

# An artificial neural network for five different assay systems of prostate-specific antigen in prostate cancer diagnostics

**Carsten Stephan, Henning Cammann\*, Hellmuth-Alexander Meyer, Christian Müller†, Serdar Deger, Michael Lein and Klaus Jung**

*Department of Urology, \*Institute for Medical Informatics, and †Department of Clinical Chemistry and Laboratory Medicine, Charité-Universitätsmedizin Berlin, Berlin, Germany*

Accepted for publication 29 February 2008

All authors are members of the Biomarkers of Prostate Cancer Group, CCM

Study Type – Diagnostic (exploratory cohort study)  
Level of Evidence 2a

## OBJECTIVE

To compare separate prostate-specific antigen (PSA) assay-specific artificial neural networks (ANN) for discrimination between patients with prostate cancer (PCa) and no evidence of malignancy (NEM).

## PATIENTS AND METHODS

In 780 patients (455 with PCa, 325 with NEM) we measured total PSA (tPSA) and free PSA (fPSA) with five different assays: from Abbott (AxSYM), Beckman Coulter (Access), DPC (Immulite 2000), and Roche (Elecys 2010) and with tPSA and complexed PSA

(cPSA) assays from Bayer (ADVIA Centaur). ANN models were developed with five input factors: tPSA, percentage free/total PSA (%fPSA), age, prostate volume and digital rectal examination status for each assay separately to examine two tPSA ranges of 0–10 and 10–27 ng/mL.

## RESULTS

Compared with the median tPSA concentrations (range from 4.9 [Bayer] to 6.11 ng/mL [DPC]) and especially the median %fPSA values (range from 11.2 [DPC] to 17.4% [Abbott], for tPSA 0–10 ng/mL), the areas under the receiver operating characteristic curves (AUC) for all calculated ANN models did not significantly differ from each other. The AUC were: 0.894 (Abbott), 0.89 (Bayer), 0.895 (Beckman), 0.882 (DPC) and 0.892 (Roche). At 95% sensitivity the

specificities were without significant differences, whereas the individual absolute ANN outputs differed markedly.

## CONCLUSIONS

Despite only slight differences, PSA assay-specific ANN models should be used to optimize the ANN outcome to reduce the number of unnecessary prostate biopsies. We further developed the ANN named 'ProstataClass' to provide clinicians with an easy to use tool in making their decision about follow-up testing.

## KEYWORDS

prostate cancer, PSA, artificial neural network, prostate biopsy

## INTRODUCTION

Measuring the PSA level has limitations as a screening test for prostate cancer (PCa). Especially in the 2–10 ng/mL PSA level 'grey zone', this serum test alone cannot distinguish between PCa and BPH [1]. Measurements of the two major molecular forms of PSA, free PSA (fPSA) and complexed PSA (cPSA), have been shown to improve specificity compared with total PSA (tPSA) alone [2,3]. Using the ratio of fPSA to tPSA (%fPSA) in the tPSA range of 4–10 ng/mL, 20–25% of unnecessary biopsies can be avoided [4,5]. For tPSA values of <4 ng/mL the use of the ratio of PSA level to prostate volume [6] and %fPSA have also been reported to increase specificity [7,8]. Furthermore, cPSA has been shown to be a

better first-line test than tPSA but only the ratio of cPSA/tPSA (%cPSA) performs equivalently to the %fPSA [9].

As the %fPSA and tPSA are influenced by factors such as prostate volume [10–12] and age [10,11], different multivariate logistic regression models [13,14] and artificial neural networks (ANN) [15–20] have been introduced to improve cancer specificity. The use of different ANN models to detect PCa has been reviewed recently [21]. All ANN and logistic regression models show an improvement in specificity when compared with %fPSA, but they use partially different input data such as prostate volume indexes or DRE status. Most importantly, these models are not comparable due to the use of different PSA assays and PSA ranges [13,15,16,20]. It is

highly recommended, that %fPSA should not be calculated when the tPSA and fPSA levels are obtained from different manufacturers because interpretive criteria for %fPSA may not apply to unproven combinations [5,22,23]. This could have unintended consequences for the number of recommended biopsies, especially with %fPSA values near the thresholds [24,25]. Stephan *et al.* [26] recently published a comparison in almost 4500 men using five different PSA assay-specific ANN models on separate populations showing better performance than the first 'ProstataClass' ANN model, which was built with one assay only [20]. However, only parallel measurements and comparisons of the different assays using the same cohort of samples could answer the question of whether PSA assay-specific ANN models are

necessary, or if one general ANN may be established for use with any PSA assay combination [26]. Therefore, we used data from our previous study [27] that compared five different tPSA and fPSA assay systems to elaborate on how PSA assay-specific ANN models answer the following questions: (i) Are there differences of diagnostic accuracy of ANN models dependent on the PSA assay systems?; (ii) Are these differences comparable to those between tPSA and %fPSA?; and (iii) Is it possible to create one ANN for all PSA assays or is there a need for separate, assay-specific ANNs?

