The new prostate cancer grading system does not improve prediction of clinical recurrence after radical prostatectomy: results of a large, two-center validation study

Dell'Oglio, P.; Karnes, R.J.; Gandaglia, G.; Fossati, N.; Stabile, A.; Moschini, M.; ... ; Briganti, A.

The Prostate 77:263–273 (2017)

The New Prostate Cancer Grading System Does Not Improve Prediction of Clinical Recurrence After Radical Prostatectomy: Results of a Large, Two-Center Validation Study

Paolo Dell'Oglio,[1,2]* Robert Jeffrey Karnes,[3] Giorgio Gandaglia,[1] Nicola Fossati,[1] Armando Stabile,[1] Marco Moschini,[1,3] Vito Cucchiara,[1] Emanuele Zaffuto,[1] Pierre I. Karakiewicz,[2] Nazareno Suardi,[1] Francesco Montorsi,[1] and Alberto Briganti[1]

[1]*Department of Urology and Division of Experimental Oncology, URI, Urological Research Institute, IRCCS San Raffaele Scientific Institute, Milan, Italy*

[2]*Cancer Prognostics and Health Outcomes Unit, University of Montreal Health Center, Montreal, Canada*

[3]*Department of Urology, Mayo Clinic, Rochester, Minnesota*

BACKGROUND. A new prostate cancer (PCa) grading system (namely, Gleason score-GS-$\leq$6 vs. 3 + 4 vs. 4 + 3 vs. 8 vs. $\geq$9) was recently proposed and assessed on biochemical recurrence (BCR) showing improved predictive abilities compared to the commonly used three-tier system (GS $\leq$6 vs. 7 vs. $\geq$8). We assessed the predictive ability of the five-tier grade group (GG) system on harder clinical endpoint, namely clinical recurrence (CR).

METHODS. Between 2005 and 2014, 9,728 clinically localized PCa patients were treated with radical prostatectomy (RP) at two tertiary referral centers. Kaplan–Meier curves, multivariable Cox regression analyses, and concordance index (C-index) were used to assess CR after treatment according to four Gleason grade classifications at biopsy and RP: Group 1: $\leq$6 versus 7 versus $\geq$8; Group 2: $\leq$6 versus 3 + 4 vs. 4 + 3 versus $\geq$8; Group 3: $\leq$6 versus 7 versus 8 versus $\geq$9; Group 4: $\leq$6 versus 3 + 4 versus 4 + 3 versus 8 versus $\geq$9. Same analyses were repeated in patients who had BCR (n = 1,624). Decision curve analyses were performed to evaluate and compare the net benefit associated with the use of the four Gleason grade classifications.

RESULTS. Overall, 443 (4.6%) patients had CR. The hazard ratio of the GS 3 + 4, 4 + 3, 8, and $\geq$9 relative to GS $\leq$6 were 3.63, 5.93, 11.44, 18.08 and 4.93, 9.99, 15.31 and 25.12 in the pre- and post-treatment models, respectively. The C-index of the five-tier GG system was slightly higher relative to the other 3 Gleason grade classifications both in the pre- (range: 0.001–0.006) and post-treatment models (range: 0–0.008). Similar findings were observed when we focused our analyses in patients with BCR after RP. The use of the five-tier GG system did not result into higher net-benefit relative to the other three Gleason grade classifications.

CONCLUSIONS. The difference in accuracy between the five-tier GG system and the other Gleason grade classifications, using CR as an endpoint, is clinically negligible. Current evidence suggests that the five-tier GG system represents a simplified

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of interest: The authors declare no conflicts of interest in preparing this article.

*Correspondence to: Paolo Dell'Oglio, MD, Department of Urology and Division of Experimental Oncology, Urological Research Institute, IRCCS San Raffaele Scientific Institute, Via Olgettina 60, Milan 20132, Italy. E-mail: paolo.delloglio@gmail.com

Received 3 July 2016; Accepted 20 September 2016

DOI 10.1002/pros.23265

Published online 18 October 2016 in Wiley Online Library (wileyonlinelibrary.com).

user-friendly scheme available for patient counseling rather than a new histopathological diagnostic system that improves the prediction of CR. *Prostate 77:263–273, 2017.*

KEY WORDS: **clinical recurrence; Epstein; Gleason grade; prostate cancer; validation**

INTRODUCTION

The Gleason system developed by Donald Gleason and the Veterans Administration Cooperative Urologic Research Group in the 1960s [1,2] was modified firstly during the International Society of Urological Pathology (ISUP) conference in 2005 [3]. Recently, a new prostate cancer (PCa) grading system has been proposed [4]. This new grading system consists of a five-grade group (GG): GG 1: Gleason score (GS) ≤ 6; GG 2: GS $3+4=7$; GG 3: GS $4+3=7$; GG 4: GS 8; GG 5: GS ≥ 9. The main reasons why this new grading system was introduced were to (i) improve patient counseling since the most differentiated PCa currently corresponds to a GS 6 which may lead to patient misunderstandings about his oncological profile; (ii) stratify PCa patients with GS 7 and 8–10 into respectively GS $3+4$ versus $4+3$ and GS 8 versus GS 9–10, that tend to have different prognosis [5–8].

To date, only three studies [9–11] assessed and validated the five-tier GG system on biochemical recurrence (BCR) in PCa patients treated with radical prostatectomy (RP), supporting the inclusion of this new scheme as part of the pathology report for prostate biopsy and RP, as suggested at the 2014 ISUP meeting [12].

The objective of the current study was to validate the five-tier GG system on a stronger outcome, namely clinical recurrence (CR), relying on contemporary PCa patients treated with RP within two tertiary care referral centers. We hypothesized that the five-tier GG system has higher predictive ability and superior clinical benefit relative to the commonly used three-tier system (Gleason score ≤ 6 vs. 7 vs. ≥ 8).

