Is Extraprostatic Extension of Cancer Predictable? A Review of Predictive Tools and an External Validation based on a Large and a Single Center Cohort of Prostate Cancer Patients

Bernardo Rocco Associate Professor of Urology ,
Maria Chiara Sighinolfi Phd, MD ,
Marco Sandri Contract Professor , Ahmed Eissa MD ,
Ahmed Elsherbiny , Ahmed Zoeir ,
Harvey Tadzia Clinical Research Coordinator ,
Hariharan Palayapalayam Clinical Fellow of Urology ,
Darian Kameh Medical Laboratory Director ,
Rafael Coelho Associate professor of urology ,
Stefano Puliatti Urology Resident Physician ,
Paola Zuccolotto Full Professor ,
Rodolfo Montironi Full Professor of Pathology Head ,
Peter Wiklund Full Professor of Urology , Salvatore Micali ,
Giampaolo Bianchi Full Professor of Urology , Vipul Patel

PII: S0090-4295(19)30316-4
DOI: https://doi.org/10.1016/j.urology.2019.03.019
Reference: URL 21522

To appear in: *Urology*

Received date: 23 January 2019
Accepted date: 21 March 2019

Please cite this article as: Bernardo Rocco Associate Professor of Urology , Maria Chiara Sighinolfi Phd, MD , Marco Sandri Contract Professor , Ahmed Eissa MD , Ahmed Elsherbiny , Ahmed Zoeir , Harvey Tadzia Clinical Research Coordinator , Hariharan Palayapalayam Clinical Fellow of Urology , Darian Kameh Medical Laboratory Director , Rafael Coelho Associate professor of urology , Stefano Puliatti Urology Resident Physician , Paola Zuccolotto Full Professor , Rodolfo Montironi Full Professor of Pathology Head , Peter Wiklund Full Professor of Urology , Salvatore Micali , Giampaolo Bianchi Full Professor of Urology , Vipul Patel , Is Extraprostatic Extension of Cancer Predictable? A Review of Predictive Tools and an External Validation based on a Large and a Single Center Cohort of Prostate Cancer Patients, *Urology* (2019), doi: https://doi.org/10.1016/j.urology.2019.03.019

Is Extraprostatic Extension of Cancer Predictable? A Review of Predictive Tools and an External Validation based on a Large and a Single Center Cohort of Prostate Cancer Patients

Bernardo Rocco[1], Maria Chiara Sighinolfi[2], Marco Sandri[3], Ahmed Eissa[4], Ahmed Elsherbiny[4], Ahmed Zoeir[4], Harvey Tadzia[5], Hariharan Palayapalayam[6], Darian Kameh[7], Rafael Coelho[8], Stefano Puliatti[9], Paola Zuccolotto[10], Rodolfo Montironi[11], Peter Wiklund[12], Salvatore Micali[1], Giampaolo Bianchi[13], Vipul Patel[14]

1. Associate Professor of Urology, Department of Urology, Ospedale Policlinico e Nuovo Ospedale Civile S.Agostino Estense Modena, University of Modena and Reggio Emilia, Modena, Italy

2. Phd, MD, Department of Urology, Ospedale Policlinico e Nuovo Ospedale Civile S.Agostino Estense Modena, University of Modena and Reggio Emilia, Modena, Italy

3. Contract Professor, Data Methods and Systems Statistical Laboratory, University of Brescia, Brescia, Italy

4. MD, Urology Department Faculty of Medicine – Tanta University- Egypt. - Fellow in Urology, Department of Urology, Ospedale Policlinico e Nuovo Ospedale Civile S.Agostino Estense Modena, University of Modena and Reggio Emilia, Modena, Italy

5. Clinical Research Coordinator, Global Robotics Institute, Florida Hospital-Celebration Health Celebration, University of Central Florida School of Medicine, Orlando, FL, USA

6. Clinical Fellow of Urology, Global Robotics Institute, Florida Hospital-Celebration Health Celebration, University of Central Florida School of Medicine, Orlando, FL, USA

7. Medical Laboratory Director, Global Robotics Institute, Florida Hospital-Celebration Health Celebration, University of Central Florida School of Medicine, Orlando, FL, USA

