

Preoperative Neural Network Using Combined Magnetic Resonance Imaging Variables, Prostate Specific Antigen, and Gleason Score to Predict Prostate Cancer Recurrence after Radical Prostatectomy

Vassilis Poulakis^{a,*}, Ulrich Witzsch^a, Rachele de Vries^a, Volker Emmerlich^b, Michael Meves^b, Hans-Michael Altmannsberger^c, Eduard Becht^a

^aDepartment of Urology and Pediatric Urology, Krankenhaus Nordwest, Teaching Hospital of the Johann-Wolfgang-Goethe University Frankfurt, Steinbacher Hohl 2-26, D-60488 Frankfurt am Main, Germany

^bDepartment of Radiology, Krankenhaus Nordwest, Teaching Hospital of the Johann-Wolfgang-Goethe University Frankfurt, Frankfurt am Main, Germany

^cPathologic Institute, Krankenhaus Nordwest, Teaching Hospital of the Johann-Wolfgang-Goethe University Frankfurt, Frankfurt am Main, Germany

Accepted 6 July 2004

Available online 28 July 2004

Abstract

Objective: An artificial neural network analysis (ANNA) was developed to predict the biochemical recurrence more effectively than regression models based on the combined use of pelvic coil magnetic resonance imaging (pMRI), prostate specific antigen (PSA) and biopsy Gleason score in patients with clinically organ-confined prostate cancer after radical prostatectomy (RP).

Methods: Two-hundred-and-ten patients undergoing retropubic RP with pelvic lymphadenectomy were evaluated. Predictive study variables included clinical TNM classification, preoperative serum PSA, biopsy Gleason score, transrectal ultrasound (TRUS) findings, and pMRI findings. The predicted result was a biochemical failure (PSA ≥ 0.1 ng/ml). Using a five-way cross-validation method, the predicted ability of ANNA for a validation set of 200 randomly selected patients was compared with those of Cox regression analysis and “Kattan nomogram” by area under the receiver operating characteristic curve (AUC) analysis.

Results: Seventy-three patients (35%) failed at median follow-up of 61 (mean: 60, range: 2–94) months. Using similar input variables, the AUC of ANNA (0.765, 95% Confidence Interval [CI]: 0.704–0.825) was comparable ($p > 0.05$) to those for Cox regression (0.738, 95%CI: 0.691–0.819) and Kattan nomogram (0.728, 95%CI: 0.644–0.819). Contrarily, adding the pMRI findings, the ANNA is significantly ($p < 0.05$) superior to any other predictive model (0.897, 95%CI: 0.841–0.977). The Gleason score represented the most influential predictor (relative weight: 2.4) of PSA recurrence, followed by pMRI (2.2), and PSA (2.0).

Conclusion: ANNA is superior to regression models to predict accurately biochemical recurrence. The relative importance of pMRI and the utility of ANNA to predict the PSA failure in patients referred for RP must be confirmed in further trials.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Prostate; Prostate neoplasm; Recurrence; Prostate resonance imaging; Neural network; Artificial neural network analysis; Prediction

1. Introduction

Several preoperative nomograms have been developed to predict prostate cancer (PCa) stage and risk of disease progression after attempted curative therapy

* Corresponding author. Tel. +49 69 7601 3917; Fax: +49 69 7601 3648.

E-mail address: Poulakis@em.uni-frankfurt.de, vpoulakis@aol.com (V. Poulakis).



[1–5]. Unfortunately, only few studies have predicted the probability of PSA recurrence within 5 years of treatment with more than 70% accuracy [1], which leaves room for improvement. The most popularized nomograms using pretreatment PSA level, clinical stage, and biopsy Gleason sum, the Kattan nomogram [2], was also internationally validated [3]. The addition of other clinical parameters has failed to significantly improve the accuracy of predictive tools [4]. Especially in patients with PSA ≤ 10 ng/ml, clinical stage and pretreatment PSA, the primary predictive parameters in these tools, may be inaccurate in prediction of PSA outcome and reflect primarily the presence of benign prostatic hyperplasia [6,7]. Therefore, there is an imminent need for simplified predictive tools that include novel preoperative variables that can extend the clinical performance of the available nomograms [1].

The objective of this study was to evaluate whether a preoperative staging based on the results of pelvic phase-array MRI (pMRI) can effectively contribute to the prediction of PSA outcome in patients with localized PCa after radical prostatectomy (RP). A artificial neural network analysis (ANNA) was designed to assess if the pMRI findings in a combined staging approach with preoperative PSA and biopsy Gleason score, excluded clinical stage, can improve the predictive ability of conventional predictive models.