MATERIALS AND METHODS

Study Population

Twenty-nine thousand and fifty-three patients with clinically localized PCa treated with RP at San Raffaele Hospital (Milan, Italy) and Mayo Clinic (Rochester, Minnesota) were assessed. Due to changes in Gleason grading at the 2005 [3] and 2014 ISUP meeting [12], we exclusively focused our analyses on patients diagnosed with PCa and treated with RP between 2005 and 2014 with complete data regarding biopsy GS, pathologic GS, and follow-up. This resulted in a final cohort of 9,728 consecutive assessable PCa patients. At the time of surgery, 9,256 patients (95.1%) received pelvic lymph node dissection [13,14], based on the clinical judgment of the surgeons and/or pre-operative tool [15] predicting the presence of lymph node invasion (LNI).

Variable Definition

Covariates consisted of age at surgery, pre-operative prostate-specific antigen (PSA) values, year of surgery, clinical stage (T1 vs. T2 vs. T3/T4 vs. unknown), biopsy GS, pathologic stage (T2 vs. T3a vs. T3b vs. T4 vs. unknown), pathologic GS, surgical margin status (negative vs. positive vs. unknown), and LNI. Biopsy and RP GSs were assigned by pathologists with genitourinary expertise at each institution.

None of the patients included in the study received neo-adjuvant hormonal therapies. On the contrary, adjuvant therapies consisted of either adjuvant radiotherapy (aRT) and/or adjuvant hormone-deprivation therapy (aHT) that were administered based on the clinical judgement of the treating physician, according to patient and tumor characteristics. Adjuvant RT was defined as local radiation delivered to the prostatic bed and to the seminal vesicles alone or to the whole pelvic fossa according to the presence of LNI at final pathology, as previously described [16,17]. For this study, adjuvant treatment was defined as treatment given within 3 months following RP, in patients with an undetectable PSA (<0.10 ng/ml).

During follow-up serum PSA was measured every 3 months for the 1st year, biannually between the 2nd and the 5th years after surgery, and annually thereafter. BCR was defined as a PSA value ≥ 0.2 ng/ml on two subsequent measurements. Clinical recurrence was defined as positive imaging during follow-up after the onset of BCR. All patients with CR included in this study underwent staging imaging procedures after BCR. These consisted of bone scan and/or computed tomography (CT) and/or abdominal magnetic resonance imaging and/or 11C-choline positron emission tomography/CT scan [18].

End-Point

The outcome of our study was represented by the CR, defined as local and/or nodal recurrence (recurrence in the prostatic bed and/or pelvic lymph

nodes), retroperitoneal nodal recurrence, systemic recurrence (skeletal and/or visceral relapse).

Statistical Analyses

Statistical analyses, as well as reporting and interpretation of the results, were conducted according to established guidelines [19] and consisted of five steps. First, medians and interquartile ranges (IQR) or frequencies and proportions were reported for continuous or categorical variables, respectively.

Second, we estimated 5-year CR-free survival rates, in the overall population (n = 9,728) using Kaplan–Meier method according to four possible Gleason grade classifications at biopsy (pre-treatment models), as well as according to four possible Gleason grade classifications at RP (post-treatment models). Same analyses were repeated exclusively in patients who had BCR after RP (n = 1,624). The four Gleason grade classifications we planned to assess were as follows:

(1) Group 1: ≤6 versus 7 versus ≥8 (three-tier system);
(2) Group 2: ≤6 versus 3 + 4 versus 4 + 3 versus ≥8;
(3) Group 3: ≤6 versus 7 versus 8 versus ≥9;
(4) Group 4: ≤6 versus 3 + 4 versus 4 + 3 versus 8 versus ≥9 (five-tier GG system [4,12]).

Log-rank test was used to assess univariable differences in CR by GS.

Third, multivariable Cox regression analyses were used to predict CR using each of the four Gleason grade classifications. The pre-treatment models were adjusted for age at surgery, pre-operative PSA, and clinical stage (T1 vs. T2 vs. T3/T4) both in the overall population and in patients with BCR after surgery. The post-treatment models were adjusted for age at surgery, pre-operative PSA, pathologic stage (pT2 vs. pT3a vs. pT3b vs. pT4), surgical margin status, LNI, number of lymph nodes removed, aRT, and aHT in the overall population. The post-treatment models in patients who had BCR were adjusted for the same variables and for time to BCR too.

Fourth, leave-one-out cross-validation was used to construct the concordance index (C-index) [20] to assess the discrimination of the four Gleason grade classifications at biopsy (pre-treatment models) and RP (post-treatment models).

Fifth, in order to evaluate the clinical impact of the five-tier GG system relative to the other Gleason grade classifications, we relied on decision curves analyses as described by Vickers and Elkin proposed model, to evaluate and compare the net benefit [21].

Finally, in a way of sensitivity analysis and to confirm the robustness of our findings, all the aforementioned analyses were repeated in the subgroup of patients who did not receive adjuvant treatment (overall: n = 8,694; patients with BCR: n = 1,262).

All statistical tests were performed using the RStudio graphical interface v.0.98 for R software environment v.3.0.2 (R Foundation, Vienna, Austria). All tests were two-sided with a significance level set at P-value <0.05.

RESULTS

Patient Characteristics

Demographics and tumor characteristics of the cohort are reported in Table I. Overall, 5,575 (57.3%), 2,323 (23.9%), 918 (9.4%), 513 (5.3%), and 399 (4.1%) patients had biopsy GS ≤6, 3 + 4, 4 + 3, 8, ≥9, respectively. Overall, 4,034 (41.5%), 3,406 (35%), 1,226 (12.6%), 390 (4%), and 672 (6.9%) patients had pathologic GS ≤6, 3 + 4, 4 + 3, 8, ≥9, respectively. The majority of patients had negative surgical margin (80.9%) and no LNI (88.5%). Overall, 536 (5.5%) and 718 (7.4%) were treated with aRT and aHT, respectively.

The median follow-up after RP in patients without CR was 5.8 years (IQR: 3.3–12.7). During the study period, 1,624 (16.7%) developed BCR after RP. Of these, 443 (4.6%) harbored CR.

