External validation of the novel International Society of Urological Pathology (ISUP) Gleason grading groups in a large contemporary Canadian cohort

Helen Davis Bondarenko[*1]; Marc Zanaty*[1]; Sabrina S. Harmouch[1]; Cristina Negrean[1]; Raisa S. Pompe[1,2]; Daniel Liberman[1]; Naeem Bhojani[1]; Pierre I. Karakiewicz[1]; Kevin C. Zorn[1]; Assaad El-Hakim[1,3]
[*]*Co-first authors*
[1]Division of Urology, Department of Surgery, Centre hospitalier de l'université de Montréal (CHUM), Montreal, QC, Canada; [2]Martini-Klinik Prostate Cancer Centre, University Hospital Hamburg-Eppendorf, Hamburg, Germany; [3]Division of Robotic Urology, Department of Surgery, Hôpital du Sacré Coeur de Montréal, Montreal, QC, Canada

***

Abstract

Introduction: We sought to test the discriminatory ability of the 2014 International Society of Urological Pathology (ISUP) Gleason grading groups (GGG) for predicting biochemical recurrence (BCR) after robot-assisted radical prostatectomy (RARP) in a large, contemporary, Canadian cohort.

Methods: A total of 621 patients who underwent RARP in two major Canadian centres were identified in a prospectively maintained Canadian database between 2006 and 2016. Followup endpoint was BCR. Log-rank test, univariable, and multivariable Cox regression analyses were used.

Results: Mean followup was 27.9 months. All five ISUP GGG independently predicted BCR. Statistically significant differences in BCR rates were found between GGG 2 and GGG 3 strata (p<0.001). No statistically significant differences in BCR rates were found between GGG 4 and GGG 5 strata (p=0.3). Relative to GGG 1, the GGG 2, GGG 3, GGG 4, and GGG 5 yielded a 1.10-, 3.44-, 4.18-, and 4.74-fold hazard ratio (HR) increment in BCR, respectively.

Conclusions: This population-based Canadian cohort study confirms the added discriminatory property of the novel ISUP grading, specifically for GGG 2 and GGG 3 strata. No difference, however, was observed between GGG 4 and GGG 5, likely due to the lower number of patients in these groups. As such, after external validation, the 2014 ISUP GGG appears to retain clinical prognostic significance in a Canadian population.

Introduction

Since its introduction by Donald Gleason and the Veterans Administration Cooperative Urologic Research Group in 1966, the Gleason score has been the most universally accepted grading system for prostate cancer (PCa). [1] After multiple revisions, including the 2005 ISUP consensus [2], the original Gleason score (GS) which consisted of 25 possibilities evolved to a traditional three-tiered Gleason grading (TGG). Due to the lack of granularity in the TGG strata, in 2013, Pierorazio et al. [3] introduced a novel five-tiered Gleason grading groups (GGG), suggesting better discrimination and finer definition of risk based on biochemical recurrence (BCR) outcomes. [4] More specifically, there was a distinct separation of the intermediate TGG (7), into GGG 2 (3+4) and GGG 3 (4+3), as well as the high risk TGG into GGG 4 (8) and GGG 5 (9,10). In 2014, the ISUP consensus redefined the TGG, accepting the novel, five-tiered GGG as the new standard reporting system [4, 5] Several groups, including Epstein et al. [6], have externally validated the GGG with BCR as their primary endpoint. [7-9] In 2016, Pompe et al. [4] externally validated the GGG with prostate cancer specific mortality as their end point, in patients treated with external beam radiation therapy, radical prostatectomy (RP), brachytherapy or no local treatment. In 2016, the ISUP GGG system was accepted as the new standard for grading PCa by the World Health Organization. [5]

However, to the best of our knowledge, no group has validated the novel GGG in PCa patients treated by robotic assisted radical prostatectomy (RARP) using a contemporary Canadian cohort. As such, we sought to externally validate the novel GGG in a Canadian cohort of men with prostate cancer treated with RARP.

