

Head-to-head comparison of prostate cancer risk calculators predicting biopsy outcome

Nuno Pereira-Azevedo^{1,2*}, Jan F. M. Verbeek^{1*}, Daan Nieboer^{1,3}, Chris H. Bangma¹, Monique J. Roobol¹

¹Department of Urology, Erasmus University Medical Center, Rotterdam, The Netherlands; ²Department of Urology, Centro Hospitalar do Porto, Porto, Portugal; ³Department of Public Health, Erasmus University Medical Center, Rotterdam, The Netherlands

Contributions: (I) Concept and design: N Pereira-Azevedo, JF Verbeek, MJ Roobol; (II) Administrative support: None; (III) Provision of study material or patients: N Pereira-Azevedo; (IV) Collection and assembly of data: N Pereira-Azevedo; (V) Data analysis and interpretation: JF Verbeek, D Nieboer, N Pereira-Azevedo, MJ Roobol; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*These authors contributed equally to this work.

Correspondence to: Dr. Nuno Pereira-Azevedo, MD. Department of Urology, Room NA-1706, P.O. Box 2040, Erasmus MC, University Medical Center Rotterdam, Rotterdam, 3000 CA Rotterdam, The Netherlands. Email: nuno@pereira-azevedo.com.

Background: Multivariable risk calculators (RCs) predicting prostate cancer (PCa) aim to reduce unnecessary workup (e.g., MRI and biopsy) by selectively identifying those men at risk for PCa or clinically significant PCa (csPCa) (Gleason ≥ 7). The lack of an adequate comparison makes choosing between RCs difficult for patients, clinicians and guideline developers. We aim to perform a head-to-head comparison of seven well known RCs predicting biopsy outcome.

Methods: Our study comprised 7,199 men from ten independent contemporary cohorts in Europe and Australia, who underwent prostate biopsy between 2007 and 2015. We evaluated the performance of the ERSPC RPCRC, Finne, Chun, ProstataClass, Karakiewicz, Sunnybrook, and PCPT 2.0 (HG) RCs in predicting the presence of any PCa and csPCa. Performance was assessed by discrimination, calibration and net benefit analyses.

Results: A total of 3,458 (48%) PCa were detected; 1,784 (25%) men had csPCa. No particular RC stood out predicting any PCa: pooled area under the ROC-curve (AUC) ranged between 0.64 and 0.72. The ERSPC RPCRC had the highest pooled AUC 0.77 (95% CI: 0.73–0.80) when predicting csPCa. Decision curve analysis (DCA) showed limited net benefit in the detection of csPCa, but that can be improved by a simple calibration step. The main limitation is the retrospective design of the study.

Conclusions: No particular RC stands out when predicting biopsy outcome on the presence of any PCa. The ERSPC RPCRC is superior in identifying those men at risk for csPCa. Net benefit analyses show that a multivariate approach before further workup is advisable.

Keywords: Prostate biopsy; prostate cancer (PCa); prostate-specific antigen (PSA); risk calculators (RCs); overdiagnosis

Submitted Dec 22, 2017. Accepted for publication Dec 22, 2017.

doi: 10.21037/tau.2017.12.21

View this article at: <http://dx.doi.org/10.21037/tau.2017.12.21>

Introduction

Prostate cancer (PCa) is the most prevalent cancer in the western world for men and the second most common cause of death in men worldwide (1). Long-term follow-up from the European Randomised Study of Screening for Prostate

Cancer (ERSPC) has shown a significant reduction in PCa specific mortality applying prostate-specific antigen (PSA) based screening (2). This, next to the fact that PSA is a well-developed, easy to implement, and cheap test, made PSA testing the mainstay in the decision for further

Table 1 Overview of the included predictors per RC to discriminate men at risk for any PCa and csPCa

Risk calculator	PSA	Age	PV	DRE	Free PSA	Fam.	Race	Prev Bx	IPSS	SD
Risk calculator, any PCa										
ERSPC RPCRC	+		+	+				+		
Finne	+		+	+	+					
Chun	+	+		+	+					+
Karakiewicz	+	+		+	+					
PCPT 1.0	+			+		+		+		
PCPT 2.0	+	+		+		+	+	+		
PCPT 2.0 + freePSA	+	+		+	+	+	+	+		
ProstataClass	+	+	+	+	+					
Sunnybrook	+	+		+	+	+	+		+	
Risk calculator, clinical significant PCa										
ERSPC RPCRC	+		+	+				+		
PCPT HG	+	+		+		+	+	+		
PCPT 2.0	+	+		+	+	+	+	+		
Sunnybrook	+	+		+	+	+	+		+	

Age in years. *, Analysed PCPT 2.0 models with free PSA and without free PSA. csPCa, clinically significant prostate cancer; DRE, digital rectal examination suspicious or not; Fam, family history or not of prostate cancer; free PSA, free proportion of total PSA; IPSS, International Prostate Symptom Score (0–35); Prev bx, previous negative biopsy; PSA, prostate-specific antigen, in ng/mL; PV, prostate volume, in mL; SD, sample density (prostate volume/number of biopsy cores taken).

clinical workup (i.e., prostate biopsy). However, any choice of a PSA cut-off involves a trade-off between sensitivity and specificity. Lowering the PSA cut-off would improve test sensitivity, but also reduce specificity, leading to far more false-positive tests and unnecessary interventions. Additionally, many of the cancers detected may never become clinically evident, thereby leading to overdiagnosis and overtreatment (3,4).