2. Material and methods

Between January 1995 and February 1999, 229 consecutive patients who underwent retropubic RP with staging pelvic lymphadenectomy for clinically localized PCa underwent preoperatively a pMRI. Patients received postoperatively immediate adjuvant hormone or radiation therapy before a documented disease recurrence ($n = 17$, 7%) and patients who were lost from follow-up ($n = 2$, 1%) were excluded from the study. Thus, 210 patients were eligible for clinical evaluation. No patient received any neo-adjuvant therapy.

Preoperatively, all patients underwent serum PSA examination (Abbott Laboratories, Abbott Park, IL, USA) and transrectal ultrasound (TRUS) guided systematic sextant needle biopsy using the automatic biopsy gun [8]. In cases of palpable nodularity (stage cT2 or higher, $n = 170$, 81%) or of TRUS abnormalities suspected of PCa (stage TRUS T3, $n = 100$) additional cores were directly obtained from these areas [8]. Tumor grade was determined according to the Gleason grading system. All biopsy and histological specimens were reviewed by one pathologist using the 1997 TNM classification [9] blindly. The RP specimens were studied using the modified Stanford protocol [10]. No distinction was made between focal and established capsular penetration. The clinical stage based solely on digital rectal examination (DRE) was assigned by two urologists according to TNM classification [9]. The biopsy results were not incorporated in the clinical staging.

Anatomic retropubic RP was performed according to a standard technique [11]. The Head of Department (E.B.) supervised all the operations, thus assuring the uniformity.

Three urologists (V.P., U.W., E.B.) trained in TRUS conducted the assessment using a 7.5 MHz transducer (Kretz, Germany). The patient was examined in the lithotomy position. Imaging included sagittal/transverse views of the prostate and periprostatic area from the level of the apex to the seminal vesicles. The gland volume, tumor location and diameter, echo structure of the lesion, extracapsular extension (ECE), seminal vesicle invasion (SVI); and periprostatic fat involvement were assessed.

To minimize artifacts, pMRI was performed 21–56 (mean: 35, median: 31) days after biopsy. The mean interval between pMRI and RP was 3 (median: 3, range: 1–21) days. For all studies, a 1 Tesla magnet was used (Magnetom Impact, Siemens, Germany) with a pelvic phased-array coil. The technical parameters were standardized and reported elsewhere [12,13]. Two radiologists performed all the prospective readings unaware of other each findings. ECE was defined as an infiltration of the periprostatic fat, irregular bulging associated with disruption of the capsule, focal thickening, capsular retraction, and pericapsular spicula. Generally, tumors >2 cm in contact with the prostate capsule but not-penetrating it and at the apex any irregular bulging were considered extraprostatic disease [12–18]. SVI was suspected on a low signal intensity focus in normally bright tissue [12–18].

Patients were evaluated 4 weeks after surgery then at 3 months for 2 years followed by an every-6-month follow-up visit thereafter. At each visit, a DRE was performed and serum PSA level obtained. Disease recurrence was established by the presence of any detectable PSA level (i.e. ≥ 0.1 ng/ml) after RP. The median follow-up was 61 (mean: 60, range: 2–94) months.

A commercially available software (Statistica-Neural-NetworksTM-4.0, StatSoft, Hamburg, Germany) was used to create a feed forward, back propagation error adjustment neural network [23]. ANNA models were developed and validated using a five-way cross-validation method using an algorithm that maintained equal frequency of outcomes in all the sets. In relatively small patient groups, the cross-validation method allows us to have the maximum patient number in validation groups for comparison with the other predictive models [22]. The 210 available cases were split randomly into five sets of 40 cases using an algorithm that maintained equal frequency of outcomes in both sets. The remaining 10 cases were set aside for use as a test data set to test the ANNA during training to identify the early stopping point. The purpose of the test set is to monitor the progress of ANNA model during training. Each of the five sets of 40 cases was labeled as data set A, B, C, D, or E. We used five ANNA for training: four of the five sets as training data and the remaining set as a validation set. For example, for the first ANNA model, using set A of 40 cases as validation test data, sets B, C, D, and E were combined and used as a training set of 160 cases. In the second ANNA, a model trained with sets A, C, D, and E was used set B to be validated and so on. The input variables were classified as either categorical or continuous depending on variable characteristics and software compatibility. Early termination was employed to prevent “network-overtraining”. Wilk’s generalized likelihood ratio test was used to determine the relative importance of each input variable [23].