Methods

Data source

After institutional-review board approval, data were extracted from a prospectively maintained Canadian database of patients who underwent RARP, by one of two high-volume fellowship trained surgeons in two large academic Canadian Centers, Hôpital du Sacré Coeur de Montréal and Hôpital Saint Luc du Centre Hospitalier de l'Université de Montréal, between 2006 and 2016.

Study population

621 patients diagnosed with adenocarcinoma of the prostate and treated with RARP between 2006 and 2016, were identified. BCR was defined as two consecutive PSA values of ≥ 0.2 ng/dl, or the use of salvage external beam radiation therapy and/or salvage androgen deprivation therapy. Preoperative biopsy grades, as well as post-operative pathological grades were compared to BCR post-RARP. Clinical and pathological GS were categorized either according to the TGG of 6,7, and 8–10 or according to the novel GGG: 1 (6), 2 (3+4),3 (4+3), 4 (8), and 5 (9–10).

Covariates

For the clinical Gleason Grade (CGG), covariates included age at diagnosis, BMI, clinical stage, PSA, prostate volume, and percentage of positive cores. For the pathological Gleason

Grade (PGG), covariates included age at diagnosis, BMI, presence of extracapsular extension (ECE), surgical margin (SM) status, and lymph node status.

Statistical analyses

The Kaplan–Meier method was used for BCR analyses and the log rank test to compare survival between groups. Uni- and multivariable Cox regression analyses were performed to test the impact of different GGG strata on BCR. In the clinical GGG, multivariable analyses were adjusted for preoperative PSA and clinical stage (T1, T2 or T3/4). In the pathological GGG, multivariable analyses were adjusted for preoperative PSA, surgical margin status, and pathological stage (pT2, pT3a, pT3b or pT4). Harrell's C-index was used to assess the discriminatory ability of four possible Gleason grading models: the TGG model (6, 7 and 8-10), the TGG and 2 separate groups for grade 7 model (6, 3+4, 4+3 and 8–10), the TGG and 2 separate groups for grade 8 -10 model (6 ,7, 8 and 9–10) and the novel GGG. All statistical tests were two-sided with a level of significance set at $p < 0.05$. Analyses were performed using the R software environment for statistical computing and graphics (version 3.3.0).

Results

Table 1 summarizes the perioperative baseline patient and tumor characteristics. Overall, mean age was 60.34 (95%CI: 59.83-60.85) with a mean follow-up of 27.9 months. For all 621 patients treated with RARP, the observed 3-year and 5-year BCR free survival was 86.7% and 80.7%, respectively. In Kaplan–Meier analyses, 2-year BCR-free survival rates for clinical GGG 1-5 were 98.3% (95% CI: 96.4-100), 89.3% (95% CI: 84.9-93.9), 86.9%(95% CI: 78.7-96), 80.6% (95% CI: 67.2-96.6), and 65.3% (95% CI: 44.3-96.1), respectively (figure 1). 2-year BCR free survival rates for pathological GGG 1-5 were 95% (95% CI: 88.3-99.3), 93.4% (95% CI: 92.5-97.6), 82.4% (95% CI: 70-94.6), 70.5% (95% CI: 54.6-93.4), and 61.5% (95% CI: 43.6-87.8), respectively (figure 2). In Kaplan-Meier analyses for the clinical GGG, statistically significant differences in BCR rates were observed between GGG 1 and GGG 2 ($p < 0.001$), as well as, between GGG 4 and GGG 5 ($p = 0.02$). In the pathological GGG, statistical significant differences in BCR rates were identified between GGG 2 and GGG 3 ($p < 0.001$). Furthermore, multivariable Cox-regression analyses (Table 2), adjusted for preoperative PSA, surgical margin status, and pathological stage (pT2, pT3a, pT3b or pT4) revealed a three-fold higher hazard ratio (HR) for GGG 3 compared to GGG 2 (HR=2.99 CI 1.42-6.29 , $p < 0.001$). However, no difference was found between HR for GGG 4 and GGG 5 (HR=1.28 CI 0.52-3.11, $p = 0.92$).

