monika-Kutzner-Stiftung (to C.S., K.J., H.C.).

References

1. Polascik TJ, Oesterling JE, Partin AW (1999) Prostate specific antigen: a decade of discovery—what we have learned and where we are going. J Urol 162:293–306
2. Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, Minasian LM, Ford LG, Lippman SM, Crawford ED, Crowley JJ, Coltman CA Jr (2004) Prevalence of prostate cancer among men with a prostate-specific antigen level < or = 4.0 ng per milliliter. N Engl J Med 350:2239–2246
3. Lilja H, Christensson A, Dahlen U, Matikainen MT, Nilsson O, Pettersson K, Lövgren T (1991) Prostate-specific antigen in serum occurs predominantly in complex with alpha 1–antichymotrypsin. Clin Chem 37:1618–1625

4. Stenman UH, Leinonen J, Alfthan H, Rannikko S, Tuhkanen K, Alfthan O (1991) A complex between prostate-specific antigen and alpha 1–antichymotrypsin is the major form of prostate-specific antigen in serum of patients with prostatic cancer: assay of the complex improves clinical sensitivity for cancer. Cancer Res 51:222–226
5. Catalona WJ, Partin AW, Slawin KM, Brawer MK, Flanigan RC, Patel A, Richie JP, deKernion JB, Walsh PC, Scardino PT, Lange PH, Subong EN, Parson RE, Gasior GH, Loveland KG, Southwick PC (1998) 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 279:1542–1547
6. Partin AW, Catalona WJ, Southwick PC, Subong EN, Gasior GH, Chan DW (1996) Analysis of percent free prostate-specific antigen (PSA) for prostate cancer detection: influence of total PSA, prostate volume, and age. Urology 48:55–61
7. Stephan C, Stroebel G, Heinau M, Lenz A, Roemer A, Lein M, Schnorr D, Loening SA, Jung K (2005) The ratio of prostate-specific antigen (PSA) to prostate volume (PSA density) as a parameter to improve the detection of prostate carcinoma in PSA values in the range of <4 ng/mL. Cancer 104:993–1003
8. Catalona WJ, Partin AW, Finlay JA, Chan DW, Rittenhouse HG, Wolfert RL, Woodrum DL (1999) Use of percentage of free prostate-specific antigen to identify men at high risk of prostate cancer when PSA levels are 2.51 to 4 ng/mL and digital rectal examination is not suspicious for prostate cancer: an alternative model. Urology 54:220–224
9. Jung K, Stephan C, Elgeti U, Lein M, Brux B, Kristiansen G, Rudolph B, Hauptmann S, Schnorr D, Loening SA (2001) Molecular forms of prostate-specific antigen in serum with concentrations of total prostate-specific antigen <4 μg/l—are they useful tools for early detection and screening of prostate cancer? Int J Cancer 93:759–765
10. Lee CT, Scardino PT (2001) Percent free Prostate-specific antigen for first-time prostate biopsy. Urology 57:594–598
11. Stephan C, Lein M, Jung K, Schnorr D, Loening SA (1997) Can prostate specific antigen derivatives reduce the frequency of unnecessary prostate biopsies? [Letter]. J Urol 157:1371
12. Haese A, Graefen M, Noldus J, Hammerer P, Huland E, Huland H (1997) Prostatic volume and ratio of free-to-total prostate specific antigen in patients with prostatic cancer or benign prostatic hyperplasia. J Urol 158:2188–2192
13. Mettlin C, Chesley AE, Murphy GP, Bartsch G, Toi A, Bahnson R, Church P (1999) Association of free PSA percent, total PSA, age, and gland volume in the detection of prostate cancer. Prostate 39:153–158
14. Stephan C, Lein M, Jung K, Schnorr D, Loening SA (1997) The influence of prostate volume on the ratio of free to total prostate specific antigen in serum of patients with prostate carcinoma and benign prostate hyperplasia. Cancer 79:104–109
15. Lein M, Koenig F, Jung K, McGovern FJ, Skates SJ, Schnorr D, Loening SA (1998) The percentage of free prostate specific antigen is an age-independent tumour marker for prostate cancer: establishment of reference ranges in a large population of healthy men. Br J Urol 82:231–236
16. Carlson GD, Calvanese CB, Partin AW (1998) An algorithm combining age, total prostate-specific antigen (PSA), and percent free PSA to predict prostate cancer: results on 4298 cases. Urology 52:455–461
17. Virtanen A, Gomari M, Kranse R, Stenman UH (1999) Estimation of prostate cancer probability by logistic regression: free and total prostate-specific antigen, digital rectal examination, and heredity are significant variables. Clin Chem 45:987–994
18. Babaian RJ, Fritsche H, Ayala A, Bhadkamkar V, Johnston DA, Naccarato W, Zhang Z (2000) 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 56:1000–1006
