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 y and the predicted probabilities p: $Brier = \frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2$. 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) $Brier_{scaled} = 1 - \frac{Brier}{Brier_{max}}$ was reported in the study, where $Brier_{max} = \bar{p} \cdot (1 - \bar{p})$ and $\bar{p} =$ ave(p) 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

$$d = \frac{(p_D - p_{EV})}{\sqrt{(p_D(1-p_D) + p_{EV}(1-p_{EV}))/2}},$$

where p_{ED} and p_D denote the prevalence in EV and development populations, respectively; for continuous variables d is defined as

$$d = \frac{(\bar{x}_D - \bar{x}_{EV})}{\sqrt{(s_D^2 + s_{EV}^2)/2}},$$

where $\bar{x}_{EV}$, s_{EV}^2, $\bar{x}_D$, s_D^2 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.

3.3 External validation

The characteristics of patients in the validation cohort are summarized in **Table 2**. The analysis of the prostatectomy specimen revealed that 1,365 (21.5%) and 1,803 (28.4%) patients had pT3a and wEPE, respectively. The inclusion and exclusion criteria of each predictive tool were respected and the total number of patients used for EV for each of them is reported in **Table 1**. The degree of balance for each covariate in the validation and derivation cohorts is summarized in **Supplementary Table 2**.

As far as discrimination, when considering the event which any singular model had been developed for, the AUC at EV ranged from 0.610 to 0.801. The nomogram developed by Ohori et al [15] showed the highest AUC similar to the one reported by authors (0.801 versus 0.806, respectively). Moreover, Tsuzuki [17] (AUC 0.787), Satake [22] (AUC 0.783), Chung [4] (AUC 0.772) and Jeong [24] (AUC 0.715) were among the top five models as regards the discrimination of the event they intended to predict, even though they are in the "acceptable predictive performance" [35].

We then tested each nomogram independently from what it was originally developed for, to predict both wEPE and pT3a status. Interestingly, all the PTs showed a better discrimination for wEPE rather than for pT3a status. **Figure 1** summarizes the discriminative performance of all the models when predicting wEPE and pT3a.

As regards the calibration, **Supplementary Figures 3-23** show the curves for all the PTs except for Tsuzuki et. al. [17] because it lacks some essential data for the calculation of predicted probabilities. Most of the PTs showed poor calibration with tendency towards

overestimation. Regarding the ECI, Tosoian [26] (0.148), Naito [20] (0.184), Chung [4] (0.294), Egawa [12] (0.314), and Ohori [15] (0.405) showed the best calibration considering the event they were developed to predict.

As far as the SBS is concerned, Chung et al [4] showed the best performance (SBS = 0.204). Furthermore, Ohori [15] (SBS = 0.142), and Satake [22] (SBS = 0.100), are among the top performing PTs.

The popularity of the PTs was assessed based on its number of citation in Google Scholar per year (total number of citation/number of years) in order to give a more realistic information about their popularity (**Table 1**). The most cited PTs are Partin 1997 [11], Partin 1993 [9], and Partin 2001 [14]. The popularity of those PTs seems not to relate to their predictive performance on an external cohort.

4. Discussion

EPE prediction is crucial for surgical planning as the localization and quantification of a possible EPE could allow tailoring the surgical approach on cancer's characteristics. Although, multiparameteric Magnetic Resonance Imaging (mpMRI) has gained great acceptance as a useful diagnostic and staging tool for PCa [27], its sensitivity in predicting EPE appears to be low (0.57) [39]. Furthermore, the incremental value of adding mpMRI parameters to the currently available PTs is debatable [40, 41]. On the other hand, the standardized use of intra-operative frozen section has been proposed for NSRP; however, its role is still debatable and has not gained widespread popularity in the clinical practice yet. Furthermore, visual and tactile assessment during surgery is only partially reliable and not reproducible [17, 42, 43].

After Partin's innovative idea to create a statistical tool to predict pathological stage [9], several authors developed other PTs [4, 7, 9-26]; however, the majority lacks of appropriate EV.

Independent EV studies are uncommon, with only a 16% probability for a PT to be externally validated by different authors within 5 years of development [29]. Most of the published EV studies are based on small sample sizes and therefore not reliable; actually, an EV should include a minimum of 100 events and 100 non-events [28].

It is noteworthy that EPE is not only prone to diagnostic pitfalls and interobserver variability, but it is also characterized by heterogeneous definitions, often elusive and equivocal. In this study, we included fourteen papers considering EPE as pure pT3a disease [9-14, 16, 19-21, 23-26] whereas the remaining ones looked for the global presence of disease out of the prostate regardless SVI (wEPE) [4, 15, 17, 18, 22].

In this setting, a large cohort was used to perform the EV of all the available PTs of EPE published since 1993 considering both definitions, and all of them have been externally validated for the prediction of both events (pT3a and wEPE).

As far as methodology is concerned, AUC is commonly used due to its user-friendly output [35]. Considering the EV, Ohori's [15] PT was the only model with a discriminative performance exceeding 0.8 (AUC 0.801). This supports the results of Clement [44] who compared the performance of Ohori's [15] and Steuber's [18] PTs reporting an AUC of 0.80 and 0.78, respectively. Seven models reached an "acceptable" AUC [4, 10, 17-19, 22, 24-26], whereas the discrimination of the remaining 11 models was "poor".

The analysis also showed that the Partin Tables are the most popular PTs with the highest number of citations (327.15 citation/year for all the versions of Partin Tables) [4, 10, 17-19, 22, 24-26]. Particularly, Partin tables 1997 [11] is the most cited PT and the most commonly externally validated, with 102.43 citations/year [45-50].

