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

European Association of Urology

Prostate Cancer

A Contemporary Prostate Biopsy Risk Calculator Based on Multiple Heterogeneous Cohorts

Donna P. Ankerst[a,b,*], *Johanna Straubinger*[a], *Katharina Selig*[a], *Lourdes Guerrios*[c], *Amanda De Hoedt*[d], *Javier Hernandez*[b], *Michael A. Liss*[b], *Robin J. Leach*[b], *Stephen J. Freedland*[d,e], *Michael W. Kattan*[f], *Robert Nam*[g,h], *Alexander Haese*[i], *Francesco Montorsi*[j,k], *Stephen A. Boorjian*[l], *Matthew R. Cooperberg*[m], *Cedric Poyet*[n], *Emily Vertosick*[o], *Andrew J. Vickers*[o]

[a] *Department of Mathematics, Technical University of Munich, Garching, Munich, Germany;* [b] *Department of Urology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA;* [c] *Department of Surgery, Urology Section, Veterans Affairs Caribbean Healthcare System, San Juan, Puerto Rico;* [d] *Section of Urology, Durham Veterans Administration Medical Center, Durham, NC, USA;* [e] *Department of Surgery, Cedars-Sinai Medical Center, Los Angeles, CA, USA;* [f] *Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA;* [g] *Division of Urology, Sunnybrook Health Sciences Centre, Sunnybrook Research Institute, University of Toronto, Toronto, Ontario, Canada;* [h] *Institute of Health Policy, Management, & Evaluation, University of Toronto, Toronto, Ontario, Canada;* [i] *Martini-Clinic Prostate Cancer Center, University Clinic Eppendorf, Hamburg, Germany;* [j] *Division of Oncology/Unit of Urology, URI, IRCCS Hospital San Raffaele, Milano, Italy;* [k] *Department of Medicine, Vita-Salute San Raffaele University, Milano, Italy;* [l] *Department of Urology, Mayo Clinic, Rochester, MN, USA;* [m] *Departments of Urology and Epidemiology & Biostatistics, University of California San Francisco, San Francisco, CA, USA;* [n] *Department of Urology, University Hospital Zurich, University of Zurich, Zurich, Switzerland;* [o] *Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA*

Article info

Article history:
Accepted May 3, 2018

Associate Editor:
Giacomo Novara

Keywords:
Digital rectal exam
Family history
High-grade disease
Prostate cancer
Prostate-specific antigen
Risk prediction

Abstract

Background: Prostate cancer prediction tools provide quantitative guidance for doctor-patient decision-making regarding biopsy. The widely used online Prostate Cancer Prevention Trial Risk Calculator (PCPTRC) utilized data from the 1990s based on six-core biopsies and outdated grading systems.
Objective: We prospectively gathered data from men undergoing prostate biopsy in multiple diverse North American and European institutions participating in the Prostate Biopsy Collaborative Group (PBCG) in order to build a state-of-the-art risk prediction tool.
Design, setting, and participants: We obtained data from 15 611 men undergoing 16 369 prostate biopsies during 2006–2017 at eight North American institutions for model-building and three European institutions for validation.
Outcome measurements and statistical analysis: We used multinomial logistic regression to estimate the risks of high-grade prostate cancer (Gleason score $\geq$7) on biopsy based on clinical characteristics, including age, prostate-specific antigen, digital rectal exam, African ancestry, first-degree family history, and prior negative biopsy. We compared the PBCG model to the PCPTRC using internal cross-validation and external validation on the European cohorts.
Results and limitations: Cross-validation on the North American cohorts (5992 biopsies) yielded the PBCG model area under the receiver operating characteristic curve (AUC) as 75.5% (95% confidence interval: 74.2–76.8), a small improvement over the AUC of 72.3% (70.9–73.7) for the PCPTRC ($p < 0.0001$). However, calibration and clinical net benefit were far superior for the PBCG model. Using a risk threshold of 10%, clinical use of the PBCG model would lead to the equivalent of 25 fewer biopsies per 1000 patients without missing any high-grade cancers. Results were similar on external validation on 10 377 European biopsies.
Conclusions: The PBCG model should be used in place of the PCPTRC for prediction of prostate biopsy outcome.
Patient summary: A contemporary risk tool for outcomes on prostate biopsy based on the routine clinical risk factors is now available for informed decision-making.

* Corresponding author. Department of Mathematics, Technical University Munich, Boltzmannstr 3, 85748 Garching, Germany. Tel. +49 89 289 18388.
E-mail address: ankerst@tum.de (D.P. Ankerst).

https://doi.org/10.1016/j.eururo.2018.05.003
0302-2838/

1. Introduction

The decision to conduct a prostate biopsy for the suspicion of prostate cancer is far from trivial, with potential consequences including sepsis, over-diagnosis of indolent disease or, conversely, delayed diagnosis of an aggressive cancer [1,2]. The new emphasis on shared-decision making for medical procedures has increased interest in decision tools to allow improved explanation of risk during the physician–patient interaction [3]. Many of the risk tools were developed on large comprehensive cohorts, where statistical modeling integrated the influences of established risk factors, such as prostate-specific antigen (PSA), digital rectal exam (DRE), race, and age, on biopsy outcomes.