Other relevant pre-biopsy clinical information, next to the serum PSA level, has been incorporated into so-called risk calculators (RCs) to enable a more accurate assessment of a patient’s individual PCa risk (5). A recent systematic review identified 127 unique RCs in the field of PCa. The conclusion was that RCs outperform PSA alone in avoiding unnecessary biopsies, that not all of the RCs have the ability to selectively identify those men at risk of having clinically significant PCa (csPCa) (defined as Gleason score ≥3+4) and that external validation studies and head-to-head comparisons are lacking (6). RCs are part of the European Association of Urology (EAU), European Society for Radiotherapy & Oncology (ESTRO), European Society of

Urogenital Radiology (ESUR), and International Society of Geriatric Oncology (SIOG) joint guidelines on PCa screening and early detection (5) but it remains a matter of personal choice whether and/or which RC to use in one’s daily clinical practice.

We aimed to address this lack of information by evaluating the performance (discrimination, calibration, and clinical impact) of the most well-known RCs developed to predict prostate biopsy outcome in a head-to-head comparison.

Methods

Participants

Our study cohort comprises of 8,649 men from ten independent contemporary cohorts (nine in Europe and one in Australia) who underwent a transrectal ultrasound (TRUS)-guided prostate biopsy between January 2007 and November 2015 (Table 1 and Acknowledgements). The cohorts were unrelated to the RCs development. Pre-

biopsy clinical data and the pathological biopsy results were obtained from electronic and paper medical charts. Data on age, previous negative biopsy, DRE (digital rectal examination; benign/suspicious), prostate volume (assessed by TRUS in mL), total PSA (tPSA), free PSA (fPSA), and biopsy outcome (PCa detected yes/no and Gleason grade) were collected for all men. Patients were included into the analyses if they met the aggregated criteria of all RCs: age between 50 and 89 years old, PSA <50 ng/mL, prostate volume between 10 and 110 mL (*Table S1*). Participants in the study underwent a TRUS-guided biopsy according to the standard clinical practice used at each participating site, which was on average 12 cores [interquartile range (IQR): 12–14] per biopsy session in the European cohorts and 30 cores (IQR: 12–30) in the Australian cohort.

Statistical analysis

Baseline characteristics of the study cohort are presented as median and interquartile range (M-IQR) or percentage for proportion type features. Missing data was imputed using Multivariate Imputation by Chained Equations (MICE) using only the values from the corresponding cohort and was imputed five times (7). Three parameters were completely missing in certain cohorts: fPSA (Den Bosch, Sydney, Breda), International Prostate Symptom Score (Den Bosch, Porto, Bordeaux, Munster, Paris, Hamburg, Rennes, Milan), and family history of PCa (Den Bosch, Porto, Bordeaux, Munster, Paris, Rennes). These missing values were imputed by using the complete information from the other cohorts in the imputation model. This strategy leads to more precise estimates of the predicted probabilities (8). After imputation, the probabilities were calculated on the basis of the predictions rules from each RC. Individual patient data was compared for men with csPCa and men without PCa using the Mann-Whitney test for continuous variables.

Risk calculators

For this meta-analysis we included seven well known RCs: Chun (9), ERSPC Rotterdam Prostate Cancer Risk Calculator RPCRC (10), Finne (11), Karakiewicz (12), ProstataClass (13), Prostate Cancer Prevention Trial (PCPT) 1.0 and 2.0 +/- free-PSA (14), and Sunnybrook (15). The first six have been externally validated in over five studies (6). All RCs use PSA and DRE as a predictive

factor, and the other predictors are displayed in *Table 1*. The RPCRC for initial and repeat biopsy (16) was adapted to the contemporary Rotterdam clinical setting (17). For each individual patient, the probabilities of having a biopsy detectable PCa and, if applicable, a csPCa, were calculated. The probabilities of the ProstataClass artificial neural network were obtained by sending a blinded database for PCa outcome to the ProstataClass developers (H. Cammann) (13). For uniform comparison between RCs men with a PSA above their advised cut-off for PSA (i.e., >20 ng/mL) were included (157 men (2.2% of the total cohort). Four models were able to separately predict csPCa (RPCRC, PCPT 2.0 +/- freePSA and Sunnybrook).

Comparison of risk calculator models

The predictive accuracy (predicting PCa and csPCa) was quantified using the area under the curve (AUC) for the receiver operator characteristic (ROC) analysis (18). A multivariable meta-analysis was performed to pool the AUCs in predicting any PCa and csPCa. Within-study correlations of the AUCs were estimated using bootstrapping. To analyze statistically significant differences in the models and taking into account the between-study heterogeneity we subsequently estimated the probability that a model has the highest AUC in a subsequent validation study. We simulated 10,000 samples from the posterior distribution to estimate this probability (19). Calibration of the RCs was pooled and explored graphically using calibrations plots. For comparison, the per center sensitivity and specificity of detecting csPCa with applying a PSA cut-off ≥ 4.0 ng/mL was calculated and graphically displayed.