The different versions of Partin Tables [9, 11, 14, 19, 21, 25, 26] showed worse discriminative performance in our EV than in the original derivation cohorts. Despite the highest popularity and the good internal AUC (0.818), the discriminative ability of Partin Table 1997 in the current EV was poor (AUC 0.675).

These findings are consistent with the ones from previous EV studies, which showed that Partin Tables' performances seem to worsen when applied to different populations. The transportability of such models to other geographical areas (Sweden, UK, France, Italy and Austria) showed poor performance [45-48]. On the contrary, some authors showed acceptable discriminative performance in German and North American patients [49, 50].

Despite the clinical importance of the PTs calibration [51], our study showed poor calibration with tendency towards overestimation of the EPE risk in most of the predictive models, consistent with other EV studies [46, 48]. In this setting, recalibration can be considered for poorly calibrated PTs (regardless their predictive performance) before their introduction into the clinical practice [51].

As far as the overall performance is concerned, all the PTs exhibited moderate performance on SBS, with Chung [4] providing the best predictive performance (SBS = 0.204) for the event it was developed to predict (wEPE).

These differences of the PTs performances between the development and the validation studies may be explained by the temporal, geographical and domain differences, which in turn may affect the measurement of variables and outcomes, the case-mix (like age and tumor characteristics) and the sample size [28].

Interestingly, when considering the different definitions of EPE, all the models had a worse discrimination capability for the prediction of pure pT3a status rather than for the prediction

of the wEPE, including those developed to predict the pure pT3a status specifically. There is no clear explanation of this, however, it can be speculated that the wEPE status is easier to predict because includes also patients with SVI, and therefore with a probably higher burden of concomitant EPE, that it is usually easier to be detected [7].

Interestingly, only four models provide a side-specific-risk of EPE, to aid surgical decision toward a unilateral vs bilateral NSRP [15, 17, 18, 22]. It is noteworthy to mention that, beyond the prediction of the presence or of the laterality of EPE, none of the PTs estimates the amount of disease out of the prostatic capsule nor provide a decision rule to grade the preservation of NVB. More recently the PRECE tool [7] was specifically developed to this purpose, providing a decision rule to grade the dissection. However, it has not been validated in this study since it had been developed using the present cohort. The PRECE tool [7] has not been externally validated yet to our knowledge.

Finally, the primary goal of the current study was to provide a broader image about the currently available PTs and their performance using a single and large cohort of PCa patients, which, might help to identify the limitations and challenges encountered in developing new PTs in the contemporary era. This EV study showed that the discriminative performance for most of the included PTs ranged from poor to acceptable discrimination with poor calibration suggesting that new clinical, pathological and radiological predictors might be integrated in the development of new PTs to improve their performance. In these settings, mpMRI could have the potentials in improving the performance of predictive tools [41, 52]. Furthermore, some authors suggested that ^{68}Gallium-prostate specific membrane antigen-positron emission tomography/CT or MRI (^{68}Ga-PSMA-PET/CT or MRI), has a good potential for preoperative prediction of EPE [53, 54], thus may be identification of ^{68}Ga-PSMA-PET/CT variables may improve the predictive performance of PTs. Moreover, Dean et al [55], demonstrated that the

quantification of Gleason pattern 4 (total length of Gleason pattern 4 across all cores) is an independent predictor of pathological adverse events after RP, suggesting that its addition to PTs may improve its predictive performance.

Strength:

- The simultaneous EV of all the PTs of EPE published in the last 25 years.
- Evaluating the models' performance using both discrimination, calibration and SBS.
- Large and heterogeneous cohort of patients as validation dataset.

Limitations:

- Our PCa cases are representative of patients mostly coming from the USA; however, it remains uncertain if they are representative for patients from different populations.
- Absence of central review of the biopsy specimen and central PSA measurement, since the cohort comes from a large referral center.
- The use of a single center cohort of patients may limit the generalizability of our results; however, it should be remarked that as a large referral center patients come from different parts of the world.

5. Conclusion

To the authors' knowledge, this is the largest cohort of patients used to externally validate the available PTs of EPE developed since 1993. However, the included PTs may have shown acceptable results based on the variables used and the time when they were developed, the current EV study raises several concerns about the limitations of almost all the existing PTs of EPE in the contemporary settings. In the era of precise and personalized medicine, surgeons have to consider these limitations when using these PTs to plan a NSRP and to

identify more reliable covariates, to improve the predictive performance of newly developed PTs; the inclusion of mpMRI variables could be a future perspective to face the concern of EPE prediction and surgical management.

TAKE HOME MESSAGE:

Despite the use of predictive models is widespread and recommended by most of the international guidelines, surgeons have to be aware about their moderate to poor predictive performance for a pT3a disease and the consequent risk of applying those decision–making tools in a population other than the development one.

REFERENCES

1. Sanda MG, Cadeddu JA, Kirkby E, et al. Clinically Localized Prostate Cancer: AUA/ASTRO/SUO Guideline. Part II: Recommended Approaches and Details of Specific Care Options. *The Journal of urology.* 2018;199:990-997.

2. 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. *European urology.* 2017;71:618-629.

3. Alemozaffar M, Regan MM, Cooperberg MR, et al. Prediction of erectile function following treatment for prostate cancer. *Jama.* 2011;306:1205-1214.

4. Chung JS, Choi HY, Song HR, et al. Preoperative nomograms for predicting extracapsular extension in Korean men with localized prostate cancer: a multi-institutional clinicopathologic study. *Journal of Korean medical science.* 2010;25:1443-1448.

5. Rabbani F, Stapleton AM, Kattan MW, Wheeler TM, Scardino PT. Factors predicting recovery of erections after radical prostatectomy. *The Journal of urology.* 2000;164:1929-1934.

6. Druskin SC, Liu JJ, Young A, et al. Prostate MRI prior to radical prostatectomy: effects on nerve sparing and pathological margin status. *Research and reports in urology.* 2017;9:55-63.