      The univariate analyses Harrell's C-index showed a higher discriminatory ability of the novel GGG when compared to the TGG in both clinical and pathological GGG (Table 3). Similarly, the novel GGG appears to have a stronger discriminatory ability in the multivariable analyses for the pathological GGG. However, in the clinical GGG, the multivariable analyses demonstrated a better discriminatory ability in theTGG and 2 separate groups for grade 8 -10 model (6 vs 7 vs 8 vs ≥9) rather than the novel GGG model. Moreover, a C-index increase of 0.002 from the TGG model to the TGG and 2 separate groups for grade 8-10 was noted in univariate analysis. Similarly, a small C-index increase of 0.005 was observed in multivariate analysis of the Pathological Gleason Grade classification

for these groups. The latter may be due to the low prevalence of GS 9–10, constituting only 2.7% of the study cohort.

Discussion

Since its introduction in 1966, the GS has been the strongest predictor of cancer outcomes in prostate cancer. [1] After multiple revisions, including the 2005 ISUP consensus, [2] the original GS evolved to a traditional three-tiered Gleason grading (TGG). [10] Due to the heterogeneity of the TGG strata, specifically, the lack of distinction between GS 7 (3+4 and 4+3) and GS (8-10), in 2013, Pierorazio et al. [3] proposed a novel five tiered GGG. After analyzing 7,869 radical prostatectomies (RP) specimens between 1982 and 2011, they concluded that the ISUP groups had a better discriminatory ability compared to the TGG. Their 5-year BCR Free Survival rates for men with ISUP 1-5 at RP was 96.6, 88.1, 69.7, 63.7, and 34.5%, respectively (p=0.001). [3] Separation of GS 7 into two prognostic groups was further warranted given the favorable prognosis of Gleason 3+4 (GGG 2, 2-year BCR Free Survival of 93.6%), whereas Gleason 4+3 carcinomas (GGG 3) behaved more similarly to GS 8 (ISUP 4).

Since 2014, the ISUP GGG has been externally validated in 9 studies, including, one multi-institutional review, [6] 2 European , 2 American population-based cohorts, [4, 8, 11, 12] as well as, in two single center cohorts. [7, 13] All studies demonstrated significant differences between the 5 GGG. Furthermore, they confirmed that better prognostication was accomplished with distinction between GS 3+4 and 4+3 instead of GS 7[4]. For example, He et al.[11] found that each Gleason grading strata approximately doubled the risk for Prostate-cancer specific mortality (PCSM) in 331320 PCa patients that underwent RP, radiotherapy or other treatments between 2006 and 2012. Similarly, in a large North American population-based cohort of 91565 patients, Pompe et al.[4] observed that the 8-year PCSM-Free Survival rates differed significantly between the five GGG. For example, the 8-year PCSM-Free Survival rates for pathological GGG 1-5 were 99.5% (95%CI: 99.4– 99.6), 99.1% (95%CI: 98.9– 99.2),97.4% (95%CI: 97– 97.9), 95.2% (95%CI: 94.5– 96), and 85.8% (95%CI: 84.4– 87.2), respectively (p<001).

To the best of our knowledge, this is the first study which aims to externally validate the novel ISUP GGG in a contemporary Canadian cohort treated with RARP. Several of our findings were noteworthy.

First, 2-year BCR-free survival rates for clinical GGG 1-5 were 98.3, 89.3, 86.9, 80.6 and 65.3% (figure 1) and 2-year BCR-free survival rates for pathological GGG 1-5 were 95, 93.4, 82.4, 70.5 and 61.5% respectively (figure 2). These findings are consistent with Pierorazio et al. [3] who noted that 2-year BCR-free survival rates for men with clinical GGG 1-5 were 97.1, 90.6, 79.9, 70.9, and 51.5%, (p < 0.001) and that 2-year BCR-free survival rates for pathological GGG 1-5 98.8, 93.6, 85.6, 73.7 and 58.5%, respectively (p< 0.001).

Second, we observed statistically significant BCR rates difference between GGG 1 and GGG 2 (p<0.001), and GGG 4 and GGG 5 (p=0.02) in the clinical GGG and between GGG 2 and GGG 3 (p<0.001) in the pathological GGG. Other reports such as Epstein et al. [6] noted that 5-year BCR-free survival rates in 20,845 North American RP and 5,501 RT

patients were 96%, 88%, 63%, 48%, and 26% for GGG 1-5, respectively. Similar results were also reported by Pierorazio et al. [3] and Pompe et al. [4], among others.