In the first model, three input parameters, i.e. preoperative serum PSA level, clinical TNM classification and biopsy Gleason score, were used. In the second model, the pMRI findings were added. In the third model, a simplified model using the previous parameters without the clinical classification parameters was constructed. For

Table 1

Distribution of patients' age, clinical classification, biopsy Gleason score, preoperative serum PSA, and pMRI findings by pathological stage (non-organ-confined disease; $n = 102$, 49%) and PSA failure ($n = 73$, 35%) in $n = 210$ men with preoperatively clinically localized prostate cancer

| | No. of subjects (%) | Postoperative pathological findings and follow-up | |
|---------------------------------------|---------------------|--|------------------------------------|
| | | No. with non-organ-confined disease, $n = 102$ (%) | No. with PSA failure, $n = 73$ (%) |
| Age | | | |
| 41–50 | 2 (1) | 1 (50) | 1 (50) |
| 51–60 | 39 (19) | 16 (41) | 13 (33) |
| 61–70 | 141 (67) | 70 (50) | 49 (35) |
| 71–80 | 28 (13) | 15 (54) | 10 (36) |
| Clinical TNM classification | | | |
| cT1a | 2 (1) | 0 | 0 |
| cT1b | 4 (2) | 1 (25) | 1 (25) |
| cT1c | 34 (18) | 8 (24) | 13 (38) |
| cT2a | 84 (40) | 35 (42) | 25 (30) |
| cT2b | 71 (34) | 47 (66) | 25 (35) |
| cT3a | 15 (7) | 11 (73) | 9 (60) |
| Biopsy Gleason score | | | |
| 2–4 | 19 (9) | 3 (16) | 1 (5) |
| 5 | 53 (25) | 14 (26) | 9 (17) |
| 6 | 81 (39) | 42 (52) | 28 (35) |
| 7 | 33 (16) | 20 (61) | 17 (52) |
| 8–10 | 24 (11) | 23 (96) | 18 (75) |
| Preoperative PSA level (ng/ml) | | | |
| 0–4.0 | 15 (8) | 2 (13) | 1 (7) |
| 4.1–10 | 95 (45) | 27 (28) | 20 (21) |
| 10.1–20 | 67 (32) | 46 (69) | 30 (45) |
| >20 | 32 (15) | 27 (84) | 22 (69) |
| Pathologic findings | | | |
| Extracapsular extension | 102 (49) | – | 62 (61) |
| Seminal vesicle invasion | 40 (19) | – | 34 (85) |
| Positive surgical margins | 50 (24) | 41 (82) | 43 (86) |
| Lymph node involvement | 19 (9) | 15 (71) | 16 (84) |
| Pelvic MRI findings | | | |
| Extracapsular extension | 81 (39) | 67 (83) | 50 (62) |
| Seminal vesicle invasion | 38 (18) | 29 (76) | 22 (58) |
| pMRI T2 | 129 (61) | 32 (25) | 23 (18) |
| pMRI T3a | 43 (20) | – | 20 (47) |
| pMRI T3b | 38 (18) | – | 30 (80) |
| TRUS findings | | | |
| Extracapsular extension | 100 (48) | 45 (45) | 28 (28) |
| Seminal vesicle invasion | 37 (18) | 20 (55) | 19 (51) |
| TRUS T2 | 110 (52) | 57 (52) | 20 (18) |
| TRUS T3a | 54 (26) | – | 26 (48) |
| TRUS T3b | 46 (22) | – | 27 (59) |

construction of model 4, the TRUS findings were added to the variables of model 1. The output variable was PSA failure.

Cox regression analysis was performed in the validation data ($n = 200$) of ANNA to define the correlation between each variable and PSA failure. Regression coefficients were used to devise nomograms predicting PSA recurrence-free survival. Each coefficient-based nomogram was subsequently subjected to predictive accuracy testing. Because the data were censored, the traditional area under the receiving operating characteristic curve (AUC) is problematic, and Harrell's measure of concordance was used [24]. For the comparison of predictive ability, the AUC was used [25]

Comparisons of actuarial survival were made using the log-rank test and actuarial freedom from PSA failure was calculated using the Kaplan–Meier method. A p -value of <0.05 was considered statistically significant.

3. Results

At RP, the median age was 66 (mean: 65.2 ± 6.1 , range: 46–79) years; the median PSA value was

Table 2

The area under receiver operating characteristic (ROC) curves (AUR) of each analysis and comparison with area of “Kattan nomogram” in the validation data set ($n = 200$)

| Predictive model | PSA failure | |
|------------------|----------------------|---|
| | Area under ROC curve | 95% Confidence Interval (CI) ^a |
| ANNA model 1 | 0.765 | 0.704–0.825 |
| ANNA model 2 | 0.897 | 0.841–0.977 |
| ANNA model 3 | 0.895 | 0.839–0.976 |
| ANNA model 4 | 0.769 | 0.701–0.833 |
| Cox model 1 | 0.738 | 0.691–0.819 |
| Cox model 2 | 0.779 | 0.714–0.836 |
| Cox model 3 | 0.777 | 0.710–0.829 |
| Cox model 4 | 0.781 | 0.722–0.836 |
| Kattan nomogram | 0.728 | 0.644–0.819 |

Input parameters for: model 1: preoperative serum PSA level, clinical TNM classification and biopsy Gleason score; model 2: the pMRI staging findings were added to the variables; model 3: a simplified model was constructed leaving out from previous parameters the clinical TNM classification; and model 4: the TRUS staging findings were added to the variables of model 1.