In addition, the clinical impact was assessed with decision curve analysis (DCA) and clinical impact curves. DCA represents the net benefit ratio, which weighs the benefits (detecting cancer) versus the harms (unnecessary biopsy) over a range of thresholds (20). Clinical impact curves show the estimated number who would be declared eligible for biopsy for each risk threshold and show the proportion of those who are cases (overall PCa and csPCa) (21). In addition, to show the potential of multivariate risk stratification when adapting to, for example, one's own hospital data, we calculated net benefit after calibration of each of the 4 models predicting csPCa using the largest clinical cohort in our series (Den Bosch; N=2,053). Analyses were performed using R statistical package, version 3.3.1, R Foundation for Statistical Computing, Vienna, Austria.

Table 2 Baseline characteristics of the independent cohorts

Cohort, city	N of cases (%)	Age, M-IQR	PSA, M-IQR	Abnormal DRE, N (%)	Previous Bx, N (%)	Any PCa, N (%)	csPCa, N (%)
Porto*, PT	568 (7.9)	66 [61–70]	6.6 (4.8–8.9)	243 [43]	73 [13]	235 [41]	164 [29]
Bordeaux, FR	1,568 (21.8)	65 [61–69]	7.0 (5.4–9.5)	616 [39]	202 [13]	845 [54]	370 [24]
Paris, FR	116 (1.6)	65 [60–71]	4.4 (3.4–5.7)	38 [33]	16 [13]	66 [59]	25 [24]
Rennes, FR	218 (3.0)	62 [58–68]	4.5 (3.4–5.8)	135 [62]	34 [16]	128 [59]	41 [20]
Milan, IT	715 (9.9)	65 [59–70]	6.1 (4.4–8.6)	128 [18]	246 [34]	281 [38]	135 [18]
Hamburg, DU	270 (3.8)	68 [62–71]	6.6 (4.6–9.5)	74 [27]	101 [37]	143 [52]	86 [31]
Munster, DU	513 (7.1)	63 [59–69]	5.2 (4.1–6.4)	59 [12]	211 [41]	253 [49]	213 [42]
Den Bosch, NL	2,053 (28.5)	64 [60–69]	7.9 (6.1–11.0)	566 [28]	428 [21]	847 [41]	368 [18]
Breda, NL	644 (9.2)	66 [62–71]	9.4 (6.9–13.5)	219 [32]	157 [24]	298 [47]	124 [20]
Sydney, AU	534 (7.4)	65 [58–69]	5.7 (4.1–8.3)	206 [39]	28 [5]	362 [68]	258 [48]
Total	7,199	65 [60–69]	6.9 (5.1–9.6)	2,284 [32]	1,496 [21]	3,458 [48]	1,784 [25]

*, given their similar protocol, the 65 patients from Évora were included in the Porto cohort. Bx, biopsy; csPCa, clinically significant prostate cancer; DRE, digital rectal examination; M-IQR, median and interquartile range.

Table 3 Clinical characteristics of men with csPCa compared with men without PCa and lrPCa

Variable	No PCa, 3,741 (52%)	lrPCa, 1,674 (23%)	csPCa, 1,784 (25%)
Age at biopsy (years)	64 [59–68]	65 [60–69]	67 [62–72]
Total PSA (ng/mL)	6.7 [5.0–9.2]	6.6 [5.0–9.0]	7.4 [5.4–11.6]
Free PSA rati	0.16 [0.12–0.21]	0.14 [0.10–0.18]	0.12 [0.08–0.16]
Abnormal DRE	673 [18]	559 [33]	1,036 [58]
PV (mL)	50 [38–66]	41 [31–55]	39 [30–50]
Positive family history	594 [16]	319 [19]	347 [20]
IPSS/AUA	3 [1–9]	2 [1–10]	3 [1–8]
Previous biopsy	977 [26]	316 [19]	198 [11]
Number of cores	12 [12–15]	12 [12–14]	12 [12–16]

Data are shown as median [IQR] or number [%]. csPCa, clinically significant prostate cancer; IPSS, International Prostate Symptom Score; IQR, interquartile range; lrPCa, low risk prostate cancer; PV, prostate volume.

Results

Of the total of 8,649 men 7,199 men (83.2%) were included in analyses. Median age was 65 years old, median PSA 6.9 ng/mL, median prostate volume 45 mL (as evaluated by TRUS) and 1,496 men (21%) underwent a previous biopsy (Table 2). PCa was diagnosed in 3,458 of 7,199 patients (48%) and 1,784 (25%) men had csPCa. PCa patients were older (median age: 65 vs. 64 years for non PCa patients, $P < 0.001$), had smaller prostate glands (41 vs. 50 mL, $P < 0.001$), and higher PSA (6.7 vs. 7.0 ng/mL, $P < 0.001$) (Table 3).

