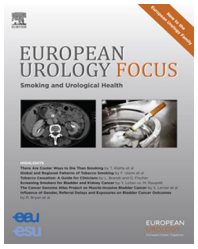
EUROPEAN UROLOGY FOCUS XXX (2016) XXX–XXX

available at www.sciencedirect.com
journal homepage: www.europeanurology.com/eufocus



Prostate Cancer

Prediction of Prostate Cancer: External Validation of the ERSPC Risk Calculator in a Contemporary Dutch Clinical Cohort

Maudy Gayet[a,b,†,*], *Christophe K. Mannaerts*[a,†], *Daan Nieboer*[c], *Harrie P. Beerlage*[a,b], *Hessel Wijkstra*[b,d], *Peter F.A. Mulders*[e], *Monique J. Roobol*[c]

[a] *Department of Urology, Jeroen Bosch Hospital, 's-Hertogenbosch, The Netherlands;* [b] *Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands;* [c] *Department of Urology, Erasmus Medical Centre, Rotterdam, The Netherlands;* [d] *Department of Urology, AMC University Hospital, Amsterdam, The Netherlands;* [e] *Department of Urology, Radboudumc University Hospital, Nijmegen, The Netherlands*

Article info

Article history:
Accepted July 21, 2016

Associate Editor:
James Catto

Keywords:
Nomogram
Decision aids
Risk stratification
Validation
Prostate cancer
Biopsy

Abstract

Background: The validity of prediction models needs external validation to assess their value beyond the original development setting.
Objective: To report the diagnostic accuracy of the European Randomized Study of Screening for Prostate Cancer (ERSPC) risk calculator (RC)3 and RC4 in a contemporary Dutch clinical cohort.
Design, setting, and participants: We retrospectively identified all men who underwent prostate biopsy (PBx) in the Jeroen Bosch Hospital, The Netherlands, between 2007 and 2016. Patients were included if they met ERSPC RC requirements of age (50–80 yr), prostate-specific antigen (PSA) (0.4–50 ng/ml), and prostate volume (10–150 ml). The probability of a positive biopsy for prostate cancer (PCa) and significant PCa (Gleason score $\geq$7 and/or higher than T2b) were calculated and compared with PBx pathology results.
Outcome measurements and statistical analysis: Evaluation was performed by calibration, discrimination, and clinical usefulness using calibration plots, area under the receiver operating characteristic curves (AUCs), and decision curve analyses (DCAs), respectively.
Results and limitations: A total of 2270 PBx sessions were eligible for final analysis. Discriminative ability of RC3 (AUC) was 0.78 and 0.90 for any PCa and significant PCa, respectively. For RC4 the calculated AUCs were 0.62 (any PCa) and 0.76 (significant PCa). The calibration plots of RC3 showed good results for both any PCa risk and significant PCa risk. In the repeat PBx group, RC4 tended to underestimate outcomes for PCa and showed moderate calibration for significant PCa. DCA showed an overall net benefit compared with PSA and digital rectal examination (DRE) alone. Limitations of this study are its retrospective single-institution design, retrospectively assessed DRE outcomes, no time restrictions between the first and repeat biopsy sessions, and no anterior sampling in the repeat PBx protocol.
Conclusions: The ERSPC RCs performed well in a contemporary clinical setting. Most pronounced in the biopsy-naive group, both RCs should be favoured over a PSA plus DRE–based stratification in the decision whether or not to perform PBx.
Patient summary: We looked at the ability of the existing European Randomized Study of Screening for Prostate Cancer risk calculator (RC), using different clinical data to predict the presence of prostate cancer in Dutch men. The RC performed well and should be favoured in the decision of whether or not to perform prostate biopsies over the conventional diagnostic pathway.

‡ These authors contributed equally to this work.
* Corresponding author. Jeroen Bosch Hospital, PO Box 90153 5200 ME, 's-Hertogenbosch, The Netherlands. Tel. +31 73 5532407; Fax: +31 73 5532373.
E-mail address: m.gayet@jbz.nl (M. Gayet).

http://dx.doi.org/10.1016/j.euf.2016.07.007
2405-4569/

1. Introduction

An estimated 1.1 million men worldwide were diagnosed with prostate cancer (PCa) in 2012, accounting for 15% of cancers in men, with 70% of them in more developed countries. PCa accounts for 6.6% of the total male cancer mortality. Incidence rates diverge, mainly because of serum prostate-specific antigen (PSA) testing [1]. First described in 1979, PSA made large-scale screening for PCa feasible. However, determination of serum PSA for diagnostic purposes lacks accuracy, with 15–25% false negatives and 60% false positives [2,3]. The likelihood of the presence of PCa is therefore preferably estimated by using additional clinical factors, such as digital rectal examination (DRE) and prostate volume (PV).

Although it has been shown that PCa-specific mortality can be reduced by 20% with PSA-based screening, population-based screening programs are not yet acceptable because of the high number needed to screen and the high number needed to treat to avoid one PCa death. More importantly, PSA-based screening results in a considerable number of unnecessary prostate biopsies (PBx) with potentially serious adverse events and leads to considerable overdiagnosis [4,5]. To achieve higher diagnostic accuracy, several nomograms and artificial neural networks (ANNs) have been developed to predict the outcome of PBx. These models have been shown to improve diagnostic accuracy compared with PSA alone [6,7]. However, it is necessary to assess the validity of these models outside the original development setting. Unfortunately, many of the published nomograms and ANNs lack external validation.