7. Patel V, Sandri M, Grasso AAC, et al. A Novel Tool for Predicting Extracapsular Extension During Graded Partial Nerve Sparing in Radical Prostatectomy. *BJU international.* 2017.

8. Sayyid R, Perlis N, Ahmad A, et al. Development and external validation of a biopsy-derived nomogram to predict risk of ipsilateral extraprostatic extension. *BJU international.* 2017;120:76-82.

9. Partin AW, Yoo J, Carter HB, et al. The use of prostate specific antigen, clinical stage and Gleason score to predict pathological stage in men with localized prostate cancer. *The Journal of urology.* 1993;150:110-114.

10. Bostwick DG, Qian J, Bergstralh E, et al. Prediction of capsular perforation and seminal vesicle invasion in prostate cancer. *The Journal of urology.* 1996;155:1361-1367.

11. Partin AW, Kattan MW, Subong EN, et al. Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer. A multi-institutional update. *Jama.* 1997;277:1445-1451.

12. Egawa S, Suyama K, Matsumoto K, et al. Improved predictability of extracapsular extension

and seminal vesicle involvement based on clinical and biopsy findings in prostate cancer in Japanese men. *Urology.* 1998;52:433-440.

13. Egawa S, Suyama K, Arai Y, et al. A study of pretreatment nomograms to predict pathological stage and biochemical recurrence after radical prostatectomy for clinically resectable prostate cancer in Japanese men. *Japanese journal of clinical oncology.* 2001;31:74-81.

14. Partin AW, Mangold LA, Lamm DM, Walsh PC, Epstein JI, Pearson JD. Contemporary update of prostate cancer staging nomograms (Partin Tables) for the new millennium. *Urology.* 2001;58:843-848.

15. Ohori M, Kattan MW, Koh H, et al. Predicting the presence and side of extracapsular extension: a nomogram for staging prostate cancer. *The Journal of urology.* 2004;171:1844-1849; discussion 1849.

16. Song C, Kang T, Ro JY, Lee MS, Kim CS, Ahn H. Nomograms for the prediction of pathologic stage of clinically localized prostate cancer in Korean men. *Journal of Korean medical science.* 2005;20:262-266.

17. Tsuzuki T, Hernandez DJ, Aydin H, Trock B, Walsh PC, Epstein JI. Prediction of extraprostatic extension in the neurovascular bundle based on prostate needle biopsy pathology, serum prostate specific antigen and digital rectal examination. *The Journal of urology.* 2005;173:450-453.

18. Steuber T, Graefen M, Haese A, et al. Validation of a nomogram for prediction of side specific extracapsular extension at radical prostatectomy. *The Journal of urology.* 2006;175:939-944; discussion 944.

19. Makarov DV, Trock BJ, Humphreys EB, et al. Updated nomogram to predict pathologic stage of prostate cancer given prostate-specific antigen level, clinical stage, and biopsy Gleason score (Partin tables) based on cases from 2000 to 2005. *Urology.* 2007;69:1095-1101.

20. Naito S, Kuroiwa K, Kinukawa N, et al. Validation of Partin tables and development of a preoperative nomogram for Japanese patients with clinically localized prostate cancer using 2005 International Society of Urological Pathology consensus on Gleason grading: data from the Clinicopathological Research Group for Localized Prostate Cancer. *The Journal of urology.* 2008;180:904-909; discussion 909-910.

21. Huang Y, Isharwal S, Haese A, et al. Prediction of patient-specific risk and percentile cohort risk of pathological stage outcome using continuous prostate-specific antigen measurement, clinical stage and biopsy Gleason score. *BJU international.* 2011;107:1562-1569.

22. Satake N, Ohori M, Yu C, et al. Development and internal validation of a nomogram predicting extracapsular extension in radical prostatectomy specimens. *International journal of urology* 2010;17:267-272.

23. Fanning DM, Yue F, Fitzpatrick JM, Watson RW. Novel predictive tools for Irish radical prostatectomy pathological outcomes: development and validation. *Irish journal of medical science.* 2010;179:187-195.

24. Jeong CW, Jeong SJ, Hong SK, et al. Nomograms to predict the pathological stage of clinically localized prostate cancer in Korean men: comparison with western predictive tools using decision curve analysis. *International journal of urology.* 2012;19:846-852.

25. Eifler JB, Feng Z, Lin BM, et al. An updated prostate cancer staging nomogram (Partin tables) based on cases from 2006 to 2011. *BJU international.* 2013;111:22-29.

26. Tosoian JJ, Chappidi M, Feng Z, et al. Prediction of pathological stage based on clinical stage, serum prostate-specific antigen, and biopsy Gleason score: Partin Tables in the contemporary era. *BJU international.* 2017;119:676-683.

27. Mehralivand S, Shih JH, Rais-Bahrami S, et al. A Magnetic Resonance Imaging-Based Prediction Model for Prostate Biopsy Risk Stratification. *JAMA oncology.* 2018.

28. Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: A review. *Journal of clinical epidemiology.*61:1085-1094.

29. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction

models is infrequent and reveals worse prognostic discrimination. *Journal of clinical epidemiology.* 2015;68:25-34.

30. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *European urology.* 2015;67:1142-1151.

31. Patel VR, Tully AS, Holmes R, Lindsay J. Robotic radical prostatectomy in the community setting--the learning curve and beyond: initial 200 cases. *The Journal of urology.* 2005;174:269-272.

32. Greene F, Page D, Fleming I, Fritz A, Bach C, Haller D. *American Joint Committee on Cancer Staging Manual. 6th ed.* New York: Springer; 2002.