## PATIENTS AND METHODS

In all 798 patients, 465 with PCa and 333 with no evidence of malignancy (NEM), within the tPSA range of 0.49–27 ng/mL based on the Access Hybritech PSA value were evaluated. We retrospectively investigated archival sera collected between 2001 and 2004. Data from 596 of these patients with tPSA concentrations of 0.49–10 ng/mL have been published [27].

Patients with PCa and NEM were histologically confirmed by 8–10 core prostate biopsies. All patients were urologically referred. Blood sampling and handling were performed as described previously [27]. The samples were taken before any diagnostic or therapeutic procedures, and sera were stored at  $-80^{\circ}\text{C}$  until analysed. Prostate volume was determined by TRUS using the prolate ellipse formula. A DRE finding not suspicious for cancer was defined as negative, and a finding suspicious for cancer as positive. The study was carried out in accordance with the standards of the local ethics board and the Helsinki Declaration of 1996.

The tPSA and fPSA (cPSA) were measured using the following commercially available kits: AxSYM (Abbott), ADVIA Centaur (Bayer Diagnostics; cPSA instead of fPSA), Access (Beckman Coulter), Immulite 2000 systems (Diagnostic Products Corp.) and Elecsys 2010 (Roche Diagnostics) as already published [27]. The cPSA values were transformed into fPSA ( $\text{fPSA} = \text{tPSA} - \text{cPSA}$ ) and %fPSA values were calculated as percentage ratios of fPSA to tPSA.

A methodological problem produced concentrations of cPSA > tPSA when measured with the ADVIA Centaur analyser in

18 patients. Consequently these patients were eliminated from further analyses. Thus, 780 patients (455 PCa and 325 NEM) were finally used for all analyses. The between-run imprecision profiles of the measurements were estimated by use of control materials supplied by the manufacturers, commercial control materials, and in-house serum pools; all interassay coefficients of variation were <8% [27].

## ANN MODELS

All 10 ANN models for the five assays and respective two tPSA ranges were constructed with the MATLAB Neural Network Toolbox (The Mathworks, Natick, MA, USA). Feed-forward back-propagation networks were built, in which the input layer consisted of five neurones for the variables tPSA, %fPSA, age, prostate volume and DRE status, with three neurones as hidden layer and one output neurone, ranged from 0 (low PCa risk) to 1 (high PCa risk). To get the best generalization of the ANN we used Bayesian regularization as used by Finne *et al.* [16]. Each ANN model was validated by the leave-one-out method (LOO), which has been previously described in detail [28,29]. In brief, the LOO method is the extreme method of cross validation with the maximum number for the training cohort because  $n$  times the training is performed in  $n-1$  patients and then it is tested in the remaining patient. The output values of the training run were then used to build the receiver operating characteristic (ROC) curve. Because the difference between the mean value of all outputs of the ANN from LOO training ( $n$  ANN models where each was built on  $n-1$  patients) and the output of one ANN built with all data ( $n$  patients) was negligible, we used the latter for all calculations.

To test our ANN classification factor (number of correctly classified patients), the patients were in another step divided into a training (75%) and test group (25%) including the same amounts of patients with PCa and NEM. The ANN models built on 75% of all patients (areas under the ROC curves (AUC) data not shown) showed no differences to those ANN models with all data (100%,  $P$  0.22–0.92). Therefore, all ANN data were given for 100% of the patients (by using Bayesian regularization for training and LOO for testing) in the respective groups except for Table 4, where 75% of the data was used to build the ANN that was then used to test 25% of unknown patients on these ANN models.

We chose thresholds of 90% and 95% sensitivity for both tPSA ranges.

We also created one ANN for all assays by using the mean of the five different tPSA and fPSA (cPSA) measurements to test the hypothesis of if one ANN can be used for all assays.

We used software SPSS version 14.0 for Windows (SPSS), MedCalc version 9.3.6.0 (MedCalc), and GraphPad Prism version 4.03 for Windows (GraphPad). The Mann–Whitney  $U$ -test and Kruskal–Wallis test of variance were used. The diagnostic validity of tPSA, %fPSA, and the different ANN models was evaluated by ROC curve analysis. The areas under the ROC curves (AUC) and the specificities at 90% and 95% sensitivity were compared by a nonparametric method using the software GraphROC 2.1 for Windows. Statistical significance was defined as  $P < 0.05$ .