Survival Estimates and Prediction of Clinical Recurrence

Figure 1A and B depict the 5-year CR free-survival rate according to four Gleason grade classifications at biopsy (pre-treatment models), as well as according to four Gleason grade classifications at RP (post-treatment models), respectively. A statistically significant difference was observed in CR-free survival rates within each Gleason grade classifications (all $P < 0.001$; Fig. 1A and B). Regarding the five-tier GG system, the 5-year CR free-survival rates were 98.7 versus 93.5 versus 90 versus 77.6 versus 69.2% in patients with biopsy GS ≤6 versus 3 + 4 versus 4 + 3 versus 8 versus ≥9 (Fig. 1A), and 99.3 versus 95.8 versus 90.3 versus 83.8 versus 69.6% in patients with RP GS ≤6 versus 3 + 4 versus 4 + 3 versus 8 versus ≥9 (Fig. 1B), respectively. Similar findings were observed when we focused our analyses in patients who recurred after RP and in those patients who did not receive adjuvant treatment (data not showed).

In multivariable analyses predicting CR (Tables II and III), each biopsy Gleason grade classifications (pre-treatment models) as well as each RP Gleason grade classifications (post-treatment models) were independent predictors of CR after accounting for different confounders. Regarding the five-tier GG

TABLE I. Descriptive Characteristics of 9,728 Prostate Cancer Patients Treated With Radical Prostatectomy at San Raffaele Hospital (Milan, Italy) and Mayo Clinic (Rochester, Minnesota) Between 2005 and 2014

Variables	Overall (n = 9,728)
Age at surgery, yrs	
Median	62
IQR	57–67
Pre-operative PSA value, ng/ml	
Median	5.6
IQR	4.2–8
Clinical stage, n (%)	
T1	5,857 (60.2)
T2	3,098 (31.8)
T3–T4	458 (4.7)
Unknown	315 (3.2)
Biopsy Gleason score, n (%)	
≤6 (GG 1)	5,575 (57.3)
3 + 4 (GG 2)	2,323 (23.9)
4+3 (GG 3)	918 (9.4)
8 (GG 4)	513 (5.3)
≥9 (GG 5)	399 (4.1)
Pathologic stage, n (%)	
T2	7,656 (78.7)
T3a	1,041 (10.7)
T3b	544 (5.6)
T4	49 (0.5)
Unknown	438 (4.5)
Pathologic Gleason Score, n (%)	
≤6 (GG 1)	4,034 (41.5)
3 + 4 (GG 2)	3,406 (35)
4 + 3 (GG 3)	1,226 (12.6)
8 (GG 4)	390 (4)
≥9 (GG 5)	672 (6.9)
Surgical margin status, n (%)	
Negative	7,867 (80.9)
Positive	1,818 (18.7)
Unknown	43 (0.4)
aRT, n (%)	
No	9,170 (94.3)
Yes	536 (5.5)
Unknown	22 (0.2)
aHT, n (%)	
No	8,988 (92.4)
Yes	718 (7.4)
Unknown	22 (0.2)
Lymph node invasion, n (%)	
pN0	8,613 (88.5)
pN1	643 (6.6)
pNx	472 (4.9)

Biopsy and Pathologic Gleason score are stratified according to the new Gleason system.
IQR, interquartile range; GG, grade group; aRT, adjuvant radiotherapy; aHT, adjuvant hormonal therapy.

system (Table IId), the hazard ratio (HR) of the GS 3 + 4, 4 + 3, 8, and ≥9 relative to GS ≤6 were 3.63, 5.93, 11.44, 18.08 and 4.93, 9.99, 15.31, and 25.12 in the pre- and post-treatment models, respectively (Table II). When the same analyses were repeated exclusively in patients with BCR (Table III) and in those who did not receive adjuvant treatment (data not showed), each biopsy Gleason grade classifications, as well as each RP Gleason grade classifications remained independent predictors of CR.

Discrimination of the Four Gleason Grade Classifications

In leave-one-out cross-validation, the C-index of the five-tier GG system was slightly higher relative to the other three Gleason grade classifications. Specifically, in the pre-treatment models (Table IVA) this increase of the C-index ranged between 0.001 and 0.006, after accounting for age at surgery, pre-operative PSA and clinical stage. In the post-treatment models (Table IVA), the C-index of the five-tier GG system was 0–0.008 higher than the other three Gleason grade classifications (Table IVA), after accounting for age at surgery, pre-operative PSA, pathologic stage, surgical margin status, LNI, number of lymph nodes removed, aRT, and aHT. Similar findings were observed when leave-one-out cross-validation was exclusively performed in patients who recurred after RP (Table IVB). Specifically, the increase of the C-index was 0.003–0.007 and 0.001–0.003 in the pre- and post-treatment models, respectively. Finally, similar findings were also observed when leave-one-out cross-validation was exclusively performed in those patients who did not receive adjuvant treatment (data not showed).

Evaluation of the Four Gleason Grade Classifications Clinical Decision Making

When compared to the select all- and select none-strategy at decision curve analysis (Fig. 2), the use of the five-tier GG system resulted into no higher net-benefit relative to the three Gleason grade classifications for clinical decision making, both in the pre- and post-treatment models. These findings were confirmed when the same analyses were repeated exclusively in patients with BCR (Fig. 2) and in those who did not receive adjuvant treatment (data not showed).