33. Schatloff O, Kameh D, Giedelman C, et al. Proposal of a method to assess and report the extent of residual neurovascular tissue present in radical prostatectomy specimens. *BJU international.* 2013;112:E301-306.

34. Abaza R. The robotic surgery era and the role of laparoscopy training. *Therapeutic advances in urology.* 2009;1:161-165.

35. Hosmer JDW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression, Third Edition*: John Wiley & Sons, Inc., Hoboken, NJ, USA; 2013.

36. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass.).* 2010;21:128-138.

37. Van Hoorde K, Van Huffel S, Timmerman D, Bourne T, Van Calster B. A spline-based tool to assess and visualize the calibration of multiclass risk predictions. *Journal of Biomedical Informatics.* 2015;54:283-293.

38. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine.* 2009;28:3083-3107.

39. de Rooij M, Hamoen EH, Witjes JA, Barentsz JO, Rovers MM. Accuracy of Magnetic Resonance Imaging for Local Staging of Prostate Cancer: A Diagnostic Meta-analysis. *European urology.* 2016;70:233-245.

40. Jansen BHE, Nieuwenhuijzen JA, Oprea-Lager DE, et al. Adding multiparametric MRI to the MSKCC and Partin nomograms for primary prostate cancer: Improving local tumor staging? *Urologic oncology.* 2018.

41. Rayn KN, Bloom JB, Gold SA, et al. Added Value of Multiparametric Magnetic Resonance Imaging to Clinical Nomograms for Predicting Adverse Pathology in Prostate Cancer. *The Journal of urology.* 2018;200:1041-1047.

42. Schlomm T, Tennstedt P, Huxhold C, et al. Neurovascular structure-adjacent frozen-section examination (NeuroSAFE) increases nerve-sparing frequency and reduces positive surgical margins in open and robot-assisted laparoscopic radical prostatectomy: experience after 11,069 consecutive patients. *European urology.* 2012;62:333-340.

43. Gillitzer R, Thuroff C, Fandel T, et al. Intraoperative peripheral frozen sections do not significantly affect prognosis after nerve-sparing radical prostatectomy for prostate cancer. *BJU international.* 2011;107:755-759.

44. Clement C, Maurin C, Villeret J, et al. [Head to head comparison of two currently used nomograms predicting the risk of side specific extra capsular extension to indicate nerve sparing during radical prostatectomy for treatment of prostate cancer]. *Progres en urologie : journal de l'Association francaise d'urologie et de la Societe francaise d'urologie.* 2014;24:581-587.

45. Jaderling F, Nyberg T, Blomqvist L, Bjartell A, Steineck G, Carlsson S. Accurate prediction tools in prostate cancer require consistent assessment of included variables. *Scandinavian journal of urology.* 2016;50:260-266.

46. Turo R, Forster JA, West RM, Prescott S, Paul AB, Cross WR. Do prostate cancer nomograms

give accurate information when applied to European patients? *Scandinavian journal of urology.* 2015;49:16-24.

47. Bhojani N, Salomon L, Capitanio U, et al. External validation of the updated partin tables in a cohort of French and Italian men. *International journal of radiation oncology, biology, physics.* 2009;73:347-352.

48. Augustin H, Isbarn H, Auprich M, et al. Head to head comparison of three generations of Partin tables to predict final pathological stage in clinically localised prostate cancer. *European journal of cancer (Oxford, England : 1990).* 2010;46:2235-2241.

49. Augustin H, Eggert T, Wenske S, et al. Comparison of accuracy between the Partin tables of 1997 and 2001 to predict final pathological stage in clinically localized prostate cancer. *The Journal of urology.* 2004;171:177-181.

50. Karakiewicz PI, Bhojani N, Capitanio U, et al. External validation of the updated Partin tables in a cohort of North American men. *The Journal of urology.* 2008;180:898-902; discussion 902-893.

51. Dalton JE. Flexible recalibration of binary clinical prediction models. *Statistics in medicine.* 2013;32:282-289.

52. Martini A, Gupta A, Lewis SC, et al. Development and internal validation of a side-specific, multiparametric magnetic resonance imaging-based nomogram for the prediction of extracapsular extension of prostate cancer. *BJU international.* 2018.

53. Thalgott M, Duwel C, Rauscher I, et al. One-stop shop whole-body (68)Ga-PSMA-11 PET/MRI compared to clinical Nomograms for preoperative T- and N-Staging of High-Risk Prostate Cancer. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine.* 2018.

