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

European Association of Urology

Platinum Priority – Prostate Cancer

Editorial by XXX on pp. x–y of this issue

Identifying the Most Informative Prediction Tool for Cancer-specific Mortality After Radical Prostatectomy: Comparative Analysis of Three Commonly Used Preoperative Prediction Models

Katharina Boehm[a,b,*], *Alessandro Larcher*[b,c], *Burkhard Beyer*[a], *Zhe Tian*[b,d], *Derya Tilki*[a,e], *Thomas Steuber*[a], *Pierre I. Karakiewicz*[b], *Hans Heinzer*[a], *Markus Graefen*[a], *Lars Budäus*[a]

[a] *Martini Clinic, Prostate Cancer Centre at University Hospital Hamburg-Eppendorf, Hamburg, Germany;* [b] *Cancer Prognostics and Health Outcomes Unit, University of Montreal Health Center, Montreal, Canada;* [c] *Division of Oncology, Unit of Urology, URI, IRCCS Ospedale San Raffaele, Milan, Italy;* [d] *Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada;* [e] *Department of Urology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

Article info

Article history:
Accepted July 27, 2015

Associate Editor:
Matthew Cooperberg

Keywords:
Radical prostatectomy
Prostate cancer
Cancer-specific survival
Predictive models
Decision curve analysis

Abstract

Background: The D'Amico risk stratification, Cancer of the Prostate Risk Assessment (CAPRA) score, and Stephenson nomogram are widely used prediction tools for biochemical recurrence and survival after radical prostatectomy (RP). These models have not been compared with respect to cancer-specific mortality (CSM) prediction.
Objective: To validate and compare the prediction tools for 10-yr CSM.
Design, setting, and participants: Overall, 2485 prostate cancer patients underwent RP in a European tertiary care center.
Outcome measurements and statistical analysis: Three preoperative models (D'Amico, CAPRA, and Stephenson) were compared in terms of their ability to predict 10-yr CSM; therefore, accuracy tests (area under the receiver operating characteristic curve [AUC]), calibration plots, and decision curve analysis (DCA) were assessed for each model.
Results and limitations: CSM at 10 yr was 3.6%. The AUC was 0.76, 0.77, and 0.80 for the D'Amico, CAPRA, and Stephenson models, respectively. In calibration plots, predicted probabilities were close to the observed probabilities for the D'Amico model but showed underestimation of CSM for the Stephenson nomogram and overestimation of CSM for the CAPRA score. DCA identified a benefit for the CAPRA score. These results apply to patients treated at a European tertiary care center.
Conclusions: Despite good discriminatory power, all tested models had some shortcomings in terms of prediction of 10-yr CSM. All three models showed good performance in North American cohorts, but our results suggested a lack of generalizability to European patients. To overcome this issue, local recalibration of the variable weights could be performed. Another possibility is the development of more universal markers that are independent of regional practice differences or, alternatively, the development of better tools to quantify clinical practice differences.
Patient summary: Prediction tools can predict cancer survival prior surgery, relying on points for age, prostate-specific antigen levels, aggressiveness, and percentage of cancer at biopsy. These tools are reliable in North American patients but have shortcomings for identifying patients at high risk of prostate cancer death in Europe.

* Corresponding author. Cancer Prognostics and Health Outcomes Unit, 264 Blvd. Rene-Levesque E. Room 228, Montreal QC H2X 1P1, Canada. Tel. +1 514 890 8000 extension 35335; Fax: +1 514 227 5103.
E-mail address: boehm@martini-klinik.de (K. Boehm).

http://dx.doi.org/10.1016/j.eururo.2015.07.051
0302-2838/

1. Introduction

Radical prostatectomy (RP) is the gold standard treatment modality for clinically localized prostate cancer (PCa) [1]. Excellent cancer-specific survival rates have been reported for the majority of the patients [2–4]. Nevertheless, some patients are at higher risk of biochemical recurrence (BCR) and cancer-specific mortality (CSM) [5]. In times of shared decision making, clinicians and patients are in need of evidence-based information about risk of recurrence and death. To simplify this task, different predictive models like nomograms, risk scores, or risk groups are available. In addition, these models are commonly used to design prospective trials among patients at particularly higher risk for an unfavorable outcome, with CSM as the end point. In clinical research, predicted probabilities derived from one or several prediction tools are used to adequately discriminate among patients at different risk of CSM, as demonstrated by Cooperberg et al [6] and Lee et al [7].

Three of the most common prediction tools are the D'Amico risk stratification scheme [8], the Cancer of the Prostate Risk Assessment (CAPRA) score [9], and the Stephenson nomogram [10]. All of these tools were initially used and validated to predict BCR after RP [9–13]. In addition, the D'Amico risk stratification scheme and the Stephenson nomogram are widely used for prediction of survival after RP [4,14]. All of these tools are based on North American patient cohorts and validated in such populations. To date, no study has directly compared the ability of the three models to predict CSM in European patients.

To address this issue, we evaluated the ability of the D'Amico risk stratification scheme, the CAPRA score, and the Stephenson nomogram to predict 10-yr CSM in a head-to-head fashion; therefore we tested discrimination and calibration and additionally relied on decision curve analysis (DCA).