DISCUSSION

In 2013, the five-tier GG system based on data from Johns Hopkins Hospital was proposed [4]. This novel classification incorporates a prognostic grade grouping which can accurately reflect patient prognosis and improve patient counseling at the same time. Consensus was reached for the adoption of this new grading system during the 2014 ISUP meeting in

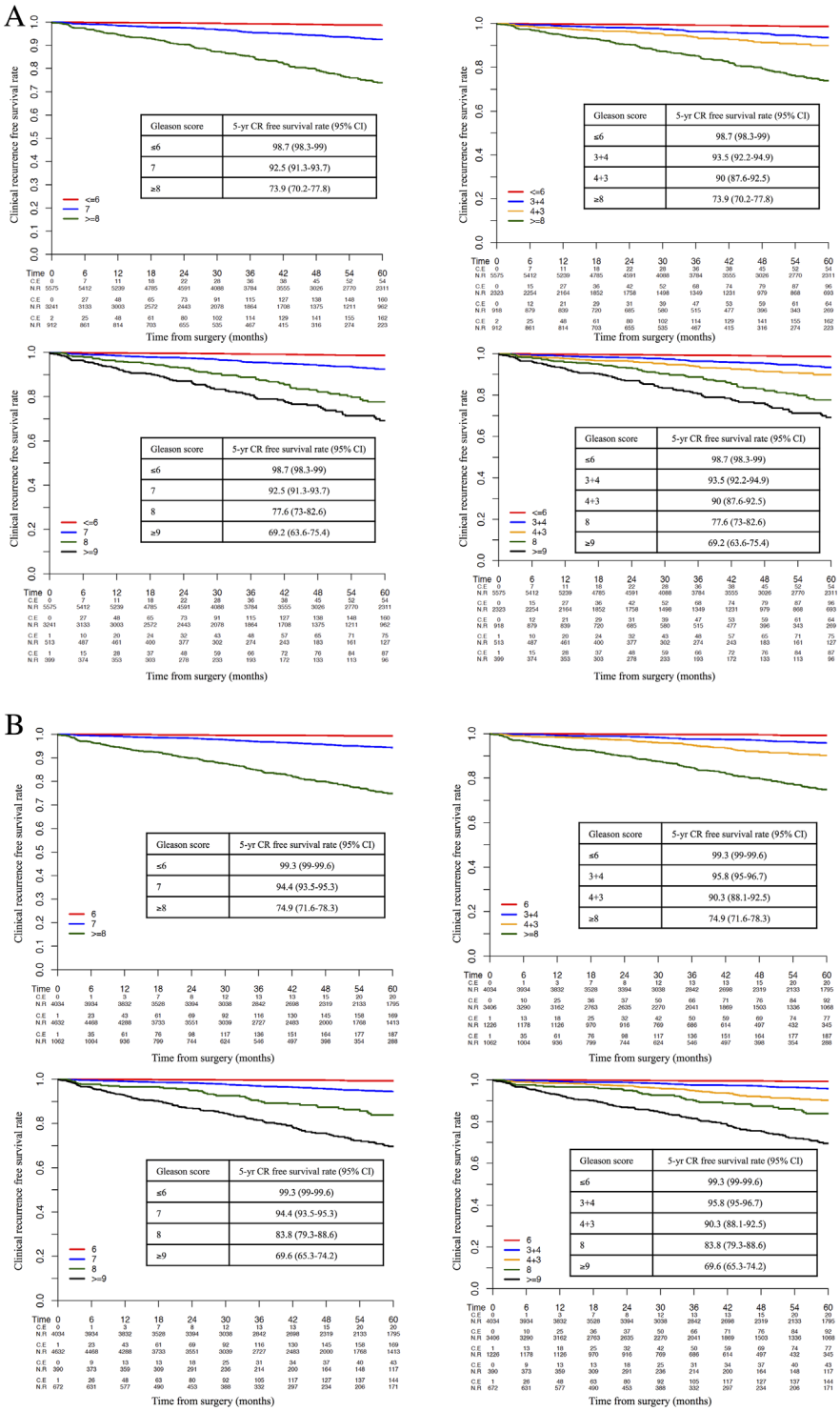

Fig. 1. Clinical recurrence free survival rate in 9,728 prostate cancer patients treated with radical prostatectomy stratified according to four Gleason grade classifications at biopsy (pre-treatment models (**A**)) and four Gleason grade classifications at radical prostatectomy (post-treatment models (**B**)).

Chicago [12]. This has led to its recently introduction into the European urological guidelines [22].

The objective of the current manuscript was to assess and validate this five-tier GG system on CR. Its rationale stems from lack of a study that assesses the predictive ability of this grading system on harder clinical endpoint than BCR in PCa patients treated with RP. We postulated that the five-tier GG system

TABLE II. Multivariable Cox Regression Analyses Predicting Clinical Recurrence in 9,728 Prostate Cancer Patients Treated With Radical Prostatectomy, According to Four Gleason Grade Classifications at Biopsy (Pre-Treatment Models) and Radical Prostatectomy (Post-Treatment Models)*

	Gleason grade at biopsy (pre-treatment models)		Gleason grade at RP (post-treatment models)	
	Multivariable		Multivariable	
	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value
a) Group 1				
≤6	1.00 (ref.)	–	1.00 (ref.)	–
7	4.24 (3.22–5.58)	<0.001	6.01 (3.91–9.18)	<0.001
≥8	13.84 (10.29–18.60)	<0.001	18.14 (11.44–28.77)	<0.001
b) Group 2				
≤6	1.00 (ref.)	–	1.00 (ref.)	–
3 + 4	3.61 (2.67–4.87)	<0.001	4.85 (3.12–7.55)	<0.001
4 + 3	5.87 (4.21–8.17)	<0.001	9.59 (6.03–15.25)	<0.001
≥8	13.99 (10.41–18.81)	<0.001	20.23 (12.76–32.05)	<0.001
c) Group 3				
≤6	1.00 (ref.)	–	1.00 (ref.)	–
7	4.27 (3.24–5.62)	<0.001	6.13 (3.99–9.38)	<0.001
8	11.33 (8.10–15.86)	<0.001	14.02 (8.42–23.36)	<0.001
≥9	17.83 (12.73–24.98)	<0.001	22.04 (13.61–35.67)	<0.001
d) Group 4				
≤6 (GG 1)	1.00 (ref.)	–	1.00 (ref.)	–
3 + 4 (GG 2)	3.63 (2.69–4.89)	<0.001	4.93 (3.17–7.67)	<0.001
4 + 3 (GG 3)	5.93 (4.25–8.26)	<0.001	9.99 (6.28–15.90)	<0.001
8 (GG 4)	11.44 (8.18–16.01)	<0.001	15.31 (9.19–25.48)	<0.001
≥9 (GG 5)	18.08 (12.91–25.34)	<0.001	25.12 (15.53–40.65)	<0.001

*The pre-treatment models were adjusted for age at surgery, pre-operative PSA and clinical stage (T1 vs. T2 vs. T3/T4). The post-treatment models were adjusted for age at surgery, pre-operative PSA, pathologic stage (pT2 vs. pT3a vs. pT3b vs. pT4), surgical margin status, lymph node invasion, number of lymph nodes removed, adjuvant radiotherapy and adjuvant androgen deprivation therapy.
GG, grade group.

has higher predictive ability and superior clinical benefit relative to the commonly used three-tier system (Gleason score ≤6 vs. 7 vs. ≥8). Our findings failed to confirm our hypotheses and warrant considerations.