8. Associate professor of urology, Instituto do Câncer do Estado de São Paulo (Icesp), SP, Brazil.

9. Urology Resident Physician, Department of Urology, Ospedale Policlinico e Nuovo Ospedale Civile S.Agostino Estense Modena, University of Modena and Reggio Emilia, Modena, Italy

10. Full Professor, Big&Open Data Innovation Laboratory University of Brescia, Brescia, Italy

11. Full Professor of Pathology Head, Genitourinary Program, Section of Pathological Anatomy, Polytechnic University of the Marche Region, School of Medicine, United Hospitals, Ancona, Italy

12. Full Professor of Urology, Department of Urology, Karolinska University Hospital, Stockholm, Sweden

13. Full Professor of Urology, Department of Urology, Ospedale Policlinico e Nuovo Ospedale Civile S.Agostino Estense Modena, University of Modena and Reggio Emilia, Modena, Italy

14. Medical Director at the Global Robotics Institute, Florida Hospital-Celebration Health Celebration, University of Central Florida School of Medicine, Orlando, FL, USA

Key Words: prostate cancer, extraprostatic extension, predictive tools, external validation

CONFLICT OF INTEREST:

Authors declare that there is no conflict of interest regarding the publication of this manuscript

Corresponding Author:

Bernardo Rocco,

Associate Professor of Urology, National Scientific Qualification Full professorship

Department of Urology, Ospedale Policlinico e Nuovo Ospedale Civile S.Agostino Estense Modena, University of Modena and Reggio Emilia, Modena, Italy

Via del Pozzo, 71 Modena, 41124, Italy.

Tel. +39 059 396 2138

E-Mail: Bernardo.rocco@gmail.com

Abstract

Our aim was to review and externally validate all the available predictive tools (PTs) predicting EPE using the area under the curve (AUC), calibration plots and scaled brier score.

A literature search was performed showing 19 models predicting EPE. External validation (EV) was carried out on 6360 prostate cancer (PCa) patients submitted to RP. Most of the PTs showed poor discrimination and unsatisfactory calibration.

The majority of the available PTs are not reliable for the prediction of EPE in populations other than the development one; thus, they may not be completely appropriate for patients' counselling or for surgical strategy preplanning.

1. Introduction

Prostate cancer (PCa) represents a major health concern of male sex. International guidelines recommend radical prostatectomy (RP) for localized PCa patients <65 years old with life expectancy >10 years [1, 2]. Erectile dysfunction is a potential drawback of RP , that has to deal with a trade-off between oncological safety and functional outcomes [3]. In 1983, Walsh introduced the nerve-sparing RP (NSRP) to improve the post-operative erectile function [4]. The AUA and the European Association of Urology (EAU) guidelines emphasize the value of NSRP for localized PCa patients seeking post-operative potency [1, 2].

The NSRP may lead to increased incidence of positive surgical margin (PSM) and subsequent biochemical recurrence [5, 6]. Thus, prediction of extraprostatic extension (EPE) of PCa is the cornerstone to determine patients' eligibility for NSRP [7]. Approximately, EPE at final pathology is found in 20% of men with clinically localized PCa [8].The pre-surgical planning has been increasingly performed using predictive tools (PTs) based on common clinical-pathological features [4, 7, 9-27] and the EAU guidelines recommend referral to externally validated PTs to select patients for NSRP [2]. Moreover, some of those models have user-friendly web access, and patients can easily consult them. However, there is an ominous gap between their potential and actual predictive performance in clinical practice [28], because of its probable optimistic performance during development and the lack of high quality external validation (EV) studies [29].

The aim of our study is to provide an accurate EV of the available PTs of EPE on a large cohort of patients.

2. Materials and methods

2.1 Search criteria

A systematic search of the Medline database was performed until December 6, 2017 using a combination of multiple keywords including: “prostate cancer”, “prostate neoplasm”, “radical prostatectomy”, “extracapsular extension”, “ECE”, “extra-capsular extension”, “extraprostatic extension”, “extra-prostatic extension”, “pathological stage”, “capsular perforation”, “organ confined”, “nomograms”, “validation”, “predictive tools”, “prediction”, and “predicting”. Three of our authors (A.E., A.E. and S.P.) were responsible for the search and the article selection process, and any discrepancies were resolved. Our inclusion criteria were: (1) original articles published in English, (2) integration of multiple variables to build a predictive tool, (3) same definitions of EPE, (4) using reproducible statistical tests without any missing data required for calculating model prediction, and (5) using variables available in our dataset.