In predicting any PCa no particular RC stood out and the pooled area under the ROC-curve (AUC) ranged between 0.64 and 0.72 (Figure 1) with Finne having the highest AUC. Substantial heterogeneity in the AUC was found between the different cohorts (range I^2 , 66–89%). In predicting csPCa the ERSPC RPCRC had the highest pooled AUC of 0.77 (95% CI: 0.73–0.80; Figure 1). After repeating this comparative analysis 10,000 times the ERSPC RPCRC had the highest probability (89%) of having the highest AUC. The probabilities of having the highest AUC in our study cohort were 6%, 3%, and 2% for PCPT 2.0 + freePSA, PCPT 2.0, and Sunnybrook RCs respectively.

The calibration plots for those RCs predicting csPCa for the pan-European data set (including all cohorts except Sydney, Australia; $n = 6,665$) are displayed in Figure 2. Three

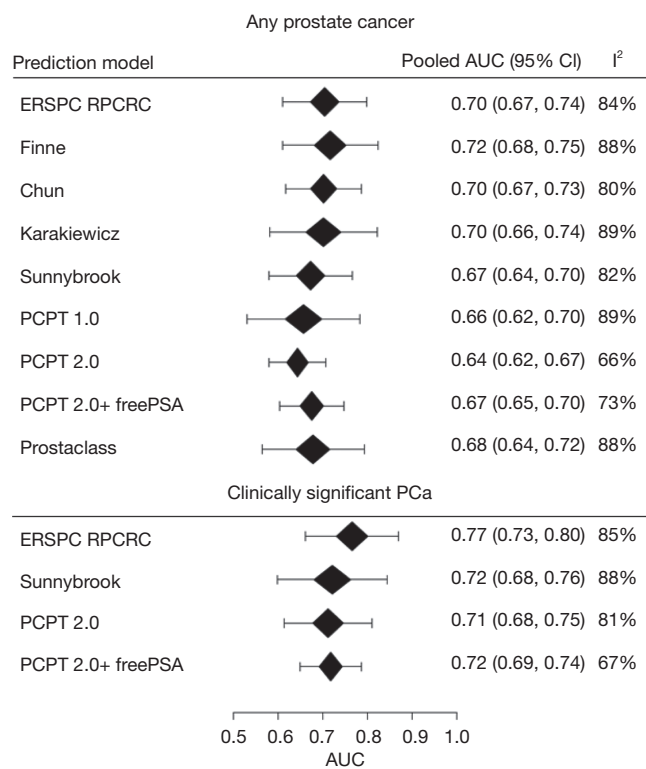


Figure 1 Meta-analysis of the pooled area under the curve (AUC) and 95% CI of various PCa risk calculators to discriminate men diagnosed with any prostate cancer (above) and clinically significant PCa (below). CI, confidence interval; PCa, prostate cancer.

models underestimated the probability of csPCa, while the ERSPC RPCRC was more accurate at low probabilities and mainly overestimated at probabilities >10%. *Figure S1* displays the decision curves and clinical impact plots per cohort applying the 4 RCs predicting csPCa. Overall the ERSPC RPCRC has the highest net benefit followed by the PCPT 2.0. Clinical impact is negligible to small starting at probabilities >10% for detecting csPCa. *Figure S2* displays the differences in sensitivity and specificity of the PSA test (cut-off ≥ 4.0 ng/mL) per center. Sensitivity and specificity range from 76% to 98% and 4% to 44% respectively. *Table 4* shows net benefit when each of the 4 RCs are calibrated. At a 4% threshold for csPCa using the ERSPC RPCRC the number of biopsies can be reduced by 32% while keeping a 95% sensitivity for detecting csPCa. Reduction and sensitivity are 8% and 99% for PCPT2.0 + freePSA, 16% and 97% for PCPT 2.0, and 25% and 95% for Sunnybrook.

Discussion

In this head-to-head comparison of seven well-known RCs predicting prostate biopsy outcome it is shown that all RCs have a moderate to well discriminatory ability when predicting any PCa (AUCs ranging from 0.64 to 0.72). Those RCs that can selectively predict csPCa show AUCs in the range of 0.71 to 0.77 with small clinical benefit in this pan-European cohort of contemporary daily clinical practice and clinical study data. Adjusting calibration shows the added value of incorporating multivariable risk prediction tools next to clinical expertise in clinical decision making. These results confirm earlier analyses on the use of multivariable prediction tools (6) and considering the substantial harm related to overdiagnosis of low risk PCa (3,22), the use of those RCs that selectively can predict csPCa is recommended.

The balance between benefit and harm of early detection of PCa is still a topic of ongoing debate. Due to the initiation of PSA screening in combination with TRUS guided systematic prostate biopsy the incidence of predominantly low risk PCa increased enormously in the 1990s. This eventually resulted in guidelines recommending no screening at all (23). However, with the available data from longitudinal studies and randomized PCa screening trials we currently have, the knowledge to improve the balance between harm and benefit recommendations have changed to using shared decision making with an individual approach towards how to screen best (24).

