Weighted Gleason scores do not outperform standard clinical Gleason scores

Antonio Cusano, BS,[1] Peter Haddock, PhD,[1] Max Jackson, BA,[1] Ilene Staff, PhD,[2] Joseph R. Wagner, MD[1]

[1]Urology Division, Hartford Healthcare Medical Group, Hartford Hospital, Hartford, Connecticut, USA

[2]Research Program, Hartford Hospital, Hartford, Connecticut, USA

CUSANO A, HADDOCK P, JACKSON M, STAFF I, WAGNER JR. Weighted Gleason scores do not outperform standard clinical Gleason scores. *Can J Urol* 2015;22(2):7709-7714.

Introduction: *Predicting patient survival rates following radical prostatectomy remains an area of clinical interest. We compared the ability of standard clinical Gleason scores and alternative 'weighted' Gleason scores to predict pathology, margin status and recurrence in prostate cancer.*

Materials and methods: *Patients who underwent robotic radical prostatectomy performed by a single surgeon between Jan 2007 - Dec 2008 were included. Tumor at the inked margin in pathologic samples was considered a positive margin. Recurrence was defined as PSA ≥ 0.2 or the institution of salvage therapy. Standard pathologic Gleason scores were recorded. The proportion of tumor in each core was used to calculate 'weighted' and 'rounded weighted ' Gleason scores. The ability of each Gleason score to predict pathology, margin status and recurrence were statistically compared.*

Results: *Of 433 cases, 281 with uniform Gleason 6 cores were excluded. One hundred and fifty-two cases had Gleason scores ≥ 7, of which complete data were unavailable for three patients. In the final cohort of 149 cases, 72 (48.3%) patients had uniformly scored biopsies, while 77 (51.7%) had biopsies with non-uniform Gleason scores. The positive margin rate and recurrence free rates were 30.2% and 77.2%, respectively.*

Analyses of the entire patient cohort, and patients with non-uniform cores, found no significant difference between the predictive capacities of each scoring system. The alternative algorithms were not shown to be better predictors of pathologic Gleason score, margin status or recurrence.

Conclusions: *Using the highest standard Gleason score of all cores to define a preoperative Gleason score remains an appropriate clinical practice.*

Key Words: Gleason score, pathologic, margin, recurrence, algorithm

Introduction

The Gleason scoring system was developed at the Minneapolis Veterans Affairs Hospital in the 1960s to clinically grade prostatic carcinoma in biopsies and prostatectomy specimens.[1,2] In 2005, the International Society of Urological Pathology endorsed a modified Gleason algorithm that has continued to be the standard scoring system in prostate pathology.[3] Gleason scores are frequently used to predict biological cancer behavior, assess patient prognosis, and guide the selection of the most appropriate treatment option. As such, the reliability of Gleason scoring in predicting the disease status of prostate pathology specimens is critical in reducing the incidence of both under and over treatment, and predicting a patient's prognosis.

Patients who present with prostate cancer are typically risk-stratified utilizing a combination of Gleason score on diagnostic biopsy, serum prostate-specific antigen (PSA) levels, and digital rectal examination. The accuracy and reliability of clinical Gleason scoring in predicting patient survival rates following radical prostatectomy remains an area of clinical interest. In patients with organ-confined disease, some studies have shown with multivariable analyses that pathologic Gleason score is the only useful prognostic factor for disease recurrence.[4] Discrepancies between clinical Gleason scores after biopsy and pathologic Gleason scores following radical prostatectomy have provided impetus to determine preoperative indices that account for these scoring disparities, with the goal of helping to better

Accepted for publication February 2015

Address correspondence to Dr. Peter Haddock, Hartford Healthcare Medical Group, Urology Division, Hartford Hospital, 85 Seymour Street, Suite 416, Hartford, CT 06106 USA

select patient treatment. Digital rectal examination findings, PSA, as well as the number of total biopsy and positive biopsy cores have each been viewed as potential predictive factors.[5-7]

In the present study, we derived two alternate derivatives of the standard Gleason score: (i) a weighted Gleason score and (ii) a rounded weighted Gleason score. In these algorithms, the percent tumor involvement for each core was used to weight the Gleason score of each individual core to derive a weighted Gleason score. We subsequently statistically compared the ability of the standard Gleason scoring system with both of the alternate Gleason scores to predict pathology, margin status and biochemical recurrence.