2.2 Reporting

The EV was performed according to the TRIPOD statement [30].

2.3 Patient population

Data of 6360 patients who underwent robotic-assisted prostatectomy (RALP) between 2008 and 2016 at the Global Robotics Institute of Celebration (FL, USA) were used as the validation dataset.

2.4 Surgical technique

All the procedures were performed by a single surgeon (VP) using the Da Vinci Surgical

System, as previously described [31].

2.5 Preoperative clinical variables analyzed

Preoperative clinical variables included patient's age, body mass index, total prostate specific antigen (PSA) level, PSA density, prostate volume, and clinical stage (American Joint Committee on Cancer (AJCC) TNM staging 1992/2002) [32]. Moreover, a side-specific clinical-T-stage was determined analyzing 11,794 prostatic lobes (6,360 patients). For example, when a patient is assigned to cT2a, the abnormally palpable lobe was considered to be stage T2a while the normal lobe was assigned to stage T1c. On the other hand, a patient with abnormally palpable tumor on both sides was considered to have cT2c in each lobe [15, 18].

2.6 Pathological analysis of prostate biopsy cores

Biopsy variables considered for each lobe were total number of cores, Gleason score, and the number of positive cores. Moreover, the percentage of positive cores and maximum percentage of cancer were considered.

2.7 Pathologic analysis of prostate specimen

Pathological analysis of specimens was described before [33] and includes:

1) The pathological T-stage (AJCC TNM Staging, 1992/2002) [32].
2) Histological pattern and Gleason score [34].
3) PSM: the presence of carcinoma on the prostatic-inked surface.
4) According to the definitions found in the literature, two distinct definitions were considered for EPE (**Supplementary Figure 1** illustrates the difference):

- pT3a: the presence of tumor beyond the confines of the prostate *without* invasion of the seminal vesicles.

- Whole EPE (wEPE): the presence of tumor beyond the confines of the prostate *regardless* the status of seminal vesicles.

5) Seminal vesicle invasion (SVI)

2.8 Statistical analysis

Receiver operating characteristics (ROC) curves were calculated to assess the ability of the prediction models to discriminate between patients with or without EPE. The area under the ROC curve (AUC) with 95% confidence interval (CI) was estimated. AUC ranges between 0.5 and 1; a value of 0.5 indicates no discrimination, $0.5 < AUC < 0.7$ poor discrimination, $0.7 \leq AUC < 0.8$ acceptable discrimination, $0.8 \leq AUC < 0.9$ excellent discrimination, $0.9 \leq AUC < 1$ outstanding discrimination, and 1 indicates perfect discrimination [35].

Calibration of the model was investigated to show the relationship between model-predicted and observed rates of EPE. Agreement between predicted and actual probabilities was assessed graphically by plotting LOESS-smoothed calibration curve together with the 45° line of perfect calibration. Deviations from the ideal line were characterized estimating intercept and slope of the line approximating the calibration curve [36]. Furthermore, the estimated calibration index (ECI) was calculated to compare the calibration of the different PTs with 0 representing perfect calibration [37].

The Brier score is the average squared difference between the actual outcomes and the predicted probabilities : $- \sum$. It is a measure of overall performance because it can be decomposed into two components: the first related to calibration and the second related to discrimination. For convenience, the scaled Brier score (SBS) ——— was reported in the study, where $-$ $-$ and $-$ indicating the average probability of the outcome. When the scaled Brier score

is negative or close to zero, the overall predictive ability of the model is worse than or similar to a non-informative model; when SBS is 1 the model returns a perfect prediction [36].

As regards the calculation of predictions, when coefficients of logistic regression were available, predicted probabilities were calculated using the formula associated to this model. However, when only a nomogram was given, the image was digitized, the coefficients of the linear functions were estimated, the single scores were added up, the logit function was applied and finally the predicted probabilities were calculated.