Third, multivariable analyses revealed a threefold higher hazard ratio (HR) for GGG 3 compared to GGG 2 (3.44 vs. 1.10, p<0.001). However, no difference was found between the HR for GGG 4 and GGG 5 (4.18 vs.4.74, p<0.013). Similarly, Pompe et al.[4] found that the GGG yielded a 1.5-fold or greater HR differences between GGG 2 and 3, and Epstein et al. [6] found that the HR for PCa progression in GGG 2 and GGG 3 relative to GGG 1 was 2.2 and 7.3 respectively. Unlike our results, both groups found twofold or greater HR differences between GGG 4 and 5. This difference in results may be due to a low sample size of GGG 4 (6.9%) and 5 (2.7%) in our cohort.

Finally, the Harrell's C-index demonstrated a higher discriminatory ability of the novel GGG compared to the TGG model in the univariate analyses for both clinical (0.692 vs 0.671, C-index delta: 0.021) and pathological GGG (0.692 vs 0.647, C-index delta: 0.045). Similarly, the novel GGG compared to the TGG model in multivariable analyses for the pathological GGG revealed a higher discriminatory ability of the novel GGG (0.808 vs 0.790, C-index delta: 0.018). Interestingly, the Harrell's C-index showed a higher discriminatory ability of the TGG with separate groups for GS 8 and GS 9-10 compared to the TGG model in the multivariable analyses for the clinical GGG (0.745 vs 0.740, C-delta: 0.005). This finding solidifies the notion that GGG 5 has worse outcomes than GGG 4. Tsao et al. [14], among others, found similar results, showing that patients treated with RP with an GGG 5 had 1.74 increased risk of death compared to GGG 4 (HR 1.74, 95% CI 1.15-2.65).

Our study has numerous strengths which distinguishes itself from other studies. First, to the best of our knowledge we are the first to have validated the ISUP GGG classification in a Canadian cohort treated with RARP. Second, we extracted our information from a prospectively maintained Canadian database, and therefore did not have the limitations associated with large retrospective databases such as missing information and coding errors. Third, even though we did not have a central pathology review, our patients were only treated at two different hospitals and as such we were able to reclassify patients according to their GS by matching the modified new score to the corresponding GS category.

Our study is not devoid of limitations. First, we relied on BCR-free survival rather than a more specific endpoint such as prostate cancer-specific mortality. Since the Gleason system was first introduced in 2005 and our methodology relies on a prospective database, this limitation was inevitable. An additional 10 years of follow-up would be required to be able to use PCSM as an endpoint and our study mean follow-up was of 27.9 months. Second, we only had 60 patients who classified as GGG 4 and GGG 5, which may be underpowered to find statistically significant results.