Materials and methods

We performed a retrospective review of our prospectively maintained Institutional Review Board (IRB)-approved radical prostatectomy database to identify consecutive robotic-assisted laparoscopic radical prostatectomies (RALRP) performed by a single surgeon between January 1, 2007 and December 31, 2008.

We subsequently identified cases with biopsy cores uniformly scored as Gleason 6. The remaining cases scored Gleason ≥ 7 formed the study cohort. Clinical data for these patients were extracted from our prostate database, inpatient hospital electronic medical records, and outpatient electronic/paper physician-recorded medical charts.

We undertook a power analysis to determine the sample size needed for comparison of the correlations of the standard and alternative calculations of clinical Gleason with surgical margins, recurrence and pathologic Gleason. This was conducted using data from patients who underwent a robotic prostatectomy at our clinical center over a period of 11 years (a total of 2006 patients). A total of 1050 patients with clinical Gleason ≥ 7 were selected (Gleason 6 patients were excluded since uniform cores are unaffected by our alternative Gleason calculations and would not be included in the analysis). After calculating the correlations of standard clinical Gleason with pathological Gleason, margins and recurrence, we determined the sample size needed to detect different effect sizes. For biochemical recurrence, the most clinically relevant outcome measure in our study, the correlation with standard Gleason was approximately 0.15. The power calculation determined that 629 patients would be needed to achieve 80% power ($p = 0.05$) for an effect size of doubling of r to 0.30. Notably, the power analysis also illustrated that thousands of patients would be required to resolve smaller differences.

Given the resource-intensive nature of retrieving the detailed information required to calculate our alternative Gleason algorithms (Gleason score for each core and % core involvement), we decided to initially obtain patient data on a subset of patients, enabling an analysis to be performed to test our hypotheses and obtain data for parameter estimates for any future power calculations. A total of 433 patients underwent robotic prostatectomy by a single surgeon between January 1, 2007 and December 31, 2008. One hundred and fifty-two patients with Gleason ≥ 7 disease were selected. 3 patients had incomplete data and 72 had uniform cores. The remaining 77 patients had non-uniform cores.

We utilized data during a 2 year period from patients for whom we also had long term (5 years) follow up information. We believe that this provided a patient cohort that is representative of the patient population treated at our clinical center.

Demographic information, clinical staging, pathological findings, operative details, and prostate-specific antigen (PSA) values were recorded prospectively. Surgical margin status was determined by pathologic evaluation of biopsy specimens immediately following RALRP. All specimens were whole-mounted and step-sectioned at 3 mm intervals with the apex and base being additionally cross-sectioned at right angles. Positive surgical margin was defined by the existence of cancer cells at the inked resection margin in the final specimen. Postoperative PSA values were obtained at 1, 3, 6, 9, 12, 18, and 24 months, and annually thereafter. Recurrence was defined as PSA ≥ 0.2 ng/mL. Patients receiving salvage therapy were also considered to have recurred at the time of treatment and included in our analysis.

While the number of biopsy cores has steadily increased at our institution over the past 10 years (in line with national trends), it was not until 2009 that 12 core TRUS biopsies became standard practice at our institution. During the study period (2007-2008), it was our standard practice to acquire 10 core biopsies. Furthermore, while all patients included in the study had their biopsy results analyzed by Hartford Hospital pathologists, not all patients had their biopsy procedure at our institute. As such, in some patients either more or less than 10 core biopsies were obtained. The median number of biopsy cores was 10. A total of 127 (85.2%) patients underwent 10 - ≥ 12 core biopsies.

In evaluating preoperative biopsy results, the percent tumor involvement for each core was used to weigh each individual core's Gleason toward a weighted Gleason score. The algorithm used to

calculate the standard, weighted and rounded-weighted Gleason sum scores is illustrated in Figure 1a. A similar process was used to calculate corresponding values for the Gleason 'major' score, Figure 1b. The standard, weighted, and rounded weighted clinical Gleason scores were correlated with pathologic Gleason score, positive margin rates, and recurrence rates using Spearman, point by serial, or Cramer's V correlation as appropriate for each respective level of data. Statistical significance was accepted to be $p < 0.05$. Statistical analyses were performed using SPSS v21.0 statistical software (SPSS, Inc., Chicago, IL, USA).