^a In case of overlapped 95%CI there is no statistically significant difference between the compared models. That means that ANNA model 2 and ANNA model 3 are equally accurate in prediction of PSA failure. Furthermore these two models are the most accurate prediction models, since their 95%CI are the highest and are not overlapped with the 95%CI of the other models.

11.8 ng/ml (mean: 12.3 ± 6.8 , range: 1.1–49). The average Gleason score of the biopsies was 6.7 ± 1.0 (range: 4–9, median: 6.5). The mean number of positive biopsies was 2.1 (range: 1–6). The distribution of preoperative parameters in non-organ-confined cancer patients and in those with PSA failure is shown in Table 1. There was no statistically significant difference in patient age among patients with organ-confined and no evidence of PSA failure and among those with extra-

prostatic cancer and PSA failure ($p = 0.4$ and $p = 0.5$, respectively). However, as the Gleason score, the clinical or pMRI stage, and the PSA increased, we noted a statistically significant increase ($p < 0.05$) in the percent of patients with non-organ-confined disease and PSA failure.

Using similar input parameters (clinical stage, PSA and Gleason score; Table 2), the AUC of ANNA model 1 for PSA-failure prediction (0.765, 95%CI: 0.704–0.825) was larger than that of Cox model 1 (0.738, 95%CI: 0.691–0.819) and Kattan nomogram (0.728, 95%CI: 0.644–0.819) but it was not statistically significant ($p > 0.05$). With addition of pMRI findings, the AUC of ANNA model 2 becomes significantly ($p < 0.05$) larger (0.897, 95%CI: 0.841–0.977). Removing the clinical staging from the model 2, the AUC of ANNA and Cox model 3 decreased only minimally to 0.895 and 0.777, respectively, showing that the clinical staging did not contribute to an increase in predictability. The ANNA model 2 showed significantly the highest accuracy in comparison to any Cox model and Kattan nomogram ($p < 0.05$). The addition of TRUS-findings, as alternative to MRI findings, results in no improvement in prediction of PSA failure (Table 2). Table 3 shows the cutoff points for the parameters used to generate ANNA and Cox model 3 (best simplified model).

The Cox regression multivariate analysis indicated that pretreatment PSA ($p < 0.0001$), biopsy Gleason score ($p = 0.006$), pMRI stage ($p = 0.002$), and clinical stage ($p = 0.03$) were all independent significant predictors of time to PSA failure. Using the three most significant preoperative parameters (PSA, pMRI findings, and biopsy Gleason score) the nomogram estimated the 5-year PSA failure rates was constructed

Table 3

Internal validity of the simplified artificial neural network analysis (ANNA) and Cox regression model 3 to predict PSA failure after 5 years of follow-up for different cutoff points in the validation data set ($n = 200$)

| Cutoff Level | Performance parameter for ANNA/Cox regression | | | | | | McNemar test (p value) for comparison with the ANNA model by cutoff level of 0.5 | |
|--------------|---|-----------------|---------|---------|--------------|--------|---|--|
| | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | Accuracy (%) | ANNA | | |
| 0.1 | 3/1 | 98/78 | 40/3 | 65/58 | 65/48 | <0.001 | <0.001 | |
| 0.3 | 37/20 | 90/70 | 67/26 | 73/62 | 72/53 | 0.001 | <0.001 | |
| 0.5 | 91/74 | 88/64 | 81/53 | 95/82 | 90/68 | – | <0.001 | |
| 0.7 | 96/80 | 57/46 | 54/44 | 96/81 | 71/58 | 0.004 | <0.001 | |
| 0.9 | 100/84 | 27/15 | 42/35 | 100/65 | 53/40 | <0.001 | <0.001 | |

PPV: positive predictive value, NPV: negative predictive value, AUC: Area under ROC curve. The cutoff level is the level of predictive probability. There are the points (the cutoff levels) over the receiver operator curve (ROC). For example, accuracy of 65% for a cutoff level of 0.5 means that the accuracy of the predictive model with a 50% probability or more for prediction of PSA recurrence is 65%.