2. Patients and methods

2.1. Study population and intervention

A total of 2485 PCa patients treated with RP in a single high-volume European center between 1992 and 2005 were included in this study. Patients with missing data for clinical and biopsy variables (eg, prostate-specific antigen [PSA] level, clinical stage, percentage of cancer in biopsy, Gleason score at biopsy) were excluded. In addition, patients treated with neoadjuvant androgen deprivation therapy (ADT) were also excluded.

RP was performed by staff urologists using an open retropubic approach, as described previously [15–18]. All patients were diagnosed by multicore transrectal ultrasound–guided biopsy. Biopsy cores were graded according to the Gleason system [19], and clinical stage was assigned by the attending urologist according to the 1992–2002 American Joint Committee on Cancer TNM guidelines. Questionnaires and death reports of the national cancer registry were used annually for follow-up. Adjuvant or salvage treatment was applied according to the treating urologist's usual practices. All data were prospectively stored in an institutional database (FileMaker Pro 10; FileMaker, Inc., Santa Clara, CA, USA).

2.2. Predictive models

The predictor variables were categorized in the same manner, as described, for D'Amico risk groups [8], the CAPRA score [20], and the Stephenson nomogram (2009) [4,10]. Specifically, the D'Amico risk groups rely on three variables. PSA levels were categorized as $\leq$10, 10.1–20, and $\geq$20 ng/ml; clinical stage categories were cT1/2a, cT2b, and cT2c/T3 or higher; and biopsy Gleason score was categorized as $\leq$6, 7, and $\geq$8.

The CAPRA score is based on five variables, categorized as follows: PSA levels of $\leq$6, 6.1–10.0, 10.1–20.0, 20.1–30, or $\geq$30 ng/ml; clinical stage of cT1/2 or cT3; biopsy Gleason score of no Gleason pattern 4, secondary Gleason pattern 4/5, or primary Gleason pattern 4/5; age of $<$50 or $\geq$50 yr; and percentage of cancer in the biopsy $<$34% or $\geq$34%.

The Stephenson nomogram is predicated on four variables: continuously coded PSA level; primary biopsy Gleason pattern $\leq$3 versus $\geq$4; secondary biopsy Gleason pattern $\leq$3 versus $\geq$4; and clinical stage cT1, cT2a, cT2b, cT2c, and cT3.

2.3. Statistical analyses

A cumulative incidence smoothed plot illustrated CSM and other-cause mortality (OCM) at 15 yr. To test discrimination of the three models, the survival analysis extension to the area under the receiver operating characteristic curve (AUC) was assessed [21]. The AUC ranges from 0.5 (chance) to 1.0 (perfect discrimination) and describes the probability in a randomly chosen pair of patients that the one with the event (eg, CSM) has a higher predicted probability of experiencing the event.

To assess miscalibration, calibration plots for prediction of 10-yr CSM were provided for all three models. Previously published risk estimates of CSM for the D'Amico risk groups [14] were plotted against the observed risk in our cohort. The estimates for the D'Amico risk stratification were derived from a North American population of 7591 men who underwent RP between 1987 and 2003 in a single tertiary care center [14].

The same approach was used for the CAPRA score [22]. The estimates for 10-yr CSM stem from a multi-institutional population (n = 10 627) from the United States (CaPSURE database, 1992–2007). Approximately 80% of this population was locally treated (approximately 50% with RP) [22].

The Stephenson nomogram also provides a risk of CSM [4]. The risk estimates for the Stephenson nomogram were derived from North American patients (n = 12 677), who underwent RP between 1987 and 2005 [4]. For each model, the predicted probability of 10-yr CSM was compared with the observed probability of 10-yr CSM.

Finally, DCA for competing risk regression was performed to compare the three models in a head-to-head fashion [23]. We relied on DCA to evaluate which model was most helpful for the identification of patients at risk of CSM at 10 yr. Decision curves depict the net benefit against the threshold probability. As an example of different potential threshold probabilities in our study, some clinical trials might include patients with a 3% risk of CSM, others if the risk is $\geq$6%.

All tests were two-tailed, and p values $<$0.05 were considered statistically significant. Statistical analyses were performed with R v.3.1.0 (R Foundation, Vienna, Austria; https://www.r-project.org).

3. Results

The patient population consisted of 2485 men treated with RP between years 1992 and 2005 at a single high-volume European center. Median follow-up time among survivors in our patient cohort was 9.5 yr (interquartile range [IQR]: 7.2–12). Among survivors, 1043 patients had follow-up $>$10 yr.

Patient characteristics are shown in Table 1. Mean age was 62 yr, and the median PSA level was 7.0 ng/ml. Distribution of Gleason scores 3 + 3, 3 + 4, 4 + 3 and $\geq$4 + 4 at biopsy was 1529 (62%), 656 (26%), 209 (8.4%), and 91 (3.7%), respectively. Proportions of clinical stage cT1, cT2a, cT2b, cT2c, and cT3

Please cite this article in press as: Boehm K, et al. Identifying the Most Informative Prediction Tool for Cancer-specific Mortality After Radical Prostatectomy: Comparative Analysis of Three Commonly Used Preoperative Prediction Models. Eur Urol (2015), http://dx.doi.org/10.1016/j.eururo.2015.07.051