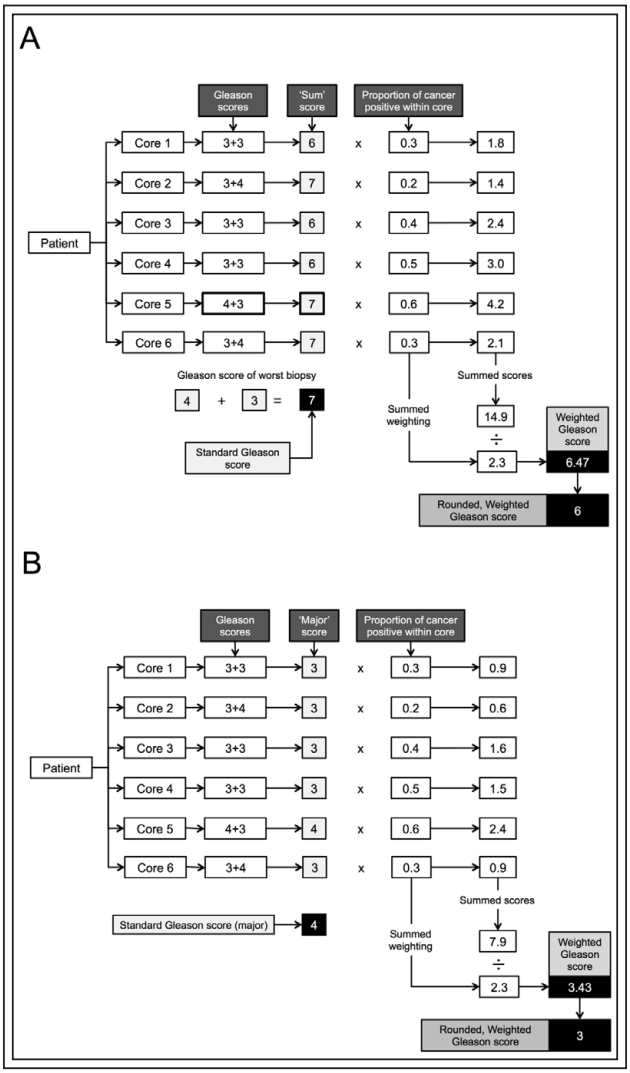

Figure 1. Algorithms used to calculate 'weighted' and 'rounded weighted' **A)** Gleason sum score and **B)** Gleason 'major' score.

Results

A total of 433 consecutive RALRPs performed by a single surgeon between January 1, 2007 and December 31, 2008 were identified and initially selected for inclusion. Of these 433 consecutive cases, 281 (64.9%) were uniformly scored as Gleason 6 and were excluded from analysis. Of the 152 remaining cases with clinical Gleason scores ≥ 7, 3 cases had incomplete data and were omitted from the study. As such, a total of 149 cases with Gleason scores ≥ 7 formed the final study cohort. A total of 6 (4%) of patients included in the study initially underwent a period of active surveillance prior to surgery. The median time that the patients were under active surveillance was 18.8 (IQR: 13.9-30.5) months. In this analysis, the biopsy just prior to surgery was utilized. The median time between biopsy and surgery was 4.3 (IQR: 3.3-5.0) months.

Of the 149 cases, 72 (48.3%) had uniform Gleason scores across all biopsy cores. In contrast, 77/149 (51.7%) of cases had non-uniform Gleason scores within individual biopsy cores. A total of 45/149 (30.2%) patients had positive margins, while 115/149 (77.2%) of patients were recurrence-free.

The median age and body mass index (BMI) of the 149 patients included in the study cohort were 61 (IQR: 56-65) years and 27.5 (IQR: 25.3-30.7) kg/m^2, respectively. Median patient follow up time was 60 (IQR: 53-64) months. At diagnosis, the majority of patients (59.1%) had PSA levels of between 4 ng/mL-10 ng/mL. At 5 years follow up, the majority of patients (81.2%) had PSA levels of < 0.2 ng/mL. Final pathology demonstrated the majority of patients (59.1%) were AJCC pathologic stage T2. Patient demographics are presented in Table 1.

We determined correlation coefficients for standard clinical Gleason score, weighted clinical Gleason score, and rounded weighted clinical Gleason score with pathologic Gleason score, positive margins, and biochemical recurrence. This was performed independently in all patients (uniform and non-uniform cores) and patients with non-uniform cores.