While both discrimination and calibration are important statistics to evaluate performance of a prediction tool we must note that discrimination cannot be easily improved while calibration can (6). An example of potentially adjusting calibration to a particular setting is shown in (25) where a model was first tested using part of the available data (calibration phase) where subsequently performance was assessed in the rest of the data (validation cohort). Based on the wide calibration prediction intervals in the current analyses it is advisable to follow such an approach where the aim should be to assess moderate calibration on the basis of center specific retrospective data on prostate biopsy outcome with a minimum of 200 prostate cancers cases (26). Subsequently these center specific adjustments for the calculated probabilities could be incorporated, for example, in the RPCRC. It is in this context important to realize that when using a purely PSA based approach considerable variation in sensitivity and specificity also

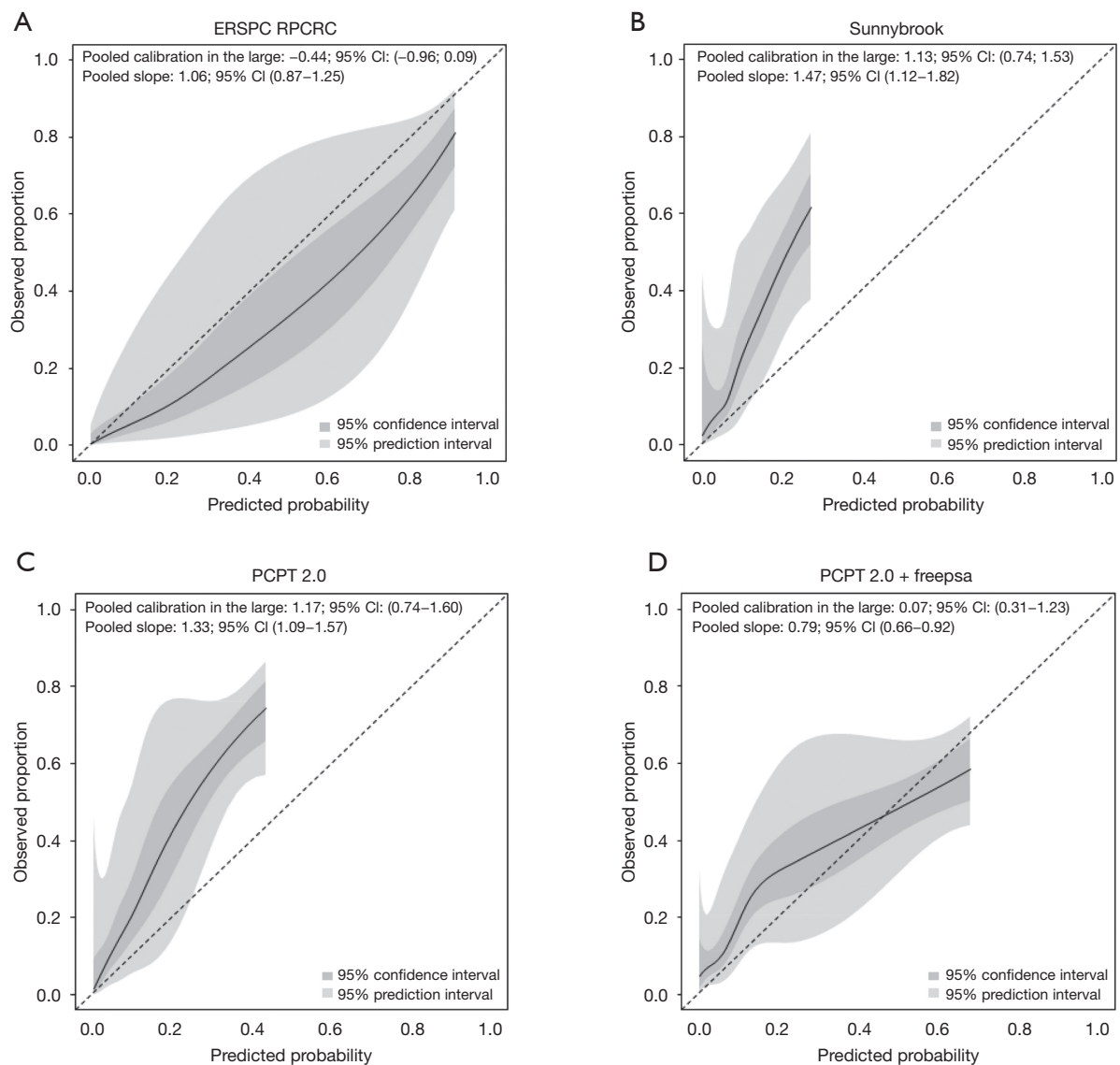


Figure 2 Pooled calibration plots for clinical significant prostate cancer using ERSPC RPCRC (A), Sunnybrook (B), PCPT 2.0 (C) and PCPT 2.0 + freePSA (D) risk calculators in the nine European cohorts. CI, confidence interval; free PSA, free proportion of total prostate-specific antigen.

exists, as was shown in *Figure S2*, something that is ignored in recommendations on applying a cut-off value to trigger prostate biopsy.

When predicting biopsy outcome, it must also be noted that especially the use of the multiparametric magnetic resonance imaging (mpMRI) in the detection of PCa and csPCa has increased considerably showing very promising results. However, while the mpMRI is advised to be used

after a negative TRUS guided systematic prostate biopsy (often solely based on an elevated PSA level), the mpMRI is more and more used before the first biopsy (27). Previous analyses with the RPCRC have however shown that upfront risk stratification on the basis of easy (and cheap) to get relevant pre-biopsy information can avoid half of mpMRIs (17).