19. Djavan B, Remzi M, Zlotta A, Seitz C, Snow P, Marberger M (2002) Novel artificial neural network for early detection of prostate cancer. J Clin Oncol 20:921–929
20. Finne P, Finne R, Auvinen A, Juusela H, Aro J, Maattanen L, Hakama M, Rannikko S, Tammela TL, Stenman U (2000) Predicting the outcome of prostate biopsy in screen-positive men by a multilayer perceptron network. Urology 56:418–422
21. Remzi M, Anagnostou T, Ravery V, Zlotta A, Stephan C, Marberger M, Djavan B (2003) An artificial neural network to predict the outcome of repeat prostate biopsies. Urology 62:456–460
22. Stephan C, Jung K, Cammann H, Vogel B, Brux B, Kristiansen G, Rudolph B, Hauptmann S, Lein M, Schnorr D, Sinha P, Loening SA (2002) An artificial neural network considerably improves the diagnostic power of percent free prostate-specific antigen in prostate cancer diagnosis: results of a 5-year investigation. Int J Cancer 99:466–473
23. Kalra P, Togami J, Bansal BSG, Partin AW, Brawer MK, Babaian RJ, Ross LS, Niederberger CS (2003) A neurocomputational model for prostate carcinoma detection. Cancer 98:1849–1854
24. Partin AW, Murphy GP, Brawer MK (2000) Report on prostate cancer tumor marker workshop 1999. Cancer 88:955–963
25. Jung K, Stephan C, Lein M, Henke W, Schnorr D, Brux B, Schürenkämper P, Loening SA (1996) Analytical performance and clinical validity of two free prostate-specific antigen assays compared. Clin Chem 42:1026–1033
26. Oberpenning F, Weining C, Brandt B, De Angelis G, Heinecke A, Hamm M, Stieber P, Hertle L, Schmid HP, Semjonow A (2002) Combining free and total prostate specific antigen assays from different manufacturers: the pitfalls. Eur Urol 42:577–582
27. Patel D, White PA, Milford WA (2000) A comparison of six commercial assays for total and free prostate specific antigen (PSA): the predictive value of the ratio of free to total PSA. BJU Int 85:686–689
28. Semjonow A, Oberpenning F, Brandt B, Zechel C, Brandau W, Hertle L (1996) Impact of free prostate-specific antigen on discordant measurement results of assays for total prostate-specific antigen. Urology 48(Suppl):10–15
29. Yurdakul G, Bangma C, Blijenberg B, van Zelst B, Wildhagen M, van der KT, Schroder F (2002) Different PSA assays lead to detection of prostate cancers with identical histological features. Eur Urol 42:154–158
30. Stephan C, Cammann H, Semjonow A, Diamandis EP, Wymenga LFA, Lein M, Sinha P, Loening SA, Jung K (2002) Multicenter evaluation of an artificial neural network to increase prostate cancer detection rate and reduce unnecessary biopsies. Clin Chem 48:1279–1287
31. Hodge KK, McNeal JE, Terris MK, Stamey TA (1989) Random systematic versus directed ultrasound guided transrectal core biopsies of the prostate. J Urol 142:71–74
32. Romics I (2004) The technique of ultrasound guided prostate biopsy. World J Urol 22:353–356
33. Woodrum D, York L (1998) Two-year stability of free and total PSA in frozen serum samples. Urology 52:247–251
34. Blijenberg BG, Yurdakul G, Van Zelst BD, Bangma CH, Wildhagen MF, Schroder FH (2001) Discordant performance of assays for free and total prostate-specific antigen in relation to the early detection of prostate cancer. BJU Int 88:545–550

35. Jung K, Stephan C, Lein M, Brux B, Sinha P, Schnorr D, Loening SA (2001) Receiver-operating characteristic as a tool for evaluating the diagnostic performance of prostate-specific antigen and its molecular forms—What has to be considered? Prostate 46:307–310
36. Stephan C, Jung K, Lein M, Sinha P, Schnorr D, Loening SA (2000) Molecular forms of prostate-specific antigen and human kallikrein 2 as promising tools for early diagnosis of prostate cancer. Cancer Epidemiol Biomarkers Prev 9:1133–1147
37. Schroder FH, Kranse R (2003) Verification bias and the prostate-specific antigen test—is there a case for a lower threshold for biopsy? N Engl J Med 349:393–395
38. Roehl KA, Antenor JA, Catalona WJ (2002) Robustness of free prostate specific antigen measurements to reduce unnecessary biopsies in the 2.6 to 4.0 ng./ml. range. J Urol 168:922–925
39. Semjonow A, Brandt B, Oberpenning F, Roth S, Hertle L (1996) Discordance of assay methods creates pitfalls for the interpretation of prostate-specific antigen values. Prostate Suppl 7:3–16
40. Sargent DJ (2001) Comparison of artificial neural networks with other statistical approaches. Cancer 91:1636–1642