Comparison of Gleason scoring using non-uniform cores

Analyses were performed on the 77 patients with non-uniform cores. For both the Gleason sum and major scores, there was no significant difference in the ability of the standard Gleason, weighted Gleason or rounded-weighted Gleason scoring algorithms to predict pathology, margin status or recurrence, Table 2. However, there was a tendency for the standard Gleason score (both the sum and major scores) to better predict pathology, margin status and biochemical

TABLE 1. **Patient demographics and surgical outcomes**

Number of patients	149
Cases (n; %)	
Uniform cores	72 (48.3)
Non-uniform cores	77 (51.7)
Patient age (yrs) (median; range)	61 (56-65)
BMI (kg/m^2) (median; IQR)	27.5 (25.3-30.7)
Follow up time: (mos) (median; IQR)	60 (53-64)
PSA (ng/mL) (n; %)	
Diagnostic	
1-4	39 (26.1)
> 4-10	88 (59.1)
> 10-20	15 (10.1)
> 20	6 (4)
Unknown	1 (0.7)
Most recent	
< 0.2	121 (81.2)
0.2-< 1	17 (11.4)
> 1	9 (6)
Unknown	2 (1.3)
Gleason score (n; %)	
Clinical	
7	121 (81.2)
8	23 (15.4)
9	5 (3.4)
Pathologic	
6	7 (4.7)
7	116 (77.9)
8	11 (7.4)
9	15 (10.1)
AJCC pathologic stage (n; %)	
T2	88 (59.1)
T3	59 (39.7)
Margins (n; %)	
Negative	104 (69.8)
Positive	45 (30.2)
Biochemical recurrence (n; %)	
No	115 (77.2)
Yes	34 (22.8)

BMI = body mass index; IQR = interquartile range
PSA = prostate-specific antigen

recurrence compared to the other two alternative Gleason scoring algorithms.

Comparison of Gleason scoring using all cores (uniform and non-uniform cores)

We compared the ability of each of the scoring algorithms to predict final pathology using data from the complete cohort of patients (i.e. patients with uniform or non-uniform cores). While not statistically significant, there was a tendency for the standard Gleason score (sum) to be a better predictor of pathology compared to the weighted and rounded-weighted Gleason scores (correlation coefficient = 0.553; p = 0.54 and 0.62, respectively; Table 3. However, there was no significant difference between the ability of each of the scoring algorithms to predict final pathology, Table 3. A similar qualitative outcome was observed for both predicting margin status and biochemical recurrence, Table 3.

Similarly, the major component of the standard Gleason score tended to be the best predictor of pathology, margins and biochemical recurrence. However, there was no significant difference between major component of the standard Gleason score and the major component of either the weighted and rounded-weighted Gleason scores to predict pathology, margin status or recurrence, Table 3.

Discussion

The Gleason scoring system is widely accepted and utilized as the standardized test to risk stratify prostate cancer. Both diagnostic and pathologic Gleason scores correlate with biochemical free prostate cancer survival rates.[8] However, significant discrepancies between diagnostic and post-radical prostatectomy Gleason scores illustrate the need to understand other preoperative features that can aid in the more accurate staging of prostate cancer at an early stage in the clinical process. Inaccuracies in determining diagnostic Gleason scores can result in the inappropriate surveillance of biologically aggressive tumors, over-treatment of indolent tumors, and in general, the utilization of treatment paradigms not ideal for a particular cancer grade.[9]

Elevated positive surgical margin rates, increased capsular, seminal vesicle, and lymphatic invasion, poor cancer-specific survival and higher rates of biochemical recurrence have all been reported for patients whose clinical Gleason score is lower than the Gleason score obtained from prostate pathology samples following radical prostatectomy.[10-12] Rates of Gleason score upgrading from biopsy Gleason score 6 have been cited to be as high as 50% following radical prostatectomy.[5]

Over the years, we have frequently questioned the dilemma of placing patients with predominantly Gleason 6 scores and a minority of Gleason 7 in the same prognostic category as patients with high volume, pure Gleason 7 disease. Our patients have justifiably questioned if their prognosis is more likely to be driven

TABLE 2. **Statistical correlations in patients with non-uniform cores comparing the ability of standard clinical Gleason scores, 'weighted' scores and 'rounded weighted' scores in predicting final pathology, margins status and biochemical recurrence. (median follow up time = 59 months; n = 77)**