## RESULTS

The data distribution of the final 780 patients for tPSA, %fPSA, volume, age, and percentage of positive DREs in patients with PCa and NEM within the clinically interesting tPSA ranges of <4, 4–10 and >10 ng/mL (based on measurements with the Access Hybritech PSA system, Beckman Coulter) are shown in Table 1. Between the five tPSA assays, the number of patients within these three tPSA ranges differed markedly. Based on important PCa incidence data obtained by Thompson *et al.* [30], which indicated a similar PCa detection rate at tPSA levels of 2–4 ng/mL compared with other data for the 4–10 ng/mL tPSA range, we used only a common tPSA range of 0–10 ng/mL for constructing the ANN models.

The different numbers of patients with tPSA values of 0–10 and >10 ng/mL is primarily a function of the standardization of specific tPSA assays as shown in Table 2. The data column named 'true' includes only those patients with values within the tPSA range of the respective assay. However, for a direct comparison between all assays and the subsequently established ANN models it was necessary to use exactly the same patients within both tPSA ranges for the five assays. Therefore we made further calculations to form so called 'cumulative' groups for both tPSA ranges (Table 2). The condition for the

**TABLE 1** Data distribution of the 780 patients for tPSA, %fPSA, volume, age and percentage of positive DRE in patients with PCa and NEM, within the three different tPSA ranges <4, 4–10 and >10 ng/mL based on measurements with the Access Hybritech tPSA assay

Variable	tPSA range, ng/mL		
	0–4	4–10	10–27
<b>Patients with Pca</b>			
<i>N</i>	68	242	145
Median (range)			
tPSA, ng/mL	2.81* (0.86–3.96)	6.88 (4.04–9.99)	14.4 (10.0–24.0)
%fPSA	14.1* (3.36–38.4)	11.6* (3.10–35.8)	9.97* (3.58–49.9)
Prostate volume, mL	33.4 (14–60)	37.5* (15–110)	39.1* (10–100)
Age, years	63 (46–77)	63* (43–79)	63* (44–77)
Positive DRE, %	62*	59*	63*
<b>Patients with NEM</b>			
<i>N</i>	143	138	44
Median (range)			
tPSA, ng/mL	2.09 (0.49–3.89)	6.38 (4.02–9.97)	15.4 (10.1–27.0)
%fPSA	23.6 (6.74–69.4)	17.7 (4.96–54.9)	15.6 (2.52–56.2)
Prostate volume, mL	42.2 (15–140)	58.3 (13–132)	71.7 (25–180)
Age, years	65 (40–85)	66 (49–84)	67 (54–80)
Positive DRE, %	7	11	7

\* $P < 0.001$  vs patients with NEM.

'cumulative' group was that at least one of the five tPSA concentrations was within the respective range. Therefore the number of patients in the 0–10 and >10 ng/mL ranges were higher (651 and 235, respectively) instead of 591 or 189 when only the Access tPSA concentrations were used for categorization. Therefore, some patient data (106 of 780, 13.6%) were used twice to build the ANN for the tPSA ranges <10 and >10 ng/mL. As seen in Table 2, all the comparisons between the 'true' and 'cumulative' groups show small differences for tPSA, %fPSA and the ANN models in both tPSA ranges ( $P 0.14$ – $1$ ) with one exception (ANN Bayer at tPSA 10–27 ng/mL;  $P = 0.045$ ). The comparison between the five used assays (cumulative group, tPSA range 0–10 ng/mL) had a small difference for the respective AUC for tPSA ( $P 0.11$ – $0.89$ ), %fPSA ( $P 0.05$ – $0.55$ ) and ANN models ( $P 0.07$ – $0.95$ ) with exceptions for %cPSA and %fPSA (Roche), which were significantly lower than the AUC for %fPSA from the three other assays.

When comparing the AUC values of tPSA vs %fPSA for each assay, %fPSA was significantly better than tPSA with one exception (tPSA vs %cPSA, range 10–27 ng/mL,  $P = 0.122$ ). It should also be mentioned that %cPSA performed better than cPSA alone (AUC 0.772 vs 0.753;  $P = 0.045$  at tPSA

0–10 ng/mL). There was a significant improvement for all ANN models compared with %fPSA or %cPSA.

As also seen in Table 2, two 'cumulative' ANN models (Abbott and Bayer) for the 10–27 ng/mL tPSA range had significantly lower AUCs than the other ANN models. Likewise, they show the respective largest AUC in the 'true' group, which indicates a clear dependence on the lower tPSA results in each selected patient group. The AUC for the ANN with the mean values of all five assays (tPSA 0–10 ng/mL) was 0.906 and had no differences from all other ANN models.