For each PT, a comparison between the distribution of patients' characteristics in development and EV datasets (Supplementary Table 2) was performed using two-sample test of proportions Pearson's chi-squared test (categorical variables) and two-sample t-test (numerical variables). The comparability of EV and development populations was also assessed using the standardized difference [38] which is defined for dichotomous variables as

$$\frac{}{\sqrt{}}$$

where and denote the prevalence in EV and development populations, respectively; for continuous variables is defined as

$$\frac{\bar{} \quad \bar{}}{\sqrt{}}$$

where $\bar{}$, $\bar{}$, denote the mean and standard deviation in EV and development populations, respectively. For categorical variables with k>2 categories, the maximum of the k standardized differences was reported. Values of in the range $-0.1 \leq d \leq 0.1$ can be considered a sign of good balance between variable distributions in the two populations [38].

Regarding missing values, no imputation method was used and a complete-case analysis was performed. All analyses were performed using R software (version 3.4.3; R Development Core Team, Vienna, Austria).

3. Results

3.1 Search results

The search identified 748 manuscripts. The selection process consisted of two phases: (1) initial screening phase by the title and abstract to exclude irrelevant articles and this resulted in the exclusion of 674 articles and (2) full text review phase for the remaining manuscripts (74 articles) with exclusion based on appropriate reasons that resulted in the exclusion of 55 more manuscripts. Overall, our search identified 19 manuscripts describing different EPE predictive tools, accounting for a total of 44901 patients. **Supplementary Figure 2** shows a detailed analysis of the search process with reasons for exclusion.

3.2 Review results

Most of the PTs had been developed in USA [9-11, 14, 15, 17, 19, 21, 25, 26] and some of them were updates of previous versions [9, 11, 14, 19, 21, 25, 26]. The characteristics of all the PTs whether predicting pT3a or wEPE are reported in **Table 1**.

The sample size of the included studies ranged from 96 [12] to 5,730 [19, 21]. In terms of pathological staging, organ confined disease ranged from 54% [12] to 80% [16, 26]. Several predictive variables were used but only PSA level and Gleason score were considered by all the authors. **Supplementary Table 1** shows all the covariates used and the number of studies integrating them.

Despite the importance of side-specific detection of EPE, only 4 nomograms reported it [15, 17, 18, 22]. Logistic regression was the most common statistical method used for the development of those PTs (84%) [4, 9, 10, 12, 13, 15, 17-26].

In 15/19 studies (79%), the internal AUC was reported by the authors ranging from 0.420 [23] to 0.856 [12] and from 0.777 [4] to 0.840 [18] in the PTs developed to predict pT3a and wEPE, respectively.

TABLE 2: 6XPPDU\ RI WKH H[WHUQDO YDOLGDWLRQ SDWL

Number of patients	6360
Age in years (median - IQR)	62.0 (56.0 ±67.0)
BMI in Kg/m2 (median - IQR)	27.8 (25.4 ±30.5)
Race	
Caucasian	5717 (89.9%)
Black	489 (7.7%)
Other	154 (2.4%)
PSA total in ng/ml (median - IQR)	5.0 (4.0 ±7.0)
’10	5601 (88.1%)
10 ±20	625 (9.8%)
“20	128 (2%)
PSA density in ng/ml/cc (median - IQR)	0.10 (0.07 ±0.15)
Prostate volume in cc (median - IQR)	15.0 (5.0 ±20.0)
Clinical stage (digital rectal examination), n (%)	
T1	4949 (77.9%)
T2a	969 (15.3%)
T2b	249 (3.9%)
T2c	134 (2.1%)
T3-T4	52 (0.8%)
' ¶ $ P L F R F O D V V L I L F D W L R Q	
Low-risk	2,737 (43%)
Intermediate-risk	2,684 (42.2%)
High-risk	939 (14.8%)
Biopsy Gleason sum, n (%)	
5 or less	6 (0.1%)
6	3022 (47.6%)
7	2588 (40.7%)
8	485 (7.6%)
9-10	253 (4%)
Biopsy cores, n (%)	
6	701 (11%)
7-11	369 (5.8%)
12	4812 (75.7%)
13-17	283 (4.4%)
“17	195 (3.1%)
Pathological stage, n (%)	
pT2a	560 (8.8%)
pT2b	32 (0.5%)
pT2c	3965 (62.3%)
pT3a	1365 (21.5%)
pT3b	438 (6.9%)
Positive surgical margins (%)	917 (14.4%)