Conclusion

In conclusion, the ISUP Gleason Grading classification is an independent predictor of BCR in Canadian PCa patients treated with RARP. The ISUP Gleason Grading classification showed added discriminate property, especially for the GGG 2 and GGG 3 strata.
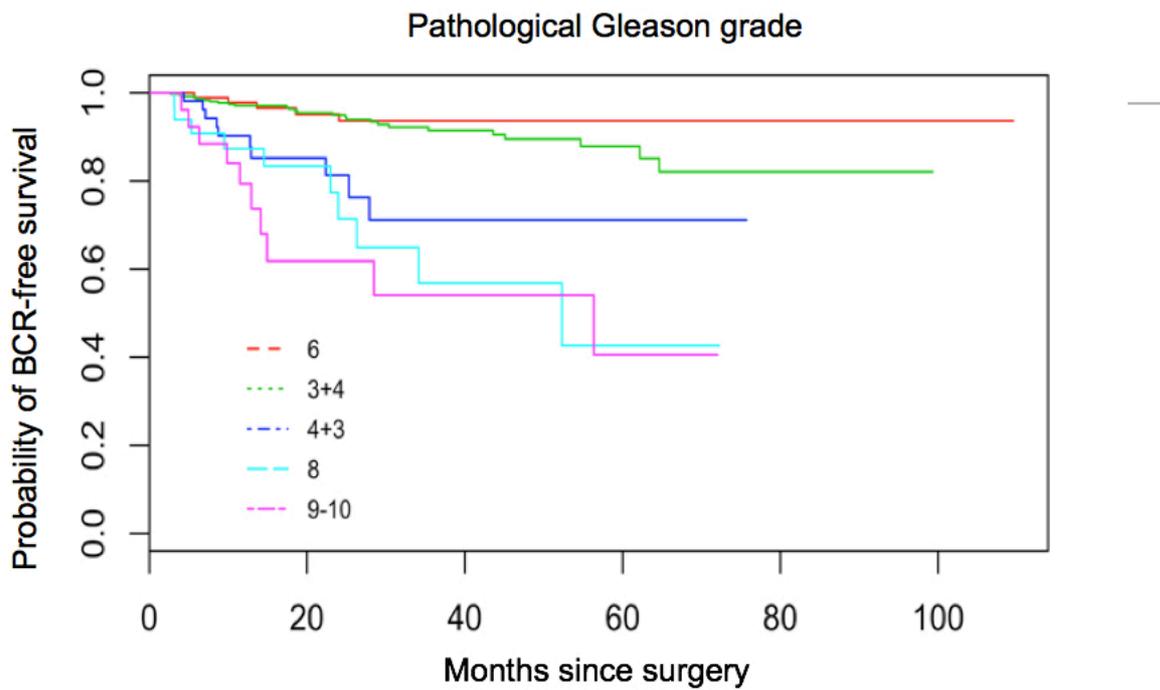
References

1. Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. J Urol. 1974;111(1):58-64.
2. 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):1228-42.
3. Pierorazio PM, Walsh PC, Partin AW, et al. Prognostic gleason grade grouping: Data based on the modified gleason scoring system. BJU Int. 2013;111(5):753-60.
4. Pompe RS, Davis-Bondarenko H, Zaffuto E, et al. Population-based validation of the 2014 isup gleason grade groups in patients treated with radical prostatectomy, brachytherapy, external beam radiation, or no local treatment. Prostate. 2017;77(6):686-93.
5. Epstein JI, Egevad L, Amin MB, et al. The 2014 international society of urological pathology (isup) consensus conference on gleason grading of prostatic carcinoma: Definition of grading patterns and proposal for a new grading system. Am J Surg Pathol. 2016;40(2):244-52.
6. Epstein JI, Zelefsky MJ, Sjoberg DD, et al. A contemporary prostate cancer grading system: A validated alternative to the gleason score. Eur Urol. 2016;69(3):428-35.
7. Spratt DE, Cole AI, Palapattu GS, et al. Independent surgical validation of the new prostate cancer grade grouping system. BJU Int. 2016.
8. Berney DM, Beltran L, Fisher G, et al. Validation of a contemporary prostate cancer grading system using prostate cancer death as outcome. Br J Cancer. 2016;114(10):1078-83.
9. Dell'Oglio P, Karnes RJ, Gandaglia G, et al. The new prostate cancer grading system does not improve prediction of clinical recurrence after radical prostatectomy: Results of a large, two-center validation study. Prostate. 2016.
10. D'Amico AV, Whittington R, Malkowicz SB, et al. The combination of preoperative prostate specific antigen and postoperative pathological findings to predict prostate specific antigen outcome in clinically localized prostate cancer. J Urol. 1998;160(6 Pt 1):2096-101.
11. He J, Albertsen PC, Moore D, et al. Validation of a contemporary five-tiered gleason grade grouping using population-based data. Eur Urol. 2016.
12. Loeb S, Folkvaljon Y, Robinson D, et al. Evaluation of the 2015 gleason grade groups in a nationwide population-based cohort. Eur Urol. 2016;69(6):1135-41.
13. Spratt DE, Jackson WC, Abugharib A, et al. Independent validation of the prognostic capacity of the isup prostate cancer grade grouping system for radiation treated patients with long-term follow-up. Prostate Cancer Prostatic Dis. 2016;19(3):292-7.
14. Tsao CK, Gray KP, Nakabayashi M, et al. Patients with biopsy gleason 9 and 10 prostate cancer have significantly worse outcomes compared to patients with gleason 8 disease. J Urol. 2015;194(1):91-7.