First, we observed a statistically significant difference in terms of 5-year CR free-survival rate between all the five GG of the new grading system [4], both in the overall population and in patients who had BCR after RP. The absence of a plateau into the Kaplan–Meier curves suggests that the differences observed between all the five GG of the new grading system will be greater in future validation studies with longer follow-up.

Second, these findings persisted even after adjusting for several confounders (Tables II and III). It is important to note that the HR of the GS 4 + 3 and ≥9 were about twice as high than for GS 3 + 4 and eight both in the pre- and post-treatment models, respectively. This increase of the HR was also observed when the same analyses were repeated exclusively in patients who had BCR after RP, even thought less overwhelmingly. In consequence, our results support the separation of GS 7 into 3 + 4 and 4 + 3, as well as of GS 8–10 into 8 and 9–10 [4,9,10,12].

However, when we compared the C-index of the five-tier GG system with the other three Gleason grade classifications (Table III), virtually no change in predictive accuracy was observed. Specifically, in the overall population the increase of the C-index was at best of 0.006 and 0.008 in the pre- and post-treatment models, respectively. With regards to analyses performed in patients who had BCR after RP, the increase of the C-index was at best of 0.007 and 0.003 in the pre- and post-treatment models, respectively.

Moreover, when compared to the select all- and select none-strategy at decision curve analysis (Fig. 2), we failed to observe a net-benefit of the five-tier GG system relative to the other three Gleason grade classifications for clinical decision making.

Finally, similar findings were invariably confirmed in the subgroup analyses performed in those patients who did not receive adjuvant treatment.

TABLE III. Multivariable Cox Regression Analyses Predicting Clinical Recurrence in 1,624 Prostate Cancer Patients Who Had Biochemical Recurrence After Radical Prostatectomy, According to Four Gleason Grade Classifications at Biopsy (Pre-Treatment Models) and Radical Prostatectomy (Post-Treatment Models)*

	Gleason grade at biopsy (pre-treatment models)		Gleason grade at RP (post-treatment models)	
	Multivariable		Multivariable	
	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value
a) Group 1				
≤6	1.00 (ref.)	–	1.00 (ref.)	–
7	2.10 (1.58–2.75)	<0.001	2.49 (1.62–3.82)	<0.001
≥8	4.02 (2.99–5.39)	<0.001	4.27 (2.70–6.74)	<0.001
b) Group 2				
≤6	1.00 (ref.)	–	1.00 (ref.)	–
3 + 4	2.00 (1.48–2.70)	<0.001	2.27 (1.46–3.53)	<0.001
4 + 3	2.24 (1.61–3.11)	<0.001	2.90 (1.83–4.60)	<0.001
≥8	4.03 (3.00–5.41)	<0.001	4.40 (2.78–6.95)	<0.001
c) Group 3				
≤6	1.00 (ref.)	–	1.00 (ref.)	–
7	2.10 (1.59–2.76)	<0.001	2.52 (1.65–3.87)	<0.001
8	3.55 (2.54–4.97)	<0.001	3.29 (1.97–5.49)	<0.001
≥9	4.62 (3.31–6.45)	<0.001	4.99 (3.11–8.01)	<0.001
d) Group 4				
≤6 (GG 1)	1.00 (ref.)	–	1.00 (ref.)	–
3 + 4 (GG 2)	2.01 (1.49–2.71)	<0.001	2.29 (1.47–3.57)	< 0.001
4 + 3 (GG 3)	2.25 (1.62–3.13)	<0.001	2.97 (1.87–4.70)	< 0.001
8 (GG 4)	3.56 (2.55–4.98)	<0.001	3.38 (2.02–5.64)	< 0.001
≥9 (GG 5)	4.63 (3.31–6.46)	<0.001	5.17 (3.22–8.30)	< 0.001

*The pre-treatment models were adjusted for age at surgery, pre-operative PSA, and clinical stage (T1 vs. T2 vs. T3/T4). The post-treatment models were adjusted for age at surgery, pre-operative PSA, pathologic stage (pT2 vs. pT3a vs. pT3b vs. pT4), surgical margin status, lymph node invasion, number of lymph nodes removed, adjuvant radiotherapy, adjuvant androgen deprivation therapy, and time to BCR.
GG, grade group.

To summarize, we provided evidence for the first time that the five-tier GG system predict CR. However, we failed to observe higher predictive ability and superior clinical benefit of the new grading system relative to the commonly used three-tier system (Gleason score ≤6 vs. 7 vs. ≥8). In consequence, these findings should bear in mind before talking of a rearrangement of the most common risk stratification for PCa, namely D'Amico classification [23] that stratify PCa based on serum PSA values, clinical stage and biopsy GS into low-risk, intermediate- and high-risk groups, with GS ≤6, 7 and 8–10, respectively.