The specificities at the 90% and 95% sensitivity thresholds for all assays and both tPSA ranges are shown in Table 3. As can be seen, all ANN models improve the specificity of %fPSA only within the tPSA range 0–10 ng/mL but not for values of >10 ng/mL. The improvement is even larger between the ANN models and %fPSA compared with %fPSA and tPSA because %fPSA was significantly better than tPSA only regarding AUC but not at 90% or 95% sensitivity, while all ANN models were significantly better than %fPSA (and tPSA) in all values at tPSA 0–10 ng/mL.

There are known differences between %fPSA values, which are also seen in the absolute

values for %fPSA in the ANN models (see text below Table 3), but all ANN models showed only a small difference from each other when compared at both sensitivity thresholds. The specificities at the 90% and 95% sensitivity thresholds for the ANN with the mean values of all five assays (tPSA 0–10 ng/mL) were 60% and 71% with no significant differences from all other ANN models. However, for instance at 95% sensitivity the absolute values differed markedly between 18% (DPC) and 28% (Abbott) for %fPSA and between 0.24 (Beckman) and 0.37 (Roche) for the ANN. The same behaviour is seen for the higher tPSA range, reinforcing the questionable use of the ANN built with the mean values of the five assays.

In Table 4 the number of correctly classified patients within the 0–10 ng/mL tPSA range for tPSA, %fPSA (100% of data) and the respective ANN models (25% test population) is given (cumulative group). The ANN data were generated by using those ANN models built with 75% of all data including the ANN created for all assays by using the mean of the five different assays. The ANN created for all assays had on average the lowest rate of correctly classified patients, but still better values than %fPSA. The rate of correctly classified patients is also lower (data not shown) when using the ANN for one assay on the data of the other four assays.

The numbers of correctly classified patients within the 10–27 ng/mL range (not shown) at the 90% and 95% sensitivity thresholds for the five ANN models from the 25% of unknown patients ( $n = 58$ ) were 88% and 89% and the ANN built with the mean values of the five assays identified somewhat less patients with 84.5%.

Based on these data, we developed a new version of the 'ProstataClass' ANN (Appendix) with the possibility to choose the respective assay. The difference of the clinical usable program to our former 'ProstataClass' is the extended tPSA range of 1–27 compared with 2–20 ng/mL and the change to 90% and 95% sensitivity for the whole tPSA instead of using a high specificity at low tPSA concentrations of 2–4 ng/mL [20]. Another ANN model, which was not built on the mean values of all five assays but built on all data of the five assay combination by using each patient five times in each ANN showed a comparable AUC with 0.904 to the other

**TABLE 2** Numbers of patients and AUC for each tPSA assay within the tPSA ranges of 0–10 and >10 ng/mL for tPSA, %fPSA (%cPSA) and the ANN models for the respective 'True' and 'Cumulative' (cumul.) patient groups

Variable	Abbott		Bayer		Beckman		DPC		Roche	
	true	Cumul.	True	Cumul.	True	Cumul.	True	Cumul.	True	Cumul.
No. patients with tPSA levels (ng/mL) of:										
0–10	647	651	643	651	591	651	562	651	596	651
10–27	140	235	136	235	189	235	218	235	184	235
AUC for:										
tPSA										
0–10	0.723	0.726	0.721	0.724	0.704	0.716	0.701	0.721	0.699	0.710
10–27	0.538	0.540	0.615	0.544	0.562	0.561	0.592	0.569	0.570	0.567
%fPSA										
0–10	0.802	0.802	0.775*	0.772*	0.812	0.814	0.806	0.808	0.790	0.785
10–27	0.712	0.741	0.657*	0.687*	0.736	0.751	0.743	0.751	0.711	0.728
ANN										
0–10	0.887	0.894	0.893	0.890	0.893	0.895	0.905	0.882	0.892	0.892
10–27†	0.863	0.801†	0.891†	0.802††	0.858	0.847	0.834	0.859	0.793	0.847

The AUC for the ANN with the mean values of all five assays was 0.906 with no differences from all other ANN models. \*The cPSA values were transformed into fPSA concentrations by use of the equation  $fPSA = tPSA - cPSA$  and then the %fPSA was calculated. †AUC for Abbott and Bayer ANN were significantly lower than the three other AUC (P from <0.001 to 0.02). ††AUC for ANN of respective 'true' group with no difference (P 0.23–0.78) to respective 'cumulative' group except Bayer (P = 0.04).