Figures and Tables

Fig. 1. Kaplan-Meier analysis for biochemical recurrence (BCR)-free survival following robotic-assisted radical prostatectomy (RARP) stratified by clinical Gleason grade groups. Red line: Gleason Score 6, grade group 1. Green line: Gleason Score 3 + 4, grade group 2. Dark blue line: Gleason Score 4 + 3, grade group 3. Turquoise blue line: Gleason Score 8, grade group 4. Pink line: Gleason Score 9 and 10, grade group 5.
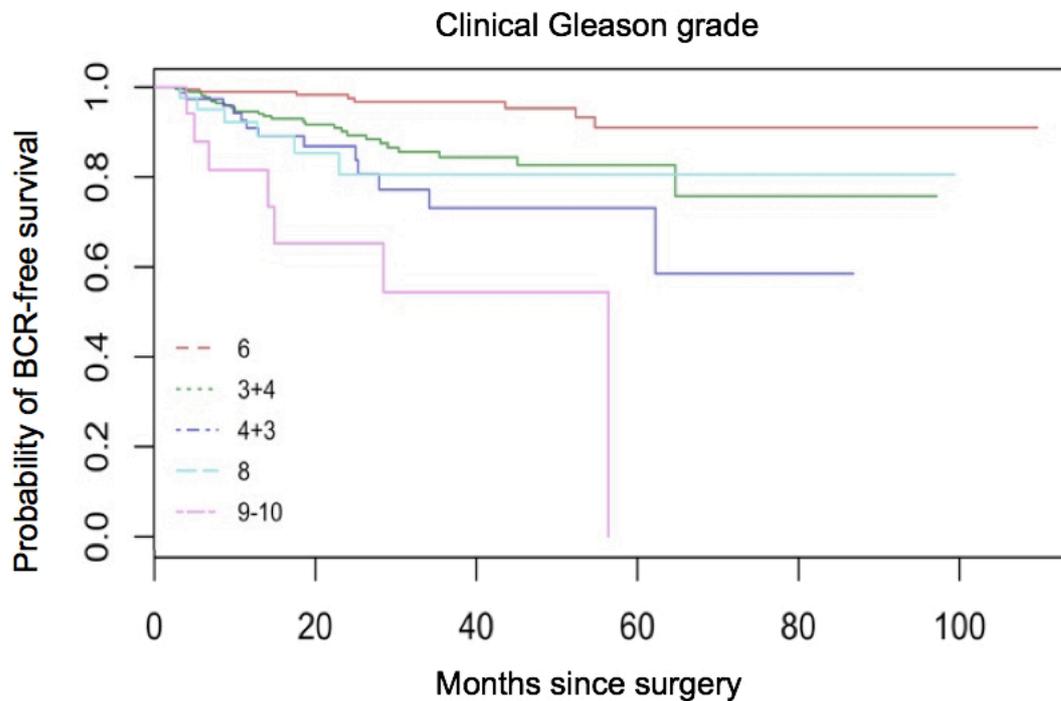
Fig. 2. Kaplan-Meier analysis for biochemical recurrence (BCR)-free survival following robotic-assisted radical prostatectomy (RARP) stratified by pathological Gleason grade groups. Red line: Gleason Score 6, grade group 1. Green line: Gleason score 3 + 4, grade group 2. Dark blue line: Gleason Score 4 + 3, grade group 3. Turquoise blue line: Gleason Score 8, grade group 4. Pink line: Gleason Score 9 and 10, grade group 5.