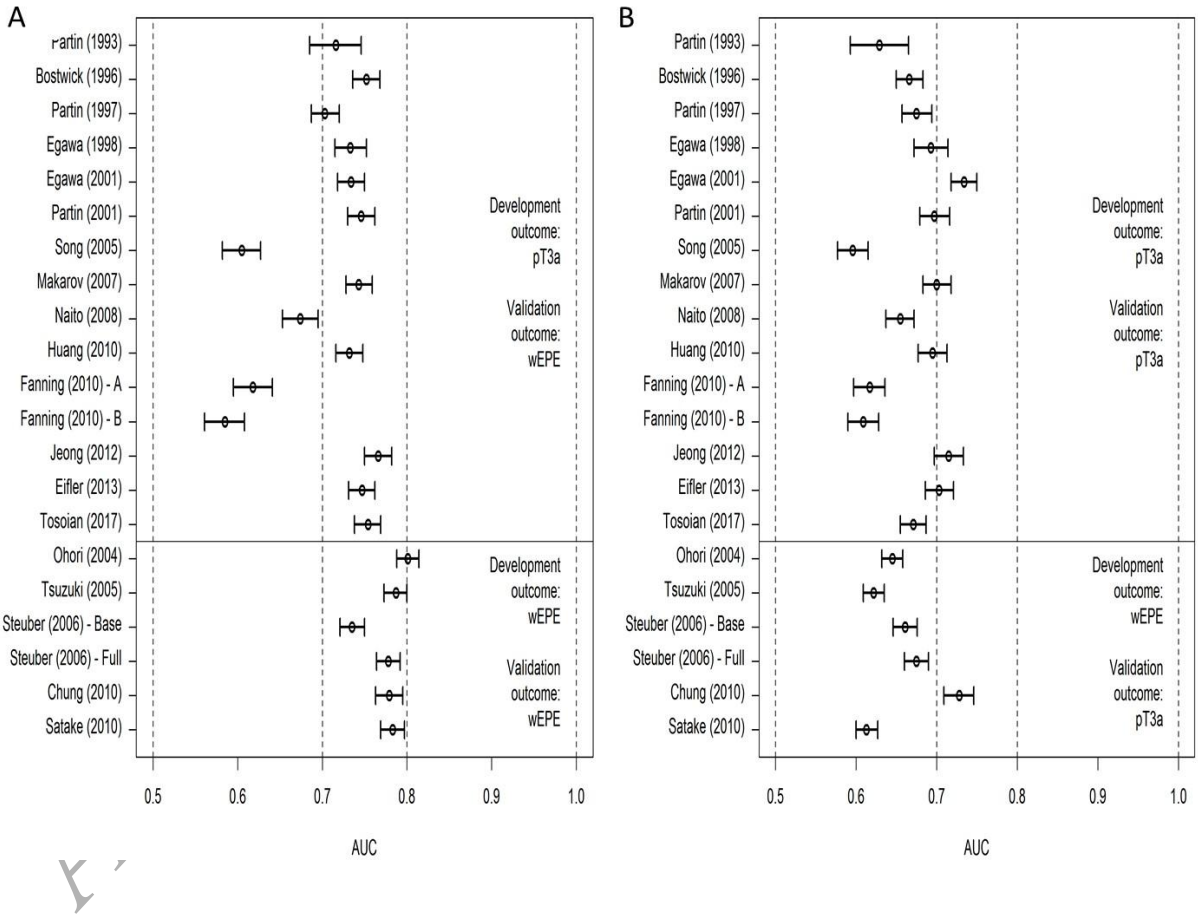
54. von Klot CJ, Merseburger AS, Boker A, et al. (68)Ga-PSMA PET/CT Imaging Predicting Intraprostatic Tumor Extent, Extracapsular Extension and Seminal Vesicle Invasion Prior to Radical Prostatectomy in Patients with Prostate Cancer. *Nuclear medicine and molecular imaging.* 2017;51:314-322.

55. Dean LW, Assel M, Sjoberg DD, et al. Clinical Usefulness of Total Length of Gleason Pattern 4 on Biopsy in Men with Grade Group 2 Prostate Cancer. *The Journal of urology.* 2019;201:77-82.

Legends:

Figure 1: (a) Summary of AUC for all the PTs when predicting whole EPE

(b) Summary of AUC for all the PTs when predicting pT3a



Supplementary material:

Supplementary Table 1: Summary of predictive parameters used and the number of authors integrating them.

Supplementary Table 2: Comparison of variables' distributions between the population used for external validation and the populations used to build predictive models. P is the p-value of the test used for comparing variables in the two populations: t-test for continuous variables (age, PSA, biopsy core positivity, maximum percent of tumor), two-sample test of proportions for binary variables (pT3a, EPE), Pearson's chi-squared test for categorical outcomes (clinical stage, Gleason score). Standardized differences d measures the degree of variables' imbalance in the two populations.

Supplementary Figure 1: Whole mount section of a radical prostatectomy specimen with the seminal vesicles. The black arrow indicates the rout of extraprostatic extension (pT3a). The red curve arrow indicates the involvement of the seminal vesicle via EPE.

Supplementary Figures 2: Detailed summary of literature search

Supplementary Figures 3-23: Assessing the discrimination and calibration of all the analyzed predictive models; the ROC curve (a) and the calibration plot (b) for the prediction of EPE independently from seminal vesicle involvement; the ROC curve (c) and the calibration plot (d) for the prediction of pT3a patients.

TABLE 1: Summary of the characteristics of the included predictive tools with the results of the external validation (AUC and Scaled Brier Score)

Reference	No. of pts	Series Characteristics	Statistical model and covariates	No. of pts / lobes used for EV	AUC reported by authors	EPE		pT3a		C/Y
						Externally validated AUC	Scaled Brier score	Externally validated AUC	Scaled Brier score	
pT3a										
Partin tables 1993 Partin AW, et al. [9] J Urol. 1993; 150: 110-114	703	**Country: USA; Outcome: pT3a (29%)** **cT (%):** T1a (4), T1b (10), T2a (46), T2b (26), T2c (9), T3a (5) ***Biopsy GS (%):*** 2-4 (9), 5 (24), 6 (43), 7 (19), 8-10 (5). **Serum PSA (%):** 0-4.0 (40), 4.0-10 (35), 10-20 (17), 20-30 (4), 30-40 (2), 40-50 (1), greater than 50 (1).	**Overall LR** - PSA - cT - Biopsy GS	*1,195 pts*	NA	0.716 (0.684 – 0.746)	0.044	0.629 (0.591 – 0.665)	-0.316	51.56
Bostwick DG, et al.[10] J Urol. 1996; 155: 1361-1367	314	**Country: USA; Outcome: pT3a (33%)** **Age (mean, yrs):** 64.9 **PSA (median, ng/ml):** 9.0 cT (%): T1c (12.4), T2a/b (32.2), T2c (55.4) **Biopsy GS (%):** 3-4 (13.4), 5-6 (43.9), 7 (31.5), 8-10 (11.1) **Pct of cancer in biopsy specimen (%):** ≤10 (36.6), 20-40 (35.7), 50-60 (16.2), ≥70 (11.5)	**Overall LR** - PSA - Pct. of cancer on biopsy	*6,354 pts*	0.770	0.752 (0.736 – 0.768)	0.120	0.666 (0.682 – 0.719)	-0.088	7.77