**TABLE 3** Specificities at the sensitivity thresholds of 90% and 95% for each assay combination within the tPSA ranges of 0–10 and 10–27 ng/mL for tPSA, %fPSA (%cPSA) and the ANN models including CIs (in parenthesis) for the cumulative group

Variable	Abbott		Bayer		Beckman		DPC		Roche	
	0–10	10–27	0–10	10–27	0–10	10–27	0–10	10–27	0–10	10–27
tPSA range, ng/mL										
90% sensitivity	41 (36–46)	15 (7.8–26)	40 (35–45)	13 (6.4–24)	42 (37–47)	19 (11–30)	43 (38–48)	21 (12–32)	40.5 (36–46)	17 (9.2–28)
95% sensitivity	28 (24–32)	3.8 (0.7–12)	28 (24–33)	11 (5.1–21)	28 (24–33)	9.4 (3.8–19)	31 (26–35)	13 (6.4–24)	27.5 (23–32)	7.5 (2.6–17)
%fPSA*†										
90% sensitivity	46 (41–51)	38 (27–50)	47 (42–52)	26 (17–38)	49.5 (44–55)	41.5 (30–54)	48 (43–53)	30 (20–42)	48 (43–53)	30 (20–42)
95% sensitivity	33* (29–38)	13† (6.4–24)	35* (30–40)	19† (11–30)	36* (31–41)	21† (12–32)	33* (29–38)	23† (14–34)	31* (26–36)	15† (7.8–26)
ANN*†										
90% sensitivity	68 (63–72)	38 (27–50)	67 (62–72)	42 (30–54)	66 (61–70)	57 (44–68)	69 (64–74)	58 (46–70)	62 (57–67)	51 (39–63)
95% sensitivity	55* (50–60)	31† (22–44)	52* (47–57)	30† (20–42)	50* (45–55)	34† (23–46)	54* (49–59)	43† (32–56)	56* (51–61)	40† (28–52)

\*The absolute %fPSA values and [ANN-values] at the 95% sensitivity threshold (tPSA 0–10 ng/mL) for the five assays are: 28.1%[0.26] (Abbott), 27.6%[0.26] (Bayer, = 72.4% of %cPSA), 22.2%[0.24] (Beckman), 18.1%[0.27] (DPC) and 20.9%[0.37] (Roche).

†The absolute %fPSA-values and [ANN-values] at the 95% sensitivity threshold (tPSA 10–27 ng/mL) for the five assays are: 26.7%[0.48] (Abbott), 27.2%[0.55] (Bayer, = 72.8% of %cPSA), 20.8%[0.43] (Beckman), 17.4%[0.47] (DPC) and 19.4%[0.45] (Roche).

assay-specific ANN models. However, the problem of very different %fPSA values was visible again making it impossible to establish a useful and clinically acceptable ANN for all

these five assays. A further problem was the inclusion of cPSA, instead of fPSA which can be solved in a separate ANN but not in one valid ANN.

## DISCUSSION

The reduction of unnecessary prostate biopsies has gained new attention by the use

**TABLE 4** The number of correctly classified patients (%) within the tPSA range of 0–10 ng/mL at 90% and 95% sensitivity if using tPSA, %fPSA ( $n = 651$ ) and the respective assay-specific ANN model with 25% of tested patients ( $n = 162$ )

Assay	ANN (mean of five assays)	Abbott	Bayer	Beckman	DPC	Roche
tPSA						
90% sensitivity	n.a.	68.4	67.9	68.2	69.3	67.9
95% sensitivity	n.a.	65.0	65.4	65.1	66.2	64.7
%fPSA						
90% sensitivity	n.a.	70.5	71.1	72.5	71.3	70.8
95% sensitivity	n.a.	67.1	67.6	68.5	66.2	66.2
ANN ( $n = 162$ )						
90% sensitivity	79.6	83.3	83.3	84.0	84.0	82.1
95% sensitivity	74.7	78.4	79.0	72.2	81.5	75.3

*n.a., not available.*

of %fPSA, which improves specificity by 10–20% compared with using tPSA levels [4,11], while other molecular PSA forms or kallikreins have shown limited clinical success [9]. The data of the present study, including recently published data on men in the tPSA range of <10 ng/mL, confirm the earlier findings on %fPSA with a slightly lower specificity increase of 4.2–11.7% compared with tPSA [31]. We are aware that improved %fPSA specificity is partially related to unique aspects of each assay. It has been published that the diagnostic power of %fPSA is weaker at low tPSA concentrations, which may partially explain the fact that accurate measurement of fPSA becomes increasingly difficult at low PSA concentrations [20,32]. Also, at tPSA concentrations of >10 ng/mL the value of %fPSA decreases while tPSA alone is already a strong predictor of PCa. Furthermore, the investigated population (screened or referred) with a possible unequal tPSA distribution may also lower the ability of %fPSA to discriminate between PCa and NEM [26]. Regardless of the assays used or the tPSA range, ANN models based on tPSA, %fPSA, age, prostate volume (indexes), and the DRE status, significantly enhance the performance of tPSA and %fPSA [15–18,20].