This study has some limitations. First, it is a retrospective study design using ten different cohorts from

Table 4 Net benefit, biopsy reduction of and missed prostate cancers with the use of four risk calculators predicting clinical significant prostate cancer at a 4% and 10% risk threshold in the Den Bosch cohort

Risk calculator	Biopsy reduction (% of total biopsy)	Missed prostate cancers		Net benefit
		Indolent (% of total indolent PCa)	csPCa (% of total csPCa)	
Threshold 4%				
ERSPC RPCRC	665 (32.4)	118 (24.6)	17 (4.6)	0.150
PCPT 2.0 + freePSA	162 (7.9)	32 (6.7)	3 (0.8)	0.147
PCPT 2.0	319 (15.5)	73 (15.2)	12 (3.3)	0.145
Sunnybrook	516 (25.1)	105 (21.9)	20 (5.4)	0.145
Threshold 10%				
ERSPC RPCRC	1,202 (58.5)	247 (51.6)	60 (16.3)	0.127
PCPT 2.0 + freePSA	890 (43.4)	216 (45.1)	61 (16.6)	0.103
PCPT 2.0	957 (46.7)	231 (48.2)	52 (14.1)	0.112
Sunnybrook	1,164 (56.7)	277 (57.8)	58 (15.8)	0.120

csPCa, clinically significant prostate cancer; PCa, prostate cancer.

populations with different background risk and different referral patterns (daily clinical practice cohorts with the risk of selective outcome reporting and clinical study cohorts with predefined eligibility criteria) as reflected in the heterogeneity of our results. On the other hand, evaluating performance of these RCs in this pan-European setting can be seen as a strength and support their use in Europe. Second, we mainly used the original RCs which were virtually all developed in the 1990s. This implies that they do not use later developed biomarkers (e.g., PHI, PCA3, the 4K panel). All these biomarkers have shown to have additional predictive value when incorporated into a prediction model and as such might be able to positively influence results (28-30).

Finally, all RCs use as endpoint csPCa based on the original Gleason grading (31). It has been shown that the new Gleason grading system better reflects disease burden (32) as does the inclusion of cribriform growth patterns in the classification of Gleason 7 PCa (16,33).

In conclusion, we performed the first head-to-head comparison of RCs predicting prostate biopsy outcome using a multicenter European and Australian population. No particular RC stood out in the discrimination of men with and without PCa. The ERSPC RPCRC showed highest discrimination when predicting clinically significant PCa. Net benefit in the available clinical cohorts was limited but can be increased by applying a simple calibration step. These outcomes support implementing next to clinical

expertise a multivariable risk prediction tool before further workup (e.g., MRI and biopsy) in men suspicious for having a clinically significant PCa.

Acknowledgements

We would like to thank the investigators, researchers and hospitals for providing the required data (listed according to the number of men they enrolled): H. Beerlage, Jeroen Bosch Hospital, 's-Hertogenbosch, NL; R. Gaston, T. Piechaud, Clinique St. Augustin, Bordeaux, FR; M. Lazzeri, San Raffaele Hospital-Turro, Milan, IT; S. Roemeling, University Medical Center Groningen: Amphia Hospital Breda, NL; D van der Schoot, Amphia Hospital, Breda, NL; I. Braga, L. Osório, V. Cavadas, A. Fraga, Porto Hospital Center, Porto, PT. E. Carrasquinho, E. Cardoso de Oliveira, Hospital Espírito Santo, Évora, PT; P. Stricker, J. Thompson, P. van Leeuwen, Garvan Institute of Medical Research, University of New South Wales, Sydney, AUS; A. Semjonow, University Hospital Munster, Munster, Germany. C. Stephan, Charite-Universitaetsmedizin, Berlin and Berlin Institute for Urologic Research, Berlin, GE; A. Haese and M. Graefen, Prostate Cancer Center, Martini Clinic, University Hamburg-Eppendorf, Hamburg, GE; S. Vincendeau, Hospital Pontchaillou, Rennes, France. A. Houlgatte, HIA Du Val De Grace, Paris, France; Special thanks to H. Cammann, Universitätsmedizin Berlin, Berlin, DE for calculating the predicted probabilities for

ProstataClass ANN.

Footnote

Conflict of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The study was approved by institutional ethics committees and informed/oral consent was taken from all the patients.