Two of the most commonly used risk tools were built on large prospective screening trials: the European Randomized Study of Screening for Prostate Cancer (ERSPC) risk tool and the Prostate Cancer Prevention Trial (PCPT) risk calculator (PCPTRC) [4,5]. Both trials were performed during the 1990s and hence based on the study populations and clinical practice standards of that time. The PCPT was a heavily screened population of primarily healthy North American white men, requiring a PSA $\leq$3 ng/ml and a normal DRE to enter the trial, annual PSA and DRE screening, and a required end-of-study biopsy at the end of 7 yr. The ERSPC comprised a near exclusively white European population also heavily screened [6,7]. These populations are not in accord with contemporary patients from diverse backgrounds who are likely to undergo limited screening during the lifetime and only ever encounter risk assessment after an elevated PSA prompts referral to a tertiary care center. Both the ERSPC and PCPT involved sextant biopsy and grading scheme current in the 1990s. Contemporary biopsies utilize 10–12 twelve cores and are subject to pathologic grading under contemporary schemes that reclassify some cancers to higher Gleason scores [8,9].

To better understand the relationships between prostate biopsy outcomes and established risk factors in heterogeneous cohorts, the Prostate Biopsy Collaborative Group (PBCG) was formed in 2009 as a consortium collecting retrospective data from 10 screening studies and tertiary referral centers [10]. Validation of both the ERSPC and PCPT risk tools revealed differences in operating characteristics across the cohorts, demonstrating that validation is a property of both the risk tool and the validation cohort [11,12]. Significant amounts of missing risk factors prohibited robust conclusions and further use from the retrospective data.

To ensure high data quality for production of a new prostate cancer risk tool based on heterogeneous contemporary populations and practice, the PBCG began prospective collection from participating centers in 2014. The new risk tool was to be modeled after the PCPTRC, with the hypothesis that such a risk tool would have better external validation for contemporary populations [13].

2. Patients and methods

Data from 11 participating sites under local internal review board approval were prospectively collected. Cleveland Clinic, Hamburg, Mayo Clinic, San Raffaele, Zurich, Memorial Sloan Kettering Cancer Center (MSKCC), and University of California San Francisco (UCSF) were participating tertiary referral centers. Durham Veterans Affairs (VA) and San Juan VA served a lower socioeconomic status population with a high percentage of African Americans and Hispanics, respectively. Sunnybrook and UT Health were consortia that include main hospitals, tertiary referral centers, and associated community urology providers. Four sites also provided retrospective data for prostate biopsies performed in 2006 or later.

This risk tool predicts three outcomes on biopsy: high-grade disease (Gleason score $\geq$7), low-grade disease (Gleason score $<$7), and benign findings (no prostate cancer). We chose the study outcomes as they reflect the clinical purpose of the tool, which is to aid biopsy decision-making. It is unequivocal that although we may not treat a man with Gleason 3 + 4, we do need to at least evaluate it before making treatment decisions. The definition of high-grade as Gleason score 3 + 4 or higher is entirely conventional, being used, in addition to PCPTRC version 2.0, for numerous studies including the Stockhom-3 model, the ERSPC Rotterdam Section update, and the Canary Prostate Active Surveillance Study [13–16].

We compared patient and biopsy characteristics between the training and validation sets using chi-square and Wilcoxon tests. We examined the relationship between prevalence of the risk factors in each cohort to the odds ratio from a logistic regression of high-grade cancer on the risk factor alone for each of the PBCG cohorts, along with results from the PCPT population of 6664 biopsies used to build the PCPTRC version 2.0 [13]. We fit a multinomial logistic model to estimate risks of high- versus low-grade versus no cancer with predictors age, PSA (logarithmically transformed), DRE, African ancestry, first-degree family history, and prior negative biopsy history to data from all eight cohorts pooled together. Prostate volume was not included in the models as this requires an invasive test, a transrectal ultrasound, to obtain. The online PCPTRC allows missing values for DRE, family history, and negative prior biopsy history using marginal models fit without these covariates, and we followed the same procedure. For both model fitting and validation, we imputed missing values of African ancestry to be non-African; a sensitivity check determined results to be similar.

We compared the validation performance of the PBCB and PCPTRC models for the prediction of high-grade cancer versus the other endpoints combined (low-grade and no cancer) in terms of discrimination measured by the area under the receiver operating characteristic curve (AUC), accuracy measured by calibration curves, and clinical utility based on net benefit. For internal validation, we used leave-one-cohort-out cross-validation, whereby each of the eight North American cohorts alternatively served as a hold-out test set for a model fit to pooled data from the seven remaining cohorts. Predictions for each use of the eight cohorts as a test set were pooled for comparison to the PCPTRC. For external validation, we compared the PBCG model fit to all eight North American cohorts to the PCPTRC on data from all three European cohorts.

3. Results

We fit the PBCG model on 5992 biopsies from eight institutions in North America and validated it on 10 377 biopsies from three institutions in Europe. Descriptions of the pooled cohorts for fitting and validation are provided in Table 1 and by individual cohort in the Supplementary Appendix. Median age (approximately 65 yr) and PSA (approximately 6 ng/ml) were fairly consistent between cohorts. The rate of positive DRE varied from 13% in MSKCC to 51% in San Juan VA; the proportion of patients with African ancestry also varied, from $<$1% at European sites to 63% at the Durham VA. Prevalence of high-grade cancer

Please cite this article in press as: Ankerst DP, et al. . Eur Urol (2018), https://doi.org/10.1016/j.eururo.2018.05.003