Taken together, there are three main results of the present study: (i) All ANN models (AUC  $\approx 0.9$ ) performed significantly better than %fPSA (AUC  $\approx 0.8$ ) and %fPSA (via %cPSA). But cPSA alone was not significantly better than %fPSA (AUC  $\approx 0.7$ ) in all AUC comparisons (Table 2). This shows a general advantage of ANNs compared with %fPSA

regardless of the assay used; (ii) The AUC and specificities at 90% and 95% sensitivity of all ANNs were not different from each other, showing the possibility to establish a powerful ANN with each (tested) assay; (iii) Compared with the assay-specific ANN models, ANN models for all assays based on the mean PSA and %fPSA had comparable ROC data, but in individual cases predicted PCa with less accuracy than the respective ANN for each tPSA assay used.

The average gain in the AUC was 0.1 improving from  $\approx 0.8$  for %fPSA to  $\approx 0.9$  for the respective ANN models. Compared with our former data from 2002 [20], using the first version of the ANN program 'ProstataClass' with an AUC increased from 0.7 (0.71) for %fPSA to 0.85 (0.84) for the ANN in the tPSA range of 4–10 (2–10) ng/mL, the improvement on ROC analyses was now smaller. The fewer patients in the present study (780 vs 1188 in 2002), retrospective vs prospective samples (in 2002), fewer positive DREs, and increasing prostate volumes (between 1996 and 2004) [6] may be responsible for this phenomenon. The discriminative power of %fPSA and an ANN decreases in larger prostate glands because BPH has a greater influence on the fPSA than a concomitant PCa.

Another factor in the diagnostic power of %fPSA is the performance of tPSA. If tPSA shows an AUC of 0.8, as published for one of the five manufacturers [26], there is only a small gap for further improvement for %fPSA or the ANN.

Another difference to the first ANN model are the thresholds used at low tPSA concentrations (<4 ng/mL). About 20–30% of patients with PCa show PSA concentrations in this low range. About 15% of 2950 biopsied men after 7 years follow-up with PSA concentrations of <4 ng/mL had a diagnosis of PCa [30]. Because there are comparable PCa detection rates at tPSA of <4 and 4–10 ng/mL, we decided to use the 90% and 95% sensitivity thresholds for the whole tPSA within the new ANN models. With our used data from 0.5 to 10 ng/mL and a PCa detection rate being almost equal from 2 to 10 ng/mL, we decided to set the clinical threshold for the lowest tPSA to use the ANN 'ProstataClass' at 1 ng/mL.

A comparison of two ANN models in two different populations (screened or referred) showed that both ANN models could not improve specificity of %fPSA in the screening population, while they did in the referred group [33]. Evaluating an ANN model based on the results of different populations is only possible with limitations, as differences in tPSA and %fPSA distributions, as well as different PSA assays or differences in the PCa detection rate from the number of cores per biopsy (6–10) may be more responsible for differences than the ANN itself [26]. The optimal number of correctly classified patients could only be achieved with the respective assay-specific ANNs. A comparison of the respective ANN models to logistic regression regarding AUC, 95% and 90% sensitivity was performed but not shown, as both methods perform equally in studies with large cohorts [34].

One point of discussion in the practical use of our program 'ProstataClass', is that each ANN acts independently of the others. This could lead to a patient with one data set getting different results for each assay combination.

Problems with higher cPSA than tPSA concentrations have recently been published by another group [25]. These methodological problems, as seen in 18 of the present 798 patients must be solved for cPSA to be accepted as a better first-line marker than tPSA. Also, the advantage for cPSA compared with tPSA regarding AUC disappeared at 90% and 95% sensitivity.

New possibilities to further improve ANN models are the inclusion of new serum markers, e.g. different kallikreins [21,28],

macrophage inhibitory cytokine 1 and migration inhibitor factor [29] or proPSA [35]. Nonetheless, these studies with new markers were only partially successful in limited tPSA ranges [28], limited patient groups [35] or only with the additional consideration of prostate volume [29]. Hopefully, proPSA which showed initial promising results [36] may further improve ANN models by including this serum marker.

To conclude, ANN models are helpful to assess the patient's risk for PCa and to decide whether a biopsy is indicated. Only assay-specific ANN models can safely optimize the number of correctly classified patients while a general ANN improves the performance of %fPSA but to a smaller extent than assay-specific ANN. Therefore, we could not establish one ANN for all assays. We developed a new version of the ANN named 'ProstataClass' to provide clinicians with an easy way to use different tPSA and fPSA assays. However, only external use of our program can show its reliability for wider clinical use. The large %fPSA variability we currently see can be decreased with the use of this ANN.