Partin tables 1997 Partin AW, et al. [11] JAMA 1997; 277: 1445-1451	4,133	**Country: USA; Outcome: pT3a (40%)** **PSA (%):** 0-4.0 (22.8), 4.1-10 (48.5), 10.1-20.0 (20.7), 20.1-30 (4.7), 30.1-40 (1.7), 40.1-50 (0.8), >50 (0.8) **cT (%):** T1a (1.8), T1b (3.6), T1c (32.9), T2a (28.7), T2b (20.6), T2c (9.6), T3a (2.8) **Biopsy GS (%):** 2-4 (5.4), 5 (16.6), 6 (50.7), 7 (21.9), 8-10 (5.4)	**Overall MLLR** - PSA - cT - Biopsy GS	*5,716 pts*	0.818	0.703 (0.686 – 0.720)	-0.036	0.675 (0.655 – 0.695)	-0.197	102.43
Egawa S, et al. [12] Urology 1998; 52: 433-440	96	**Country: Japan; Outcome: pT3a (46%)** **Age (mean, yrs):** 63.9 **PSA (mean, ng/ml):** 15.2 **cT (%):** T1c (36.5), T2a (10.4), T2b (27.1), T2c (10.4), T3 (15.6) **Biopsy GS (%):** 2-4 (15.6), 5-6 (26.0), 7 (27.1), 8-10 (31.2) **No. of cores with cancer:** 1 (26.0), 2(15.6), 3 (27.1), >3 (31.2) **Max cancer length (mm):** ≤ 3 (30.2), 3.1-10 (37.5), >10 (32.3)	**Overall LR** - PSA - cT - Biopsy GS - No. of cores with cancer - Max cancer length	*4,596 pts*	*3-variable model* 0.802 *4-variable model* 0.856	*3-variable model* 0.725 (0.707 – 0.743) *4-variable model* 0.733 (0.715 - 0.752)	*4-variable model* 0.093	*3-variable model* 0.671 (0.650 – 0.692) *4-variable model* 0.693 (0.671 - 0.714)	*4-variable model* 0.042	2.95
Egawa S, et al. [13] Jpn J Clin Oncol. 2001; 31:74-81	178	**Country: Japan; Outcome: pT3a (42%)** **Age (median, yrs):** 65 **PSA (%):** ≤ 4.0 (14), 4.1-10 (49.4), 10.1-20 (20.2), ≥ 20.1 (16.3) **cT (%):** T1c (51.7), T2a (12.3), T2b (19.7), T2c (7.3), T3a (1.7), T3b (1.7), T3c (5.6) **Biopsy core number (%):** > 6 (3.4), 6 (54.5), >6-12 (42.1) **Biopsy GS (%):** 2-4 (11.8), 5 (13.5), 6	**Overall LR** - PSA - cT - Biopsy GS	*5,726 pts*	NA	0.734 (0.716 – 0.751)	0.113	0.684 (0.664 – 0.704)	-0.034	2.47

		(18.5), 7 (34.8), 8-10 (21.3)							
Partin tables 2001 Partin AW, et al. [14] Urology 2001; 58: 843-848	5,079	**Country: USA; Outcome: pT3a (30%)** **Age (mean, yrs):** 57.9 **Race (%):** White (90), African-American (6), Other (4) **PSA (%):** 0-2.5 (7), 2.6-4.0 (10), 4.1-6.0 (27), 6.1-10 (35), >10 (21) **cT (%):** T1c (63), T2a (32), T2b (11), T2c (3) **Biopsy GS (%):** 2-4 (0.6), 4-6 (79.0), 3+4 (13.0), 4+3 (4.4), 8-10 (3.0)	**Overall MLLR** - PSA - cT - Biopsy GS	*5,689 pts*	NA	0.746 (0.729 – 0.762)	0.077	0.697 (0.678 – 0.716)	-0.083	70.29
Song C, et al. [16] J Korean Med Sci. 2005; 20: 262-266	317	**Country: Korea; Outcome: pT3a (20%)** **Age (mean, yrs):** 64.3 **PSA (%):** 0-4 (5.4), 4.1-10 (39.4), 10.1-20 (32.2), >20 (23). **cT (%):** T1a, b (5.4), T1c (35.6), T2a (44.5), T2b (13.2), T3 (1.3). **GS (%):** 2-4 (7.9), 5-6 (31.9), 7 (30.9), 8-10 (29.3).	**Overall MLLR** - PSA - cT - Biopsy GS	*5,573 pts*	0.626	0.605 (0.576 – 0.632)	0.006	0.596 (0.575 – 0.618)	-0.021	1.54
Partin tables 2007 Makarov DV, et al.[19] Urology 2007; 69:1095-1101	5,730	**Country: USA; Outcome: pT3a (22%)** **Age (mean, yrs):** 57.4 **Race (%):** White (88.8), African-American (6.5), Other (4.7) **PSA (%):** 0-2.5 (7.9), 2.6-4.0 (16.5), 4.1-6.0 (34.8), 6.1-8 (19.1),8.1-10.0 (10.1), >10 (11.6) **cT (%):** T1c (77.1), T2a (17.4), T2b (4.9), T2c (0.6)	**Overall LR** - PSA - cT - Biopsy GS	*5,675 pts*	0.696	0.743 (0.726 – 0.759)	0.108	0.700 (0.681 – 0.719)	0.022	45.27