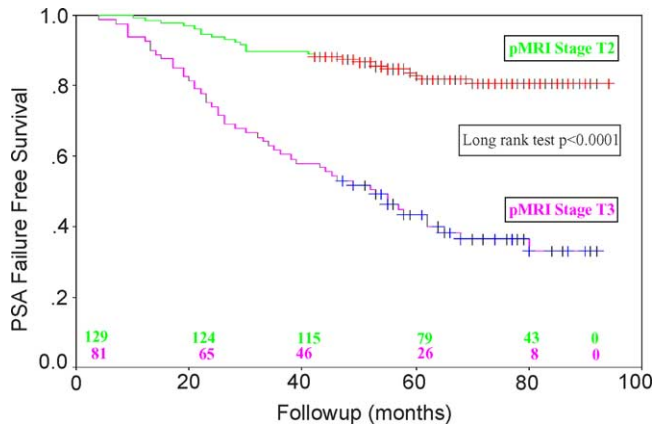


Fig. 1. Kaplan–Meier estimates of PSA failure-free survival probability for 210 patients after radical prostatectomy for clinically organ-confined prostate cancer stratified by pMRI stage (T2 or T3).

Table 4

Relative weights of the predictive variable categories for predicting PSA failure after 5 years of follow-up

| Factor | PSA failure |
|---------------------------------|-------------|
| Biopsy Gleason score | |
| 8–9 | 2.3890 |
| 7 | 1.0645 |
| 4 | –0.0096 |
| Pelvic MRI findings | |
| ECE | 2.2401 |
| SVI | 2.0819 |
| Preoperative PSA (ng/ml) | |
| >20 | 1.9976 |
| 10.01–20 | 1.2012 |
| 4.01–10 | –0.0095 |
| TRUS findings | |
| ECE | 0.0016 |
| SVI | 0.0021 |
| Clinical stage | |
| cT3a | 0.0038 |
| cT2b | 0.0012 |
| cT2a | 0.0010 |

Table 5

Performance of pMRI to identify extracapsular extension (ECE) and seminal vesicle invasion (SVI) according to the pretreatment risk group for postoperative disease recurrence [12] in all patients ($n = 210$)

| Preoperative risk group | No. of patients (total/validation set) | Sensitivity (%) | | | Specificity (%) | | | PPV (%) | | | NPV (%) | | | Accuracy (%) | | |
|-------------------------|--|-----------------|-----|------|-----------------|-----|------|---------|-----|------|---------|-----|------|--------------|-----|------|
| | | ECE | SVI | PSAf | ECE | SVI | PSAf | ECE | SVI | PSAf | ECE | SVI | PSAf | ECE | SVI | PSAf |
| Low risk | 38/36 | 43 | 100 | 80 | 84 | 97 | 73 | 38 | 67 | 53 | 87 | 100 | 90 | 76 | 97 | 75 |
| Intermediate risk | 96/91 | 73 | 50 | 96 | 96 | 99 | 90 | 94 | 83 | 82 | 81 | 94 | 98 | 85 | 94 | 92 |
| High risk | 76/73 | 65 | 76 | 91 | 40 | 85 | 75 | 85 | 76 | 94 | 27 | 85 | 67 | 57 | 82 | 88 |
| Overall | 210/200 | 66 | 73 | 91 | 87 | 95 | 88 | 83 | 76 | 81 | 73 | 94 | 95 | 77 | 90 | 90 |

Prediction of PSA failure (PSAf) after 5 years of follow-up using a simplified ANNA (model 3 by a cutoff point of 0.5) in each risk group in validation set patient group ($n = 200$). Low risk: PSA 0–4 and biopsy Gleason score 2–6, or PSA 4–10 and biopsy Gleason score 2–4. Intermediate risk: PSA 4–10 and biopsy Gleason score 5–7, or PSA 0–4 and biopsy Gleason score 7, or PSA 10–20 and biopsy Gleason score 2–7. High risk: PSA >20 and/or biopsy Gleason score ≥ 8 . PPV: positive predictive value, NPV: negative predictive value.

from predicted probabilities and 95%CI for the final model. A ROC curve for the predictive model was generated after establishing various cutoff points. Mean AUC was 0.738 ± 0.056 . Fig. 1 shows the actuarial PSA failure-free survival stratified by pMRI stage. The 5-year PSA failure-free survival rate was 82% vs. 38% ($p < 0.0001$) of patients with pMRI T2 vs. pMRI T3 disease, respectively.

Table 4 shows the data with the most weight by the ANNA when predicting the PSA recurrence. Gleason score 8–9, pMRI findings, and PSA >20 ng/ml a relative weight that was 2-fold greater than any other input factor. Using Wilk’s likelihood ratio test, Gleason score, PSA, and pMRI findings were the more significant predictors of ANNA model. Removal of the clinical stage resulted in a simplified model with an equally strong accuracy compared to the ANNA that included all variables (Tables 2 and 4).

The overall accuracy of the pMRI to predict ECE and SVI (Table 5) decreases with increasing risk group (i.e., increasing PSA and biopsy Gleason score) [12]. The overall accuracy in the intermediate-risk group for prediction of ECE and SVI is 85% and 94%, respectively. Moreover, the accuracy of a positive pMRI for ECE to predict for PSA failure is 100% in the intermediate-risk group.