#### ACKNOWLEDGEMENTS

We gratefully acknowledge Charles F. Weinzierl for helpful corrections. This work was partly supported by the Berliner Sparkassenstiftung Medizin, the ProFIT-Program, the Mildred-Scheel-Foundation (Grant 70-3295-ST1), and the Monika-Kutzner-Stiftung.

#### CONFLICT OF INTEREST

None declared.

#### REFERENCES

- Stephan C, Jung K, Diamandis EP, Rittenhouse HG, Lein M, Loening SA. Prostate specific antigen, its molecular forms, and other kallikrein markers for detection of prostate cancer. *Urology* 2002; **59**: 2–8
- Lilja H, Christensson A, Dahlen U *et al.* Prostate-specific antigen in serum occurs predominantly in complex with  $\alpha$ 1-antichymotrypsin. *Clin Chem* 1991; **37**: 1618–25
- Stenman UH, Leinonen J, Alfthan H, Rannikko S, Tuhkanen K, Alfthan O. 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* 1991; **51**: 222–6
- Catalona WJ, Partin AW, Slawin KM *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
- Jung K, Stephan C, Lein M *et al.* Analytical performance and clinical validity of two free prostate-specific antigen assays compared. *Clin Chem* 1996; **42**: 1026–33
- Stephan C, Stroebel G, Heinau M *et al.* 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* 2005; **104**: 993–1003
- Catalona WJ, Partin AW, Finlay JA *et al.* Use of percentage of free prostate-specific antigen to identify men at high risk of prostate cancer when PSA levels are 2.51–4 ng/mL and digital rectal examination is not suspicious for prostate cancer: an alternative model. *Urology* 1999; **54**: 220–4
- Jung K, Stephan C, Elgeti U *et al.* Molecular forms of prostate-specific antigen in serum with concentrations of total prostate-specific antigen <4  $\mu$ g/L: are they useful tools for early detection and screening of prostate cancer? *Int J Cancer* 2001; **93**: 759–65
- Stephan C, Jung K, Lein M, Diamandis EP. PSA and other tissue kallikreins for prostate cancer detection. *Eur J Cancer* 2007; **43**: 1918–26
- Mettlin C, Chesley AE, Murphy GP *et al.* Association of free PSA percent, total PSA, age, and gland volume in the detection of prostate cancer. *Prostate* 1999; **39**: 153–8
- Partin AW, Catalona WJ, Southwick PC, Subong EN, Gasior GH, Chan DW. Analysis of percent free prostate-specific antigen (PSA) for prostate cancer detection: influence of total PSA, prostate volume, and age. *Urology* 1996; **48** (Suppl.): 55–61
- Stephan C, Lein M, Jung K, Schnorr D, Loening SA. 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* 1997; **79**: 104–9
- Carlson GD, Calvanese CB, Partin AW. An algorithm combining age, total prostate-specific antigen (PSA), and percent free PSA to predict prostate cancer: results on 4298 cases. *Urology* 1998; **52**: 455–61
- Virtanen A, Gomari M, Kransse R, Stenman UH. Estimation of prostate cancer probability by logistic regression: free and total prostate-specific antigen, digital rectal examination, and heredity are significant variables. *Clin Chem* 1999; **45**: 987–94
- Babaian RJ, Fritsche H, Ayala A *et al.* Performance of a neural network in detecting prostate cancer in the prostate-specific antigen reflex range of 2.5–4.0 ng/mL. *Urology* 2000; **56**: 1000–6
- Finne P, Finne R, Auvinen A *et al.* Predicting the outcome of prostate biopsy in screen-positive men by a multilayer perceptron network. *Urology* 2000; **56**: 418–22
- Remzi M, Anagnostou T, Ravery V *et al.* An artificial neural network to predict the outcome of repeat prostate biopsies. *Urology* 2003; **62**: 456–60
- Kalra P, Togami J, Bansal BS *et al.* A neurocomputational model for prostate carcinoma detection. *Cancer* 2003; **98**: 1849–54
- Snow PB, Smith DS, Catalona WJ. Artificial neural networks in the diagnosis and prognosis of prostate cancer: a pilot study. *J Urol* 1994; **152**: 1923–6
- Stephan C, Cammann H, Semjonow A *et al.* Multicenter evaluation of an artificial neural network to increase prostate cancer detection rate and reduce unnecessary biopsies. *Clin Chem* 2002; **48**: 1279–87
- Stephan C, Cammann H, Meyer HA, Lein M, Jung K. PSA and new biomarkers within multivariate models to improve early detection of prostate cancer. *Cancer Lett* 2007; **249**: 18–29
- Patel D, White PA, Milford Ward A. 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* 2000; **85**: 686–9