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136:E359-86.
2. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet* 2014;384:2027-35.
3. Heijnsdijk EA, Wever EM, Auvinen A, et al. Quality-of-life effects of prostate-specific antigen screening. *N Engl J Med* 2012;367:595-605.
4. Loeb S, Vellekoop A, Ahmed HU, et al. Systematic review of complications of prostate biopsy. *Eur Urol* 2013;64:876-92.
5. Mottet N, Bellmunt J, Bolla M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur Urol* 2017;71:618-29.
6. Louie KS, Seigneurin A, Cathcart P, et al. Do prostate cancer risk models improve the predictive accuracy of PSA screening? A meta-analysis. *Ann Oncol* 2015;26:848-64.
7. van Buuren SG. Multivariate Imputation by Chained Equations in R. *J Stat Softw* 2011;45.
8. Nieboer D, Vergouwe Y, Ankerst DP, et al. Improving prediction models with new markers: a comparison of updating strategies. *BMC Med Res Methodol* 2016;16:128.
9. Chun FK, Steuber T, Erbersdobler A, et al. Development and internal validation of a nomogram predicting the probability of prostate cancer Gleason sum upgrading between biopsy and radical prostatectomy pathology. *Eur Urol* 2006;49:820-6.
10. Roobol MJ, Steyerberg EW, Kranse R, et al. A risk-based strategy improves prostate-specific antigen-driven detection of prostate cancer. *Eur Urol* 2010;57:79-85.
11. Finne P, Finne R, Bangma C, et al. Algorithms based on prostate-specific antigen (PSA), free PSA, digital rectal examination and prostate volume reduce false-positive PSA results in prostate cancer screening. *Int J Cancer* 2004;111:310-5.
12. Karakiewicz PI, Benayoun S, Kattan MW, et al. Development and validation of a nomogram predicting the outcome of prostate biopsy based on patient age, digital rectal examination and serum prostate specific antigen. *J Urol* 2005;173:1930-4.
13. Stephan C, Cammann H, Semjonow A, et al. Multicenter evaluation of an artificial neural network to increase the prostate cancer detection rate and reduce unnecessary biopsies. *Clin Chem* 2002;48:1279-87.
14. Ankerst DP, Hoefler J, Bock S, et al. Prostate Cancer Prevention Trial risk calculator 2.0 for the prediction of low- vs high-grade prostate cancer. *Urology* 2014;83:1362-7.
15. Nam RK, Toi A, Klotz LH, et al. Assessing individual risk for prostate cancer. *J Clin Oncol* 2007;25:3582-8.
16. Roobol MJ, Verbeek JF, van der Kwast T, et al. Improving the rotterdam european randomized study of screening for prostate cancer risk calculator for initial prostate biopsy by incorporating the 2014 International Society of Urological Pathology Gleason Grading and Cribriform growth. *Eur Urol* 2017;72:45-51.
17. Alberts AR, Schoots IG, Bokhorst LP, et al. Risk-based patient selection for magnetic resonance imaging-targeted prostate biopsy after negative transrectal ultrasound-guided random biopsy avoids unnecessary magnetic resonance imaging scans. *Eur Urol* 2016;69:1129-34.
18. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128-38.
19. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002;21:589-624.
20. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016;352:i6.
21. Kerr KF, Brown MD, Zhu K, et al. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. *J Clin Oncol* 2016;34:2534-40.
22. Carlsson SV, de Carvalho TM, Roobol MJ, et al. Estimating the harms and benefits of prostate cancer screening as used in common practice versus recommended

- good practice: A microsimulation screening analysis. *Cancer* 2016;122:3386-93.
23. Moyer VA, Force USPST. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2012;157:120-34.
 24. Bibbins-Domingo K, Grossman DC, Curry SJ. The US Preventive Services Task Force 2017 draft recommendation statement on screening for prostate cancer: an invitation to review and comment. *JAMA* 2017;317:1949-50.
 25. Parekh DJ, Punnen S, Sjoberg DD, et al. A multi-institutional prospective trial in the usa confirms that the 4kscore accurately identifies men with high-grade prostate cancer. *Eur Urol* 2015;68:464-70.
 26. Van Calster B, Nieboer D, Vergouwe Y, et al. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol* 2016;74:167-76.
 27. Schoots IG, Roobol MJ, Nieboer D, et al. Magnetic resonance imaging-targeted biopsy may enhance the diagnostic accuracy of significant prostate cancer detection compared to standard transrectal ultrasound-guided biopsy: a systematic review and meta-analysis. *Eur Urol* 2015;68:438-50.
 28. Roobol MJ, Vedder MM, Nieboer D, et al. Comparison of two prostate cancer risk calculators that include the prostate health index. *Eur Urol Focus* 2015;1:185-90.
 29. Vedder MM, de Bekker-Grob EW, Lilja HG, et al. The added value of percentage of free to total prostate-specific antigen, PCA3, and a kallikrein panel to the ERSPC risk calculator for prostate cancer in prescreened men. *Eur Urol* 2014;66:1109-15.
 30. Perdonà S, Cavadas V, Di Lorenzo G, et al. Prostate cancer detection in the "grey area" of prostate-specific antigen below 10 ng/ml: head-to-head comparison of the updated PCPT calculator and Chun's nomogram, two risk estimators incorporating prostate cancer antigen 3. *Eur Urol* 2011;59:81-7.
 31. Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J Urol* 1974;111:58-64.
 32. Loeb S, Folkvaljon Y, Robinson D, et al. Evaluation of the 2015 Gleason Grade Groups in a Nationwide Population-based Cohort. *Eur Urol* 2016;69:1135-41.
 33. Kweldam. CF, Kümmerlin. IP, Nieboer. D, et al. Prostate cancer outcomes of men with biopsy Gleason score 6 and 7 without cribriform or intraductal carcinoma. *Eur J Cancer* 2016;66:26-33.