4. Discussion

A simplified ANNA based on the input variables of Kattan nomogram [2] and the pMRI findings was constructed and shows high sensitivity (91%) and specificity (88%) to predict preoperatively PSA failure statistically more accurate than the Kattan nomogram or the classical Cox regression models in clinically localized PCa. The high negative predictive values of

95% made the network useful for initial evaluation before a curative attempt.

Since patients who are classifiable into low or high risk for postoperative PSA failure do not need additional studies to predict what is already known [12], this approach excludes nearly half of the patients as candidates for pMRI. The accuracy in predicting the biochemical recurrence is significantly ($p < 0.05$) higher in the intermediate patient group than in low or high risk groups (92% vs. 75% vs. 88%, respectively) and thus the performance of pMRI in this patient group may be useful tool in the management of localized PCa.

Although our primary objective was not a radiological-pathological correlation, we observed 27% false negative rate and 17% false positive rate detecting clinically occult pT3 disease. Despite the inaccuracy of pMRI to predict the pathological stage [15], the significant contribution of pMRI to predict the PSA outcome may be explained by the assumption that pMRI T3 stage serves as a possible indicator for tumor volume, which was one of the most important factors for disease recurrence [7] (tumors >3 cm in contact with the prostate capsule but not penetrating it were also classified as pMRI T3), rather than strictly identifying pT3 disease.

Our study has several differences compared to those models developed by Kattan [2,3] and Han et al. [5]. The staging based on the 1997 TNM classification instead of the 1992. Although DRE is a simple and inexpensive staging method, it underestimates tumor extent in a majority of patients (ranged from 22 to 63%), is subject to interpretational errors [26]. On the other hand, among imaging studies used in prostate cancer staging, the introduction of MRI has dramatically improved the ability of the technique to correctly define the local tumor staging, with accuracies as high as 82–88% [12–18]. In accordance to other reports [19–21], our study showed that the overall accuracy of pMRI for detecting ECE was 77%, while no significant correlation was found between the DRE and the pathological findings. Our pMRI results were significant predictors of PSA failure, but DRE was not. This confirms strongly the thesis that pMRI provides more valuable information than DRE in PCa clinical staging and adds independent information in the prediction of PSA outcome, which is not already contained in the values of the preoperative PSA and biopsy Gleason score. The clinical stage was not found to add independent information when pMRI information was included in the multivariable analysis.

Another reason why the difference in predictability between ANNA and Kattan nomogram reach statistical

significance may be the relative large number of pre-treatment parameters with various pMRI findings variables, which had a relatively non linear relationship to the pathological findings.

Generally our findings may be confined by the following limitations. The relatively small number of participants with the relatively short follow-up time of 5 years may have restricted the power of our ANNA and Cox regression models. On the other hand, since statistical power was adequate to test a single variable, power limitations unlikely underlie lack of association observed between pMRI findings and PSA outcome. The population study includes patients treated with RP. Our results may have selection bias but are not applicable in other treatment options, such as external beam radiation therapy or cryotherapie. Although PSA failure may be a surrogate end point for death from PCa for patients undergoing surgical treatment, another limitation is that the findings are based on PSA control and not patterns of failure or cause-specific survival data. In biochemical failure, these patients who fail biochemical do not die of their disease or even progress to metastasis; an aggressive therapy would not be recommended in patients, who certainly fail biochemically. All data are from the same institution, and we have no outside validation dataset. Specialists performed all pMRI and Gleason grading in our institution. The applicability of our models in the wider medical community assumes comparable histological-grading or pMRI-findings accuracy by other pathologists or radiologists, respectively. A further limitation may be related to the length of follow-up. Our models predict only to the 5-year point, beyond which recurrence could be possible. However, the vast majority of patients fail within the first 3 years after prostatectomy [7]. As our follow-up matures, our results will need re-evaluation, in order to find any missing factor, which may affect the PSA outcome. Furthermore, ANNA models are not survival curves as used in traditional statistics. There is no censoring of patients, and the models cannot plot the survival curves. Instead, they are using clinically available variables and simply providing probability estimates for recurrence at a specific follow-up time after treatment. However, it is difficult to make direct comparisons between the AUCs of our model and those of Kattan nomogram, as they have been generated from different data sets of diverse patient populations.

Although various cutoff points have been used for determining PSA recurrence after RP with the most accepted cutoff value being >0.2 ng/ml [1–3], we think that patients having PSA between 0.1 and 0.2 ng/ml are at a certain risk of PSA progression [27].

In our patient group PSA value showed a high wide window ranged between 1.1 and 49 ng/ml. Since patients who had PSA >20 ng/ml were 15% of the group studied and non-organ confined were 84% of this group one can assume that in the patient group with PSA <10 ng/ml the results would be different. However, our objective was not to discriminate patients into favorable and unfavorable prognosis groups, but to evaluate the preoperative model for PSA recurrence in the whole patient group having clinically localized PCa and who underwent RP.