The main limitation of the study where the basis for the five-tier GG system was proposed [4] is represented by the short-term follow-up (median: 2 years). In consequence, studies with longer follow-up using the new grading system were needed to validate those findings. For this reason, Epstein and colleagues [9] assessed the predictive ability of the five-tier GG system on BCR in 20,845 men treated with RP at five academic centers with a median follow-up of 3 years. The authors showed a slightly higher predictive accuracy of the five-tier GG system relative to the commonly used three-tier system. Specifically, the increase of the C-index was at best of 0.008 and 0.012 in the pre- and post-treatment models, respectively. When the same authors [9] assessed the five-tier GG system in patients treated with RT (n = 5,501), the C-index for the five-tier GG system was 0.008 at best higher than for the commonly used three-tier system. Thereafter, Loeb and colleagues [10] examined the five-tier GG system on a population-based cohort from The National Prostate Cancer Register of Sweden made up of 4,325 men who underwent RP with a median follow-up of 4.6 years. Using BCR as the primary outcome, the authors concluded that the increment of the C-index of the five-tier GG system compared to the commonly used three-tier system ranged between 0 and 0.01 in pre- and post-treatment models. Conversely, the C-index of the five-tier GG system was even lower relative to

TABLE IV. Leave-One-Out Cross Validation to Assess the Discrimination of Four Gleason Grade Classifications at Biopsy (Pre-Treatment Models) and Radical Prostatectomy (Post-Treatment Models) in the Overall Population (n = 9,728) and in Patients Who Had Biochemical Recurrence After Radical Prostatectomy (n = 1,624)*

	A		B	
	Overall (n = 9,728)		BCR yes (n = 1,624)	
	Gleason grade at biopsy (pre-treatment models)	Gleason grade at RP (post-treatment models)	Gleason grade at biopsy (pre-treatment models)	Gleason grade at RP (post-treatment models)
	Multivariable	Multivariable	Multivariable	Multivariable
a) Group 1 ≤6 versus 7 versus ≥8	0.790	0.827	0.653	0.693
b) Group 2 ≤6 versus 3 + 4 versus 4 + 3 versus ≥8	0.795	0.835	0.656	0.695
c) Group 3 ≤6 versus 7 versus 8 versus ≥9	0.791	0.828	0.657	0.694
d) Group 4 ≤6 versus 3 + 4 versus 4 + 3 versus 8 versus ≥9	0.796	0.835	0.660	0.696

*The pre-treatment models were adjusted for age at surgery, pre-operative PSA and clinical stage (T1 vs. T2 vs. T3/T4).
The post-treatment models in the overall population were adjusted for age at surgery, pre-operative PSA, pathologic stage (pT2 vs. pT3a vs. pT3b vs. pT4), surgical margin status, lymph node invasion, number of lymph nodes removed, adjuvant radiotherapy and adjuvant androgen deprivation therapy.
The post-treatment models in patients who had biochemical recurrence were adjusted for age at surgery, pre-operative PSA, pathologic stage (pT2 vs. pT3a vs. pT3b vs. pT4), surgical margin status, lymph node invasion, number of lymph nodes removed, adjuvant radiotherapy, adjuvant androgen deprivation therapy, and time to BCR.

the commonly used three-tier system when the same analyses where repeated in 1,555 patients treated with RT (−0.011). The predictive ability of the five-tier GG system on BCR was also tested by Spratt and colleagues [11] in 3,694 men treated with RP at a single institution between 1994 and 2013, with a median follow-up of 4.4 years. The authors observed an increment of the discrimination of the five-tier GG system relative to the commonly used three-tier system (0.02 and 0.06 in the pre- and post-treatment models) that was more prominent in the sub-analyses performed in patients treated after 2005 (0.04 and 0.1 in the pre- and post-treatment models). However, the sample size was smaller relative to the previous aforementioned validation studies [9,10].

Our work differs from the aforementioned studies [9–11] for several factors. First, the outcome of interest is stronger, namely CR instead of BCR. Second the follow-up is longer (median: 5.8 vs. 3 vs. 4.6 vs. 4.4 years). Third, we assessed the net benefit of the five-tier GG system that is mandatory in a validation study. Last but not least, in our multivariable models we adjusted also for age at surgery, LNI, number of lymph nodes removed, and adjuvant treatments that are important variables that should be account for.

Recently, other authors validated the five-tier GG system in conservatively [24] or RT treated [25] PCa patients, on a stronger outcome than BCR, namely PCa death. However, these authors did not assess the discrimination and/or the net-benefit of the five-tier GG system relative to the commonly used three-tier system that certainly limits their findings.

The aim of our study was not to discredit the five-tier GG system [4] and the subsequent validations studies [9–11], who should be commend for their effort to improve the clinical practice distilling grades of PCa down to the lowest number of grades, each with an unique prognosis. We agree with the authors [4,9–11] that the five-tier GG system provides a simplified user-friendly classification scheme that improves the patient counseling. Indeed, when patients are told that they have a GS 6 out of 10, leads to a logical yet incorrect assumption that their tumor is in the mid range of aggressiveness. In consequence, the prognosis may be perceived as intermediate despite the fact that GS 6 is the lowest score currently