Cite this article as: Pereira-Azevedo N, Verbeek JF, Nieboer D, Bangma CH, Roobol MJ. Head-to-head comparison of prostate cancer risk calculators predicting biopsy outcome. *Transl Androl Urol* 2018;7(1):18-26. doi: 10.21037/tau.2017.12.21

Supplementary

Table S1 Inclusion criteria of the different risk calculators

RC	Age (years)	N of biopsy cores	PSA, ng/mL	Prostate volume
ERSPC RPCRC	50–74	≥6	<50	10–110 mL
Finne	55–67	≥6	–	–
Chun	–	≥10	–	–
Karakiewicz	–	≥6	<50	–
PCPT 1.0	>55	≥6	–	–
PCPT 2.0	>55	≥6	–	–
ProstataClass	<89	≥6	<25	–
Sunnybrook	–	12	<50	–

PSA, prostate-specific antigen.

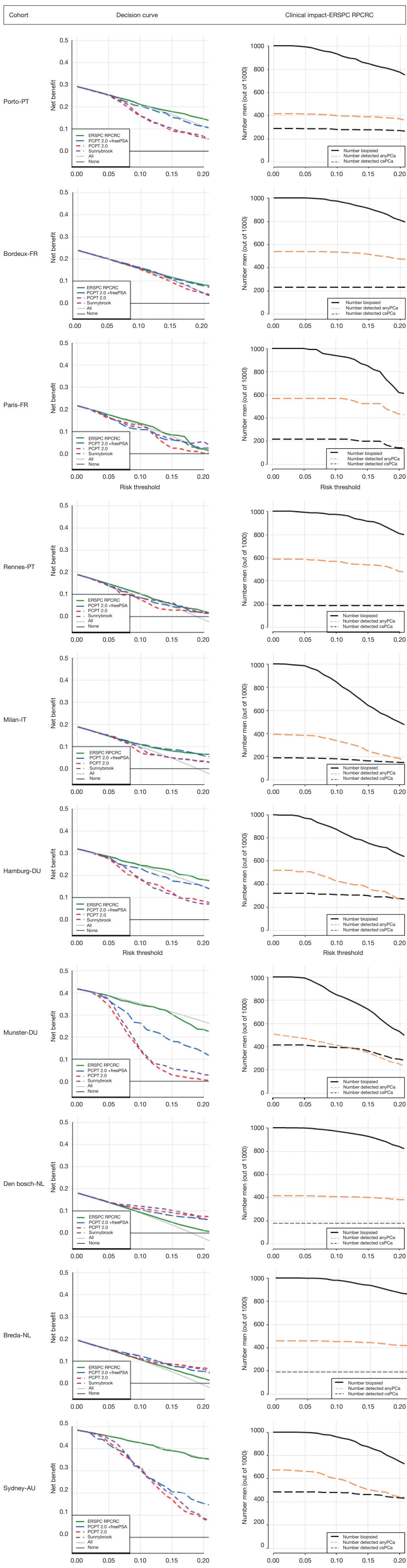
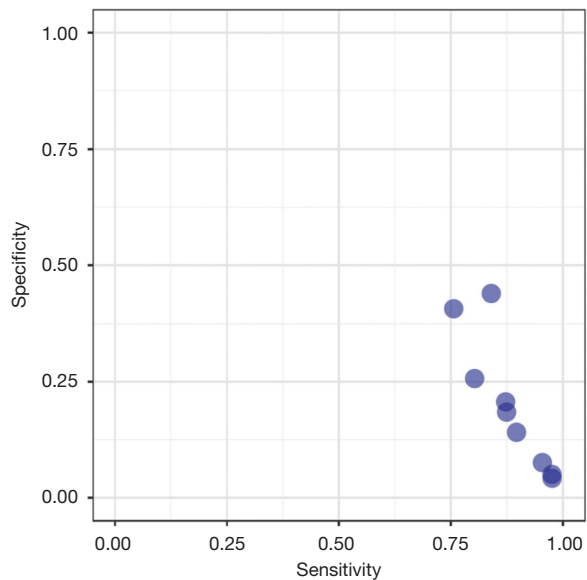


Figure S1 Decision curve for the four models predicting clinically significant prostate cancer in all the individual cohorts (left panel) and clinical impact plots (right panel) for predicting clinical significant prostate cancer (csPCa) with the ERSPC RPCRC. Of the 1,000 patients, the black solid line shows the total number of men who would receive a biopsy for each risk threshold. The orange dashed line shows how many of those would be detected with any prostate and the black dashed line shows how many would be detected with csPCa. The space between the two dashed lines indicated the number of men with indolent PCa.



Cohort	Sensitivity	Specificity
Porto	0.90	0.14
Bordeaux	0.95	0.08
Paris	0.84	0.44
Rennes	0.76	0.41
Milan	0.87	0.18
Hamburg	0.87	0.21
Munster	0.80	0.26
Den Bosch	0.98	0.05
Groningen	0.98	0.04

Figure S2 Sensitivity and specificity per study cohort of the biomarker PSA applying a cut-off for biopsy of ≥ 4.0 ng/mL and predicting the presence of clinically significant prostate cancer (defined as $GS \geq 7$). PSA, prostate specific antigen; GS, Gleason score.