Although the group of patients having Gleason sum 2-5 in preoperative biopsy cores seems to be relatively large (31%, $n = 72$), this phenomenon is not so rare in the literature [28,29], since the original patient group, in which the Kattan nomogram for the prediction of pathological stage has been evaluated, showed similar characteristics [29]. Furthermore, these are the preoperative characteristics of our patient group. Calculating the under-graded rate of 46% ($n = 88$), which mainly affected the group of patients having Gleason sum 2-5 (80%, $n = 57$ patients out of 72 with Gleason score 2–5), and the over-graded rate of 10% the Gleason sum of 2-5 in prostatectomy specimen become lower (21%, $n = 40$) similar to the patient collective of other large studies [3,30].

Although TRUS is more accurate than DRE for preoperative staging, MRI is still in our study and in other reports [31,32] superior to TRUS in detecting ECE or SVI. This reflects the finding that ANNA included only the TRUS-Stage is significantly less accurate than those incorporated the pMRI findings and comparable to those having as input variable the clinical stage (DRE findings) only.

Regarding the utility of the study, confirmatory studies will be necessary. Before pMRI can be used routinely, there must not be a cheaper available alter-

native, which when tested on a comparative multi-variable analysis and an ANNA is shown to provide the same prognostic information as pMRI.

5. Conclusions

An ANNA incorporating the pMRI-findings was significantly more accurate than the classical regression with the same input variables and the Kattan nomogram in predicting biochemical outcome after RP in our area. Biopsy Gleason score and pMRI findings were the most valuable predictor parameters. Additional prospective investigations in larger patient groups with longer follow-up with sophisticated body and endorectal MRI [19–21] will determine its ultimate role in the staging workup and predicting the PSA recurrence. In the group of patients with intermediate risk for postoperative PSA failure [12], the use of pMRI seems to enhance the pre-treatment prediction of biochemical outcome. Our preoperative model, with an AUC of 0.89, may be useful in decision-making, assisting urologists and patients in deciding whether RP is an acceptable treatment option and in identifying men at high risk of PSA recurrence who may benefit from neoadjuvant or immediate adjuvant treatment protocols.

Acknowledgements

The authors are indebted to the software company StatSoft GmbH, Hamburg, Germany, for constructing the artificial neural network and assisting in the statistical data analysis before the initial paper submission and after then during the final correction and resubmission process.

References

- [1] Han M, Partin AW. Nomograms for clinically localized prostate cancer. Part I: radical prostatectomy. *Semin Urol Oncol* 2002;20:123–30.
- [2] Kattan MW, Eastham JA, Stapleton AM, Wheeler TM, Scardino PT. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst* 1998;90:766–71.
- [3] Graefen M, Karakiewicz PI, Cagiannos I, Quinn DI, Henshall SM, Grygiel JJ, et al. International validation of a preoperative nomogram for prostate cancer recurrence after radical prostatectomy. *J Clin Oncol* 2002;20:3206–12.
- [4] D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Fondurulia J, Chen MH, et al. Clinical utility of the percentage of positive prostate biopsies in defining biochemical outcome after radical prostatectomy for patients with clinically localized prostate cancer. *J Clin Oncol* 2000;18:1164–72.
- [5] Han M, Partin AW, Zahurak M, Piantadosi S, Epstein JI, Walsh PC. Biochemical (prostate specific antigen) recurrence probability following radical prostatectomy for clinically localized prostate cancer. *J Urol* 2003;169:517–23.
- [6] Stamey TA. Preoperative serum prostate-specific antigen (PSA) below 10 $\mu\text{g/l}$ predicts neither the presence of prostate cancer nor the rate of postoperative PSA failure. *Clin Chem* 2001;47:631–4.
- [7] Stamey TA, McNeal JE, Yemoto CM, Sigal BM, Johnstone IM. Biological determinants of cancer progression in men with prostate cancer. *JAMA* 1999;281:1395–400.
- [8] Hammerer P, Huland H. Systematic sextant biopsies in 651 patients referred for prostate evaluation. *J Urol* 1994;151:99–102.