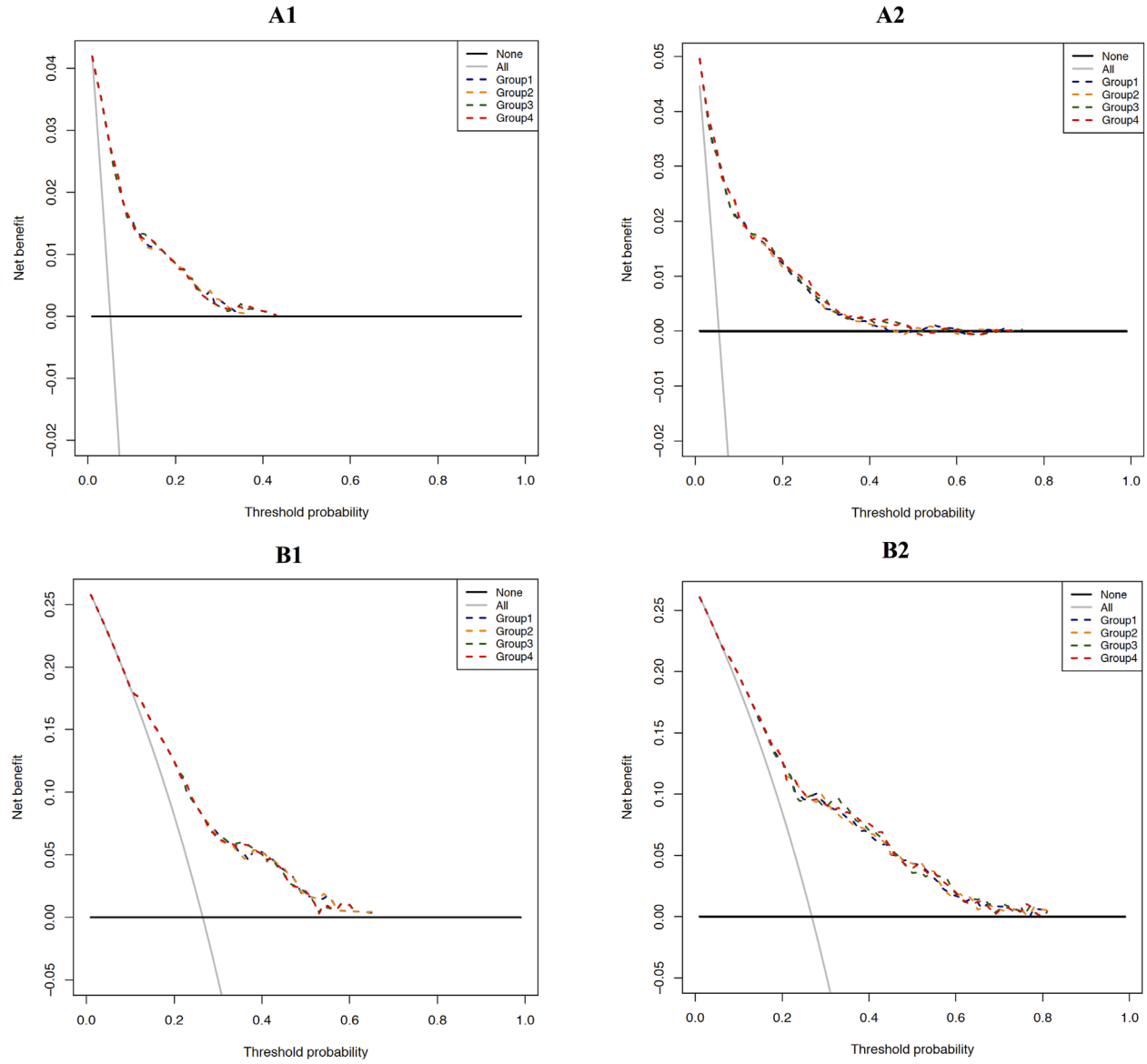

Fig. 2. Decision curve analyses depicting the net benefit of the new prostate cancer grading system (Group 4: ≤6 vs. 3 + 4 vs. 4 + 3 vs. 8 vs. ≥9) relative to three Gleason grade classifications (Group 1: ≤6 vs. 7 vs. ≥8; Group 2: ≤6 vs. 3 + 4 vs. 4 + 3 vs. ≥8; Group 3: ≤6 vs. 7 vs. 8 vs. ≥9) (**A**) in the overall population (n = 9,728) and (**B**) in patients with biochemical recurrence after radical prostatectomy (n = 1,624) according to Gleason grade classifications at biopsy (pre-treatment models (**A1**–**B1**)) and Gleason grade classifications at radical prostatectomy (post-treatment models (**A2**–**B2**)).

assigned. However, we failed to observe a clinical benefit compared to the commonly used three-tier system. At the same time, there is an impending need of new models to improve the prediction of oncological outcomes of PCa patients. Probably, the five-tier GG system [4,12] will be the cornerstone of these future models. However, even thought the Gleason system is the strongest predictor of oncological outcomes for men with PCa, it is not the single one. Indeed, the combination of this five-tier GG system with, for example, multiparametric magnetic resonance imaging as well as new genomic tissue tests might help the clinicians to develop tools not only able to aid in patient counseling but also to overwhelmingly improve predictive accuracy of the oncological outcome of interest.

Despite several strengths, our analyses are not devoid of limitations. First, they are limited by their retrospective nature. Second, no centralized pathologic review of the specimens was performed. However, this limitation is in common with previous studies that validated the five-tier GG system on

BCR [9–11]. Third, we were not able to verify our findings in a group of men treated with RT. However, this is the first study that validated the five-tier GG system on CR in PCa patients treated with RP. Fourth, despite this represents the validation study in PCa patients treated with RP with the longer follow-up available, future studies with longer follow-up are needed to validate our findings. Finally, patients included in our study received different imaging procedures at the time of recurrence according to the clinical judgment of the treating physician. This could have introduced an element of heterogeneity that might in part limit the validity of our findings.

CONCLUSION

The difference in accuracy between the five-tier GG system and the other Gleason grade classifications, using CR as an endpoint, is clinically negligible. Current evidence suggests that the five-tier GG system represents a simplified user-friendly scheme available for patient counseling rather than a new histopathological diagnostic system that improves the prediction of CR.