		Correlation coefficient				
Component of Gleason score	**Clinical index**	**Standard Gleason**	**Weighted Gleason**	**Rounded weighted Gleason**	**p (standard versus weighted)**	**p (standard versus rounded weighted)**
Sum	Pathology	0.620†	0.470†	0.499†	0.19	0.28
	Margin	0.136*	0.168§	0.172*	0.84	0.82
	Recurrence	0.206*	0.100§	0.186*	0.50	0.90
Major	Pathology	0.419†	0.408†	0.379†	0.94	0.77
	Margin	0.221*	0.171§	0.159*	0.75	0.69
	Recurrence	0.177*	0.100§	0.125*	0.62	0.67

Statistical tests used: †Spearman's Rho; *Cramers V; §Point-by-serial

by the Gleason 6 or Gleason 7 score. The paradox is equally relevant when considering patients with non-uniform Gleason 7 and 8 scores. To address this issue, we attempted to evaluate whether a weighted or rounded-weighted Gleason scoring system may offer advantages by better predicting positive margin rates and biochemical recurrence when compared to the current, standard Gleason scoring system. As shown in Figure 1, the proportion of cancer in each core was used to calculate a weighted Gleason score. We felt this would incorporate the known prognostic impact of Gleason score with the often-cited prognostic impact of number of positive cores and maximum core volume.[8,13,14] However, our data suggest no significant differences between standard, weighted, and rounded-weighted clinical Gleason scores in the ability to predict both margin status and recurrence in patients with discordant Gleason scores on diagnostic biopsy, Table 2. Although not statistically significant, the standard clinical Gleason score had the highest correlation out of the three Gleason scoring methods that we assessed. As such, our data support the continued use of the current convention of using the highest Gleason score of all cores to determine clinical staging.

Univariate and multivariate analyses have shown an association between positive surgical margins and disease-specific survival.[15,16] Similarly, biochemical recurrence correlates with disease-specific survival.[7,17,18] Therefore, we believe that positive margin rates and 5 year biochemical free survival rates are appropriate endpoints with which to measure a weighted Gleason score.

TABLE 3. **Statistical correlations in patients with uniform and non-uniform cores, comparing the ability of standard clinical Gleason scores, 'weighted' scores and 'rounded weighted' scores in predicting final pathology, margins status and biochemical recurrence. (Gleason ≥ 7; median follow up time = 60 months; n = 149)**

		Correlation coefficient				
Component of Gleason score	**Clinical index**	**Standard Gleason**	**Weighted Gleason**	**Rounded weighted Gleason**	**p (standard versus weighted)**	**p (standard versus rounded weighted)**
Sum	Pathology	0.553†	0.501	0.511†	0.54	0.62
	Margin	0.122*	0.091§	0.110*	0.79	0.92
	Recurrence	0.165*	0.093§	0.141*	0.53	0.83
Major	Pathology	0.424†	0.419†	0.406†	0.96	0.85
	Margin	0.162*	0.037§	0.115*	0.28	0.68
	Recurrence	0.146*	0.093§	0.108*	0.65	0.74

Statistical tests used: †Spearman's Rho; *Cramers V; §Point-by-serial

A limitation to our weighted Gleason algorithm is that it gives equal weight to diagnostic Gleason 3+4 and 4+3 disease. The same is true for Gleason 5+4 versus 4+5 disease. It has been clearly demonstrated that Gleason 4+3 disease is associated with a significantly worse prognosis than Gleason 3+4 disease.[8,19] In fact, there is increasing evidence the proportion of Gleason 4 and Gleason 5 disease is a better indicator of pathologic behavior than standard Gleason score.[20] We addressed this concern running analyses utilizing just the major Gleason score. An additional limitation is that while we have the percentage of cancer of each core for a certain Gleason grade (e.g. 3+4=7 or 4+5=9), the exact percent of the major and minor score was not available for most biopsies and therefore not analyzed.

Based on our observed differences in correlation between standard and weighted Gleason scoring system (which actually favor standard scoring) we calculate that > 10,000 patients would be necessary to reject the null hypothesis that there is no difference between these Gleason scoring systems. Nonetheless, despite the use of a relatively small cohort of patients, we believe that our data provide meaningful information to support the conclusion that these alternative Gleason algorithms do not provide a significant advantage over the standard Gleason scoring algorithm.