- [9] Sobin LH, Wittekind C, editors. *TNM Classification of Malignant Tumours*. 5th ed. New York: John Wiley & Sons; 1997. 170 p.
- [10] Humphrey PA. Complete histologic serial sectioning of a prostate gland with adenocarcinoma. *Am J Surg Pathol* 1993;17:468–72.
- [11] Walsh PC. Radical retropubic prostatectomy. In: Walsh PC, Retik AB, Stamey TA, Vaughan ED, editors. *Campbell's Urology*. 6th Edition. Vol. 3. Philadelphia: Saunders; 1992. p. 2865–86 [Chapter 78].
- [12] D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Schnall M, Tomaszewski JE, et al. Critical analysis of the ability of the endorectal coil magnetic resonance imaging scan to predict pathologic stage, margin status, and postoperative prostate-specific antigen failure in patients with clinically organ-confined prostate cancer. *J Clin Oncol* 1996;14:1770–7.
- [13] Siegelman ES, Schnall MD. Contrast-enhanced MR imaging of the bladder and prostate. *Magn Reson Imaging Clin N Am* 1996;4:153–69.
- [14] Hricak H, White S, Vigneron D, Kurhanewicz J, Kosco A, Levin D, et al. Carcinoma of the prostate gland: MR imaging with pelvic phased-array coils versus integrated endorectal–pelvic phased-array coils. *Radiology* 1994;193:703–9.
- [15] Wong-You-Cheong JJ, Krebs TL. MR imaging of prostate cancer. *Magn Reson Imaging Clin N Am* 2000;8:869–86.
- [16] Outwater EK, Petersen RO, Siegelman ES, Gomella LG, Chernesky CE, Mitchell DG. Prostate carcinoma: assessment of diagnostic criteria for capsular penetration on endorectal coil MR images. *Radiology* 1994;193:333–9.
- [17] Yu KK, Hricak H, Alagappan R, Chernoff DM, Bacchetti P, Zaloudek CJ. Detection of extracapsular extension of prostate carcinoma with endorectal and phased-array coil MR imaging: multivariate feature analysis. *Radiology* 1997;202:697–702.
- [18] Husband JE, Padhani AR, MacVicar AD, Revell P. Magnetic resonance imaging of prostate cancer: comparison of image quality using endorectal and pelvic phased array coils. *Clin Radiol* 1998;53:673–81.
- [19] Aziza R, Soulie M, Escourrou G, Bachaud J, Molinie L, Tollon C, et al. Local staging of prostate carcinoma with phased array MR imaging: prospective study over 5 years. *J Radiol* 2002;83:39–44.
- [20] Rorvik J, Halvorsen OJ, Albrektsen G, Erslund L, Daehlin L, Haukaas S. MRI with an endorectal coil for staging of clinically localised prostate cancer prior to radical prostatectomy. *Eur Radiol* 1999;9:29–34.
- [21] Ogura K, Maekawa S, Okubo K, Aoki Y, Okada T, Oda K, et al. Dynamic endorectal magnetic resonance imaging for local staging and detection of neurovascular bundle involvement of prostate cancer: correlation with histopathologic results. *Urology* 2001;57:721–6.
- [22] Porter CR, O'Donnell C, Crawford ED, Gamito EJ, Sentizimary B, De Rosalia A, et al. Predicting the outcome of prostate biopsy in a racially diverse population: a prospective study. *Urology* 2002;60:831–5.
- [23] Golden RM. *Mathematical methods for neural network analysis and design*. Cambridge (MA): MIT Press; 1996.
- [24] Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
- [25] Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;148:839–43.
- [26] Perrotti M, Pantuck A, Rabbani F, Israeli RS, Weiss RE. Review of staging modalities in clinically localized prostate cancer. *Urology* 1999;54:208–14.
- [27] Freedland SJ, Sutter ME, Dorey F, Aronson WJ. Defining the ideal cutpoint for determining PSA recurrence after radical prostatectomy. *Prostate-specific antigen Urology* 2003;61:365–9.
- [28] Tewari A, Issa M, El-Galley R, Stricker H, Peabody J, Pow-Sang J, et al. Genetic adaptive neural network to predict biochemical failure after radical prostatectomy: a multi-institutional study. *Mol Urol* 2001;5:163–9.
- [29] Kattan MW, Stapleton AM, Wheeler TM, Scardino PT. Evaluation of a nomogram used to predict the pathologic stage of clinically localized prostate carcinoma. *Cancer* 1997;79:528–37.
- [30] Palisaan RJ, Graefen M, Karakiewicz PI, Hammerer PG, Huland E, Haese A, et al. Assessment of clinical and pathologic characteristics predisposing to disease recurrence following radical prostatectomy in men with pathologically organ-confined prostate cancer. *Eur Urol* 2002;41:155–61.
- [31] Bates TS, Gillatt DA, Cavanagh PM, Speakman M. A comparison of endorectal magnetic resonance imaging and transrectal ultrasonography in the local staging of prostate cancer with histopathological correlation. *Br J Urol* 1997;79:927–32.
- [32] Sanchez-Chapado M, Angulo JC, Ibarburen C, Aguado F, Ruiz A, Viano J, et al. Comparison of digital rectal examination, transrectal ultrasonography, and multicoil magnetic resonance imaging for preoperative evaluation of prostate cancer. *Eur Urol* 1997;32:140–9.