		Biopsy GS (%): 5-6 (64.5), 3+4 (22.7), 4+3 (7.4), 8-10 (5.4)							
Naito S, et al. [20] J Urol. 2008; 180: 904-909	1,188	**Country: Japan; Outcome: pT3a (26%)** **Age (median, yrs):** 66 **PSA (%):** 2.5 or less (2), 2.6 – 4 (5), 4.1 – 6 (24), 6.1 – 10 (32), 10.1 or greater (37) **cT (%):** T1c (70), T2a (17), T2b (7), T2c (6) **Biopsy GS (%):** 6 or less (33), 7 (3+4) (29), 7 (4+3) (19), 8 or greater (18)	**Overall MLR** - PSA - cT - Biopsy GS	*5,861 pts*	NA	0.674 (0.651-0.696)	0.002	0.655 (0.636-0.673)	0.029	6.4
Partin tables 2010 Huang Y, et al. [21] BJU Int. 2010; 107: 1562 – 1569	5,730	**Country: USA; Outcome: pT3a (22%)** **Age (mean, yrs):** 57.4 **Race (%):** White (88.8), African-American (6.5), Other (4.7) **PSA (%):** 0-2.5 (7.9), 2.6-4.0 (16.5), 4.1-6.0 (34.8), 6.1-8 (19.1),8.1-10.0 (10.1), >10 (11.6) **cT (%):** T1c (77.1), T2a (17.4), T2b (4.9), T2c (0.6) **Biopsy GS (%):** 5-6 (64.5), 3+4 (22.7), 4+3 (7.4), 8-10 (5.4)	**Overall MLR** - PSA - cT - Biopsy GS	*5,675 pts*	0.673	0.732 (0.716 – 0.748)	0.092	0.695 (0.677 – 0.713)	-0.005	4
Fanning DM, et al.[23] Ir J Med Sci. 2010; 179:187-195	Group A: 169 Group B: 253	**Country: Ireland; Outcome: pT3a (Group A 27%, Group B 21%)** **Age (mean, yrs):** 61.0 **PSA (mean, ng/ml):** 8.0 **cT (%):** T1c (70), T2 (30) **Biopsy GS (%, group A and B):** 5-6 (66.9; 65.2), 3+4 (20.1; 20.5), 4+3 (8.9; 9.1), 8-10 (4.1; 5.2)	**Overall LR** *Group A Model* - PSA - cT - Biopsy GS *Group B Model*	*5,861 pts*	*Group A Model* 0.42 (0.34-0.50) *Group B Model* 0.42	*Group A Model* 0.618 (0.594-0.642) *Group B Model* 0.585	*Group A Model* -0.213 *Group B Model* -0.106	*Group A Model* 0.617 (0.597-0.636) *Group B Model* 0.609	*Group A Model* -0.031 *Group B Model* 0.003	0.25

			- PSA - Biopsy GS		(0.35-0.50)	(0.560-0.609)		(0.589-0.628)		
Jeong CW, et al. [24] Int J Urol. 2012; 19: 846-852	2,000	**Country: Korea; Outcome: pT3a (35%)** **Age (mean, yrs):** 65.3 **PSA (%):** ≤4 (10.5), 4.1-10.0 (52.2), >10 (37.3) **cT (%):** T1c (59.7), T2a (34.6), T2b/c (4.5), T3a (1.3) **Biopsy GS (%):** ≤6 (26.6), 3+4 (41.1), 4+3 (18.9), 8-10 (13.6)	**Overall LR** - PSA - cT - Biopsy GS - Pct. pos. cores	*5,706 pts*	0.804	0.766 (0.750 – 0.782)	0.092	0.715 (0.696 – 0.733)	-0.174	3.17
Partin tables 2013 Eifler JB, et al. [25] BJU Int. 2013; 111:22-29	5,629	**Country: USA; Outcome: pT3a (23%)** **Age (median, yrs):** 59 **Race (%):** White (81.3), African-American (11.8), Other (6.8) **PSA (%):** 0-2.5 (9.2), 2.6-4.0 (20.6), 4.1-6.0 (40.3), 6.1-10 (21.7), >10 (8.3) **cT (%):** T1c (77.8), T2a (15.9), T2b (5.5), T2c (0.8) **Biopsy GS (%):** 6 (62.9), 3+4 (22.5), 4+3 (8.8), 8 (3.9), 9-10 (2.0)	**Overall MLR** - PSA - cT - Biopsy GS	*5,689 pts*	0.702	0.747 (0.730 – 0.762)	0.109	0.703 (0.685 – 0.721)	0.018	44.6
Partin tables 2017 Tosoian JJ, et al. [26] BJU Int. 2017; 119: 676-683	4,459	**Country: USA; Outcome: pT3a (20%)** **Age (median, yrs):** 60 **Race (%):** White (79), African American (13), Hispanic (2), Asian (1), Others (5) **PSA (%):** 0-2.5 (17), 2.6-4.0 (15), 4.1-6.0 (36), 6.1-10.0 (23), more than 10 (9) **cT (%):** T1c (79), T2a (15), T2b (5), T2c (1) **Biopsy GS (%):** ≤6 (47), 3+4 (30), 4+3	**Overall MLR** - PSA - cT - Biopsy GS	*6,295 pts*	EPE vs OC 0.724 Focal EPE vs OC 0.673 Non-focal	0.754 (0.737 – 0.770)	0.124	0.671 (0.687 – 0.724)	0.061	9