- 23 Semjonow A, Oberpenning F, Brandt B, Zechel C, Brandau W, Hertle L. Impact of free prostate-specific antigen on discordant measurement results of assays for total prostate-specific antigen. *Urology* 1996; **48** (Suppl.): 10–5
- 24 Yurdakul G, Bangma C, Blijenberg B *et al.* Different PSA assays lead to detection of prostate cancers with identical histological features. *Eur Urol* 2002; **42**: 154–8
- 25 Dittadi R, Franceschini R, Fortunato A *et al.* Interchangeability and diagnostic accuracy of two assays for total and free prostate-specific antigen: two not always related items. *Int J Biol Markers* 2007; **22**: 154–8
- 26 Stephan C, Xu C, Cammann H *et al.* Assay-specific artificial neural networks for five different PSA assays and populations with PSA 2–10 ng/mL in 4480 men. *World J Urol* 2007; **25**: 95–103
- 27 Stephan C, Klaas M, Muller C, Schnorr D, Loening SA, Jung K. Interchangeability of measurements of total and free prostate-specific antigen in serum with 5 frequently used assay combinations: an update. *Clin Chem* 2006; **52**: 59–64
- 28 Stephan C, Jung K, Soosaipillai A *et al.* Clinical utility of human glandular kallikrein 2 within a neural network for prostate cancer detection. *BJU Int* 2005; **96**: 521–7
- 29 Stephan C, Xu C, Brown DA *et al.* Three new serum markers for prostate cancer detection within a percent free PSA-based artificial neural network. *Prostate* 2006; **66**: 651–9
- 30 Thompson IM, Pauler DK, Goodman PJ *et al.* Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. *N Engl J Med* 2004; **350**: 2239–46
- 31 Stephan C, Kramer J, Meyer HA *et al.* Different prostate-specific antigen assays give different results on the same blood sample: an obstacle to recommending uniform limits for prostate biopsies. *BJU Int* 2007; **99**: 1427–31
- 32 Roehl KA, Antenor JA, Catalona WJ. Robustness of free prostate specific antigen measurements to reduce unnecessary biopsies in the 2.6–4.0 ng/mL range. *J Urol* 2002; **168**: 922–5
- 33 Stephan C, Xu C, Finne P *et al.* Comparison of two different artificial neural networks for prostate biopsy indication in two different patient populations. *Urology* 2007; **70**: 596–601
- 34 Sargent DJ. Comparison of artificial neural networks with other statistical approaches: results from medical data sets. *Cancer* 2001; **91** (Suppl.): 1636–42
- 35 Stephan C, Meyer HA, Kwiatkowski M *et al.* A (-5, -7) proPSA based artificial neural network to detect prostate cancer. *Eur Urol* 2006; **50**: 1014–20
- 36 Catalona WJ, Bartsch G, Rittenhouse HG *et al.* Serum pro-prostate specific antigen preferentially detects aggressive prostate cancers in men with 2–4 ng/mL prostate specific antigen. *J Urol* 2004; **171**: 2239–44

**Correspondence:** Carsten Stephan, Department of Urology, Charité-Universitätsmedizin Berlin, Charitéplatz 1, D-10117 Berlin, Germany.  
e-mail: carsten.stephan@charite.de

**Abbreviations:** ANN, artificial neural network; PCa, prostate cancer; NEM, no evidence of malignancy; ROC, receiver operating characteristic (curve); tPSA, total PSA; fPSA, free PSA; cPSA, complexed PSA; %cPSA, percentage cPSA; %fPSA, percentage free/total PSA; AUC, area under ROC curve; LOO, leave-one-out.

## APPENDIX

The program 'ProstataClass' with the input parameters: age, tPSA, %fPSA (or cPSA if Bayer is selected), prostate volume and DRE status.

The screenshot shows the 'ProstataClass' software interface. The title bar reads 'CHARITÉ - UNIVERSITÄTSMEDIZIN BERLIN / DEPARTMENT OF UROLOGY & INSTITUTE OF MEDICAL INFORMATICS / GERMANY'. The main window title is 'Risk estimation for prostate cancer Artificial Neural Network (ANN)'. The interface includes a language selector (Deutsch/Englisch), a patient ID field, and input fields for age (50), total PSA (3.9), %fPSA (12), and prostate volume (23). There are radio buttons for DRE status (neg/pos) and a 'Compute' button. Below the input fields, a table displays the patient's data: age 50.0, total PSA 3.90, %fPSA 12.00, prostate volume 23.0, and DRE 0.0. The ANN value is 0.265, and the cut-off (Sens 95%) is 0.237, with a recommendation: 'risk - biopsy recommended'. A bar chart shows the ANN value (0.265) relative to the cut-off (0.237). The bottom of the window has buttons for 'New patient', 'Print picture', and 'Exit'. A footer note states '(\*) PSA WHO-standardized'.