In 2006, different risk calculators (RCs) based on the Dutch section of the European Randomized Study of Screening for Prostate Cancer (ERSPC) were developed using data of men with a purely PSA-driven biopsy indication and a random transrectal ultrasound (TRUS)-guided sextant biopsy scheme [8]. ERSPC RC1 and RC2 are for patient use; RC3 (plus DRE), RC4 (plus DRE), RC5, and RC6 are for use by health care professionals at different stages of the testing process. Several external validation studies have been performed for these RCs. In both European and non-European cohorts, the accuracy of prediction of positive PBx in biopsy-naive or previously biopsied men using the ERSPC RC3 or RC4 was assessed, showing area under the curve (AUC) values in the range of 0.71–0.88 [9–12]. Until now, ERSPC RC3 plus DRE and RC4 plus DRE were externally validated using an extended biopsy scheme instead of a sextant biopsy scheme in both a Swiss and Irish cohort, with AUC for PCa and significant PCa of 0.66–0.77 and 0.85, respectively, and showing sufficient to good calibration [13,14].

The aim of this study was to assess the accuracy of the ERSPC RC3 and RC4 in a contemporary Dutch clinical cohort for which biopsy indications and number of biopsies differed from the development cohort.

2. Material and methods

2.1. Study population

We retrospectively identified all men who underwent PBx due to a clinical suspicion of PCa between January 2007 and December 2015 at the Jeroen Bosch Hospital. In our institution PBx was generally performed in patients with a serum PSA level $\geq$3.0–4.0 μg/l and/or an abnormal DRE. A standardised 12-core biopsy protocol consisting of two biopsies of each base, mid-gland, and apex in the peripheral zone of the prostate was performed, with additional cores taken when needed (eg, in case of hypoechogenic lesions). We examined patient files and obtained relevant clinical and pathologic data of each patient. Patients were included in our study if PCa risk prediction was considered relevant and possible, thus patients aged 50–80 yr with a PSA level between 0.4 and 50 μg/l, PV between 10 and 150 ml, and no previous positive PBx (ie, under active surveillance). Patients with a history of PCa were excluded. For our analyses, we retrospectively converted the descriptively documented DRE findings in our cohort to clinical T stages.

The patient database was blinded by PCa diagnosis and sent to one of the ERSPC RC designers (M.J.R.) for risk outcome calculations. Probabilities of detection of PCa and significant prostate PCa (Gleason score $\geq$7 and/or T stage higher than T2b) were calculated for each patient individually using two ERSPC RCs (www.prostatecancer-riskcalculator.com). RC3 was used to calculate probabilities in biopsy-naive patients; RC4 was used for patients with previous negative biopsy sessions undergoing a repeat PBx (Supplementary Table 1). The calculated probabilities were subsequently compared with the actual biopsy results for the entire cohort.

2.2. Statistics

Differences between clinical and pathologic variables in the studied cohort were assessed using the chi-square test for categorical variables and the Mann-Whitney *U* test for continuous variables. The performance of both RCs in the clinical setting was assessed by discrimination, calibration, and clinical usefulness.

Discrimination, that is, predictive accuracy, was quantified using the receiver operating characteristics derived AUC. Calibration refers to the agreement between observed and predicted outcomes with the extent of risk of over- or underestimation of the RCs evaluated graphically using calibration plots [15].

Clinical usefulness of the RCs was evaluated by decision curve analyses (DCAs) as described previously by Vickers and Elkin and by Steyerberg et al [16,17]. DCAs determine the value (net benefit) of a prediction model by examining the theoretical relationship between the threshold probability of an event (eg, PCa at biopsy) and the relative value of false-positive and false-negative results. We compared the RC model with a PSA plus DRE–based model, also developed on original ERSPC data. We also assessed the theoretical number of (significant) cases of PCa missed, numbers of biopsies saved, and number of Gleason score 6 PCa diagnoses saved at different RC thresholds.

Statistical analyses were performed using SPSS v23.0 (IBM Corp, Armonk, NY, USA) and R v3.2.5 (R Foundation for Statistical Computing, Vienna, Austria). A $p < 0.05$ was considered to indicate statistical significance in all analyses.

3. Results

We identified 2862 prostate biopsy sessions in 2124 men. Overall, 426 biopsy sessions were omitted due to the predefined inclusion criteria. In 166 biopsy sessions (<6%), data were incomplete (PSA, $n = 2$; DRE findings, $n = 123$; TRUS PV, $n = 27$; TRUS findings, $n = 40$) and excluded from further analyses. As a result, 2270 prostate biopsy sessions (79.3%) in 1812 different men were eligible for final analysis: 73.0% biopsy-naive men and 27.0% men with a prior negative PBx.

PCa and significant PCa were detected in 44.1% and 20.3% of the biopsy-naive men ($n = 1658$). Men with PCa and

Please cite this article in press as: Gayet M, et al. Prediction of Prostate Cancer: External Validation of the ERSPC Risk Calculator in a Contemporary Dutch Clinical Cohort. Eur Urol Focus (2016), http://dx.doi.org/10.1016/j.euf.2016.07.007