		(13), 8 (6), 9-10 (4).			EPE vs OC 0.771				
EPE									
Ohori M, et al. [15] J Urol. 2004; 171: 1844-1849	Pts: 763 Lobes: 1526	**Country: USA; Outcome: EPE (30%)** **Age (median, yrs):** 61.0 **PSA (median, ng/ml):** 6.7 **cT (%):** T1c (50.5), T2a (22.9), T2b (16.8), T2c (7.6), T3a (2.2) **Biopsy GS (%):** <5 (2.1), 5 (10.2), 6 (54.5), 3+4 (19.7), 4+3 (7.6), >7 (5.9)	**Side specific LR** - PSA - cT - Biopsy GS - Pct. pos. cores - Pct of cancer in cores	*11,794 lobes*	0.806 (0.774 – 0.837)	0.801 (0.788 – 0.814)	0.142	0.645 (0.632 – 0.658)	-0.055	18.36
Tsuzuki T, et al. [17] J Urol 2005; 173: 450-453	Pts: 2,660 Lobes: 3,006	**Country: USA; Outcome: EPE (31%)** **Age (mean, yrs, OC vs EPEb vs EPEe):** 57.4, 58.6, 58.0 **PSA (mean, ng/ml, OC vs EPEb vs EPEe):** 6.0, 8.4, 8.5 **DRE (%, OC vs EPEb vs EPEe):** 22.6, 49.5, 31.3 **Biopsy GS (%, OC vs EPEb vs EPEe):** <7 (77.4; 32.7; 41.8), 3+4 (17.2; 40.2; 37.3), 4+3 (3.6; 15.5; 13.1), >7	**Side specific LR** - PSA - Biopsy GS - DRE - Average Pct of biopsy core	*11,794 lobes*	0.78	0.787 (0.769 – 0.804)	NA	0.622 (0.607 – 0.637)	NA	5.54

		(1.7; 11.6; 7.9) **Pct. of biopsy core with tumor (mean, OC vs EPEb vs EPEe):** 25.3, 41.1, 32.5 **Pct. of side specific cores with tumor (mean):** 40.2, 58.0, 44.2	with tumor - Pct of cores with tumor						
Steuber T, et al. [18] J Urol. 2006; 175: 939-944	Pts: 1,118 Lobes: 2,236	**Country: Germany; Outcome: EPE (27%)** **PSA (median, ng/ml):** 6.6 **cT (%):** T1c (82.1), T2a (10.1), T2b (4.9), T2c (2.4), T3 (0.5) **Biopsy GS (%):** 0 (39.0), 4-5 (1.9), 6 (42.2), 3+4 (11.7), 4+3 (3.9), 8-9 (1.3) **Pct. positive cores/lobe (median, %):** 33.3 **Pct. of cancer/lobe (median, %):** 3.5	**Side specific LR** *Base model* PSA; cT; biopsy GS *Full model* PSA; cT; biopsy GS; Pct. pos. cores; pct cancer in pos. cores	*8,137 lobes*	Base model 0.831 Full model 0.840	*Base model* 0.735 (0.721 – 0.750) *Full model* 0.778 (0.764 – 0.792)	*Base model* -0.173 *Full Model* -0.046	*Base model* 0.661 (0.646 – 0.676) *Full model* 0.675 (0.660 – 0.690)	*Base model* -0.170 *Full Model* -0.114	12.34
Chung JS, et al. [4] J Korean Med Sci. 2010; 25: 1443-1448	1,031	**Country: Korea; Outcome: EPE (30%)** **Age (mean, yrs):** 65.8 (EPE), 64.2(No EPE) **PSA (mean, ng/ml):** 17.1 (EPE), 9.3 (No EPE) **Prostate volume (mean, ml):** 34.7 (EPE), 37.5 (No EPE) **Biopsy GS (%):** <7 (16.3 EPE, 83.7 No EPE), 7 (35.7 EPE, 64.3 No EPE), >7 (64.0 EPE, 36.0 No EPE) **No. of positive cores (mean):** 4.5	**Overall LR** - Age - PSA - Biopsy GS - Pos. core ratio - Max pct. of cancer - PSA density	*5712 pts*	0.777 (0.762-0.803)	0.772 (0.762-0.795)	0.204	0.722 (0.709-0.746)	0.004	1.5

		(EPE), 2.6 (no EPE) **Positive core ratio (mean):** 0.49 (EPE), 0.28 (No EPE) **Max pct. of tumor in any core (mean):** 61.2 (EPE), 31.8 (No EPE)							
Satake N, et al. [22] Int J Urol. 2010; 17: 267-272	Pts: 354 Lobes: 708	**Country: Japan; Outcome: EPE (40%)** **Age (median, yrs):** 68 **PSA (median, ng/ml):** 7.4 **cT (%):** T1c (71.8), T2a (15.3), T2b (9.0), T2c (3.1) T3a (0.8) **Biopsy GS (%):** <6 (3.7), 6 (28.2), 3+4 (20.6), 4+3 (23.2), 8-10 (24.3) **Pct of positive cores/lobe (%):** 0 (34.6), 1-34 (34.2), 34.1-67 (16.0), 67.1-100 (15.3) **Max pct of cancer/lobe (%):** 0 (34.6), 1-34 (40.0), 34.1-67 (14.4), 67.1-100 (11.0)	**Side specific LR** - PSA - cT - Biopsy GS - Max pct of cancer	*11,814 lobes*	0.797	0.783 (0.768 – 0.796)	0.100	0.613 (0.600 – 0.627)	-0.040	2.63

Abbreviations: pts, patients; NA, not available; OC, organ confined; NVB, neurovascular bundle; EPE, extraprostatic extension; EPEb, extraprostatic extension in NVB; EPEe, extraprostatic extension elsewhere (no EPE in NVB); GS, Gleason Score; cT, clinical stage; PSM, positive surgical margin; PSAD, PSA density; LR, logistic regression; MLR, multinomial logistic regression; MLLR, multinomial log-linear regression; CART, Classification And Regression Tree; pos., positive; max, maximum; pct, percent: EV external validation, C/Y; citations per year

TABLE 2: Summary of the external validation patients' characteristics

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%)
D'Amico classification	
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%)