Conclusions

The standard Gleason score remains the most effective and preferred method for grading prostate cancer. In the future, biomarker assays may be of added benefit in guiding diagnosis and treatment.[21] □

References

1. Gleason DF. Classification of prostatic carcinomas. *Cancer Chemother Rep* 1966;50(3):125-128.
2. Bailar III JC, Mellinger GT, Gleason DF. Survival rates of patients with prostatic cancer, tumor stage, and differentiation—preliminary report. *Cancer Chemother Rep* 1966;50(3):129-136.
3. Epstein JI, Allsbrook WC Jr, Amin MB, et al. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason grading of prostatic carcinoma. *Am J Surg Pathol* 2005;29(9):1228-1242.
4. Bianco FJ Jr, Wood DP Jr, Cher ML, Powell IJ, Souza JW, Pontes JE. Ten-year survival after radical prostatectomy: specimen Gleason score is the predictor in organ-confided prostate cancer. *Clin Prostate Cancer* 2003;1(4):242-247.
5. Chung MS, Seung HL, Lee DH, Chung BH. Is small prostate volume a predictor of Gleason score upgrading after radical prostatectomy? *Yonsei Med J* 2013;54(4):902-906.
6. Kulkami GS, Lockwood G, Evans A et al. Clinical predictors of Gleason score upgrading: implications for patients considering watchful waiting, active surveillance, or brachytherapy. *Cancer* 2007;109(12):2432-2438.
7. Kattan MW, Wheeler TM, Scardino PT. Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer. *J Clin Oncol* 1999;17(5):1499-1507.
8. Pierorazio PM, Walsh PC, Partin AW, Epstein JI. Prognostic Gleason grade grouping: data based on the modified Gleason scoring system. *BJU Int* 2013;111(5):753-760.
9. Isariyawongse BK, Sun L, Bañez LL et al. Significant discrepancies between diagnostic and pathologic Gleason sums in prostate cancer: the predictive role of age and prostate-specific antigen. *Urology* 2008;72(4):882-886.
10. Dong F, Jones JS, Stephenson AJ et al. Prostate cancer volume at biopsy predicts clinically significant upgrading. *J Urol* 2008;179(3): 896-900.
11. Müntener M, Epstein JI, Hernandez DJ, et al. Prognostic significance of Gleason score discrepancies between needle biopsy and radical prostatectomy. *Eur Urol* 2008;53(4):767-775.
12. Moussa AS, Li J, Soriano M, Klein EA, Dong F, Jones JS. Prostate biopsy clinical and pathological variables that predict significant grading changes in patients with intermediate and high grade prostate cancer. *BJU Int* 2009;103(1):43-48.
13. Humphrey PA. Tumor amount in prostatic tissues in relation to patient outcome and management. *Am J Clin Pathol* 2009;131(1):7-10.
14. Shinohara K. Improving cancer detection by prostate biopsy: the role of core number and site. *Nat Clin Pract Urol* 2006;3(10):526-527.
15. Swindle P, Eastham JA, Ohori M, et al. Do margins matter? The prognostic significance of positive surgical margins in radical prostatectomy specimens. *J Urol* 2008;179(5 Suppl):S47-S51.
16. Pfitzenmaier J, Pahernik S, Tremmel T et al. Positive surgical margins after radical prostatectomy: do they have an impact on biochemical or clinical progression? *BJU Int* 2008;102(10):1413-1418.
17. Kotb AF, Elabbady AA. Prognostic factors for the development of biochemical recurrence after radical prostatectomy. *Prostate Cancer* 2011;2011:485189.
18. Freedland SJ, Terris MK, Csathy GS et al. Preoperative model for predicting prostate specific antigen recurrence after radical prostatectomy using percent of biopsy tissue with cancer, biopsy Gleason grade and serum prostate specific antigen. *J Urol* 2004;171(6 Pt 1):2215-2220.
19. Eifler JB, Feng Z, Lin BM, et al. An updated prostate cancer staging nomogram (Partin tables) based on cases from 2006 to 2011. *BJU Int* 2013;111(1):22-29.
20. Humphrey PA. Gleason grading and prognostic factors in carcinoma of the prostate. *Mod Pathol* 2004;17(3):292-306.
21. Sartori DA, Chan DW. Biomarkers in prostate cancer: what's new? *Curr Opin Oncol* 2014;26(3):259-264.