CUAJ – Original Research
Davis Bondarenko et al
Validation of new ISUP grading for predicting BCR

Table 1. Baseline characteristics of prostate cancer patients undergoing robotic-assisted radical prostatectomy between 2006 and 2016

Variables	
Mean age, years (95% CI)	60.34 (59.83, 60.85)
Mean body mass index, kg/m^2 (95% CI)	31.65 (31.25; 32.05)
Prostate volume, cc (95% CI)	39.04 (37.79; 40.30)
Mean preoperative PSA, ng/ml (95% CI)	6.51 (6.13; 6.88)
Biopsy Gleason score, % (n)	
6	32 (199)
3+4	45.5 (283)
4+3	12.7 (79)
8	6.9 (43)
9–10	2.7 (17)
Clinical stage, % (n)	
cT1b	0.2 (1)
cT1c	71.9 (447)
cT2a	20.6 (128)
cT2b	5.1 (32)
cT2c	0.9 (6)
cT3	1.1 (7)
Pathological Gleason score, % (n)	
LFU	0.2 (1)
No evidence of cancer	0.3 (2)
6	16.6 (103)
3+4	64.1 (398)
4+3	9.2 (57)
8	5.5 (34)
9–10	4.2 (26)
Pathological stage, % (n)	
pT0	0.3 (2)
pT2a	7.7 (48)
pT2b	5.8 (36)
pT2c	53.8 (334)
pT3a	27.2 (169)
pT3b	5.0 (31)
pT4	0.2 (1)
ECE, % (n)	31.2 (194)
PSM, % (n)	23.6 (147)
SVI, % (n)	4.6 (29)
Mean followup, months	27.9
BCR rate, % (n)	9.9 (62)

BCR: biochemical recurrence; CI: confidence interval; ECE: extracapsular extension; LFU: lost at followup; PSA: prostate-specific antigen; PSM: positive surgical margins; SVI: seminal vesicle invasion;

Table 2. Univariate and multivariable Cox-regression analyses according to different Gleason grading classifications for the clinical and pathological Gleason grade											
	Clinical Gleason grade						Pathological Gleason grade					
	Univariate			Multivariable			Univariate			Multivariable		
	HR	95% CI	p	HR	95% CI	p	HR	95% CI	p	HR	95% CI	p
6	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
3+4	3.47	1.58–7.65	<0.001	3.15	1.41–7.00	<0.001	1.53	0.58–3.98	0.383	1.10	0.41–2.93	0.842
4+3	5.90	2.44–14.29	<0.001	5.65	2.33–13.71	<0.001	5.14	1.75–15.09	0.002	3.44	1.13–10.41	0.028
8	4.99	1.72–14.42	<0.001	4.54	1.56–13.19	<0.001	8.15	2.78–23.91	<0.001	4.18	1.34–13.08	0.013
9–10	15.96	5.75–44.30	<0.001	13.34	4.69–37.95	<0.001	10.50	3.58–30.80	<0.001	4.74	1.49–15.06	<0.001
6	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
7	3.98	1.86–8.53	<0.001	3.65	1.69–7.88	<0.001	1.92	0.75–4.89	0.171	1.34	0.51–3.49	0.545
8–10	7.90	3.26–19.12	<0.001	7.00	2.86–17.15	<0.001	9.25	3.46–24.69	<0.001	4.28	1.48–12.34	<0.001

Multivariable clinical Gleason Cox model adjusted for preoperative prostate-specific antigen (PSA) and clinical stage (T1, T2, or T3/4). Multivariable pathological Gleason Cox model adjusted for PSA, surgical margin status, and pathological stage (pT2, pT3a, pT3b, or pT4). CI: confidence interval; HR: hazard ratio; Ref: reference.

Table 3. Discriminatory ability of the four Gleason grading classifications				
	Clinical Gleason grade		**Pathological Gleason grade**	
	Univariate	**Multivariable**	**Univariate**	**Multivariable**
TGG (6 vs. 7 vs. ≥8)	0.671	0.740	0.647	0.790
TGG and 2 separate groups for grade 7 (6 vs. 3+4 vs. 4+3 vs. ≥8)	0.686	0.730	0.677	0.793
TGG and 2 separate groups for grade 8–10 (6 vs. 7 vs. 8 vs. ≥9)	0.673	0.745	0.649	0.789
Novel GGG (6 vs. 3+4 vs. 4+3 vs. 8 vs. ≥9)	0.692	0.724	0.692	0.808

GGG: Gleason grade groups; TGG: traditional three-tiered Gleason grading.