REFERENCES

1. Mellinger GT, Gleason D, Bailar J, 3rd. The histology and prognosis of prostatic cancer. J Urol 1967;97:331–337.
2. Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. J Urol 1974;111:58–64.
3. Epstein JI, Allsbrook WC, Jr., Amin MB, Egevad LL, Committee IG. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. Am J Surg Pathol 2005;29:1228–1242.
4. Pierorazio PM, Walsh PC, Partin AW, Epstein JI. Prognostic Gleason grade grouping: Data based on the modified Gleason scoring system. BJU Int 2013;111:753–760.
5. Chan TY, Partin AW, Walsh PC, Epstein JI. Prognostic significance of Gleason score 3 + 4 versus Gleason score 4 + 3 tumor at radical prostatectomy. Urology 2000;56:823–827.
6. Burdick MJ, Reddy CA, Ulchaker J, Angermeier K, Altman A, Chehade N, Mahadevan A, Kupelian PA, Klein EA, Ciezki JP. Comparison of biochemical relapse-free survival between primary Gleason score 3 and primary Gleason score 4 for biopsy Gleason score 7 prostate cancer. Int J Radiat Oncol Biol Phys 2009;73:1439–1445.
7. Stark JR, Perner S, Stampfer MJ, Sinnott JA, Finn S, Eisenstein AS, Ma J, Fiorentino M, Kurth T, Loda M, Giovannucci EL, Rubin MA, Mucci LA. Gleason score and lethal prostate cancer: Does 3 + 4 = 4 + 3? J Clin Oncol 2009;27:3459–3464.
8. Tsao CK, Gray KP, Nakabayashi M, Evan C, Kantoff PW, Huang J, Galsky MD, Pomerantz M, Oh WK. Patients with biopsy gleason 9 and 10 prostate cancer have significantly worse outcomes compared to patients with gleason 8 disease. J Urol 2015;194:91–97.
9. Epstein JI, Zelefsky MJ, Sjoberg DD, Nelson JB, Egevad L, Magi-Galluzzi C, Vickers AJ, Parwani AV, Reuter VE, Fine SW, Eastham JA, Wiklund P, Han M, Reddy CA, Ciezki JP, Nyberg T, Klein EA. A contemporary prostate cancer grading system: A validated alternative to the Gleason score. Eur Urol 2016;69:428–435.
10. Loeb S, Folkvaljon Y, Robinson D, Lissbrant IF, Egevad L, Stattin P. Evaluation of the 2015 Gleason grade groups in a nationwide population-based cohort. Eur Urol 2016;69:1135–1141.
11. Spratt DE, Cole AI, Palapattu GS, Weizer AZ, Jackson WC, Montgomery JS, Dess R, Zhao SG, Lee JY, Wu A, Kunju LP, Talmich E, Miller DC, Hollenbeck BK, Tomlins SA, Feng FY, Mehra R, Morgan TM. Independent surgical validation of the new prostate cancer grade grouping system. BJU Int 2016. doi: 10.1111/bju.13488
12. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA, Grading C. 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;40:244–252.
13. Abdollah F, Suardi N, Gallina A, Bianchi M, Tutolo M, Passoni N, Fossati N, Sun M, dell'Oglio P, Salonia A, Karakiewicz PI, Rigatti P, Montorsi F, Briganti A. Extended pelvic lymph node dissection in prostate cancer: A 20-year audit in a single center. Ann Oncol 2013;24:1459–1466.
14. Abdollah F, Gandaglia G, Suardi N, Capitanio U, Salonia A, Nini A, Moschini M, Sun M, Karakiewicz PI, Shariat SF, Montorsi F, Briganti A. More extensive pelvic lymph node dissection improves survival in patients with node-positive prostate cancer. Eur Urol 2015;67:212–219.
15. Briganti A, Larcher A, Abdollah F, Capitanio U, Gallina A, Suardi N, Bianchi M, Sun M, Freschi M, Salonia A, Karakiewicz PI, Rigatti P, Montorsi F. Updated nomogram predicting lymph node invasion in patients with prostate cancer undergoing extended pelvic lymph node dissection: The essential importance of percentage of positive cores. Eur Urol 2012;61:480–487.
16. Cozzarini C, Montorsi F, Fiorino C, Alongi F, Bolognesi A, Da Pozzo LF, Guazzoni G, Freschi M, Roscigno M, Scattoni V, Rigatti P, Di Muzio N. Need for high radiation dose (> or = 70 gy) in early postoperative irradiation after radical prostatectomy: A single-institution analysis of 334 high-risk, node-negative patients. Int J Radiat Oncol Biol Phys 2009;75:966–974.
17. Dell'Oglio P, Suardi N, Boorjian SA, Fossati N, Gandaglia G, Tian Z, Moschini M, Capitanio U, Karakiewicz PI, Montorsi F, Karnes RJ, Briganti A. Predicting survival of men with recurrent prostate cancer after radical prostatectomy. Eur J Cancer 2016;54:27–34.
18. Nini A, Gandaglia G, Fossati N, Suardi N, Cucchiara V, Dell'Oglio P, Cazzaniga W, Luzzago S, Montorsi F, Briganti A. Patterns of clinical recurrence of node-positive prostate cancer and impact on long-term survival. Eur Urol 2015;68:777–784.
19. Vickers AJ, Sjoberg DD, European U. Guidelines for reporting of statistics in European urology. Eur Urol 2015;67:181–187.
20. Rushing C, Bulusu A, Hurwitz HI, Nixon AB, Pang H. A leave-one-out cross-validation SAS macro for the identification of markers associated with survival. Comput Biol Med 2015;57:123–129.
21. Vickers AJ, Elkin EB. Decision curve analysis: A novel method for evaluating prediction models. Med Decis Making 2006;26:565–574.

22. Mottet N, Bellmunt J, Briers E, Bolla M, Cornford P, De Santis M, Henry A, Joniau S, Lam T, Mason MD, Matveev V, van der Poel H, van der Kwast TH, Rouvière O, Wiegel T. Guidelines on prostate cancer. Eur Urol 2016. https://uroweb.org/guideline/prostate-cancer/
23. D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, Tomaszewski JE, Renshaw AA, Kaplan I, Beard CJ, Wein A. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. JAMA 1998;280:969–974.
24. Berney DM, Beltran L, Fisher G, North BV, Greenberg D, Moller H, Soosay G, Scardino P, Cuzick J. Validation of a contemporary prostate cancer grading system using prostate cancer death as outcome. Br J Cancer 2016;114:1078–1083.
25. Spratt DE, Jackson WC, Abugharib A, Tomlins SA, Dess RT, Soni PD, Lee JY, Zhao SG, Cole AI, Zumsteg ZS, Sandler H, Hamstra D, Hearn JW, Palapattu G, Mehra R, Morgan TM, Feng FY. 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;19:292–297.