Grade Groups Provide Improved Predictions of Pathological and Early Oncologic Outcomes Compared with Gleason Score Risk Groups

Samer Kirmiz, Ji Qi, Stephen K. Babitz, Susan Linsell, Brian Denton, Karandeep Singh,* Gregory Auffenberg, James E. Montie and Brian R. Lane†

From the College of Human Medicine, Michigan State University (SK, BRL) and Spectrum Health Hospital System (SKB, BRL), Grand Rapids, and Departments of Urology (JQ, SL, BD, GA, JEM) and Learning Health Sciences (KS), University of Michigan Medical School, Ann Arbor, Michigan

Abbreviations and Acronyms

ADT = androgen deprivation therapy
BCR = biochemical recurrence
EPE = extraprostatic extension
GG = Grade Group
GS = Gleason score
ISUP = International Society of Urological Pathology
MUSIC = Michigan Urological Surgery Improvement Collaborative
N1 = positive lymph nodes
PCa = prostate cancer
PSA = prostate specific antigen
RP = radical prostatectomy
RT = radiation therapy
SVI = seminal vesical invasion

Purpose: The GG (Grade Group) system was introduced in 2013. Data from academic centers suggest that GG better distinguishes between prostate cancer risk groups than the Gleason score (GS) risk groups. We compared the performance of the 2 systems to predict pathological/recurrence outcomes using data from the MUSIC (Michigan Urological Surgery Improvement Collaborative).

Materials and Methods: Patients who underwent biopsy and radical prostatectomy in the MUSIC from March 2012 to June 2017 were classified according to GG and GS. Outcomes included the presence or absence of extraprostatic extension, seminal vesical invasion, positive lymph nodes, positive surgical margins and time to cancer recurrence (defined as postoperative prostate specific antigen 0.2 ng/ml or greater). Logistic and Cox regression models were used to compare the difference in outcomes.

Results: A total of 8,052 patients were identified. When controlling for patient characteristics, significantly higher risks of extraprostatic extension, seminal vesical invasion and positive lymph nodes were observed for biopsy GG 3 vs 2 and for GG 5 vs 4 ($p < 0.001$). Biopsy GGs 3, 4 and 5 also showed shorter time to biochemical recurrence than GGs 2, 3 and 4, respectively ($p < 0.001$). GGs 3, 4 and 5 at radical prostatectomy were each associated with a greater probability of recurrence compared to the next lower GG ($p < 0.001$). GG (vs GS) had better predictive power for extraprostatic extension, seminal vesical invasion, positive lymph nodes and biochemical recurrence.

Conclusions: GG at biopsy and radical prostatectomy allows for better discrimination of recurrence-free survival between individual risk groups than GS risk groups with GGs 2, 3, 4 and 5 each incrementally associated with increased risk.

Accepted for publication August 29, 2018.

The corresponding author certifies that, when applicable, a statement(s) has been included in the manuscript documenting institutional review board, ethics committee or ethical review board study approval; principles of Helsinki Declaration were followed in lieu of formal ethics committee approval; institutional animal care and use committee approval; all human subjects provided written informed consent with guarantees of confidentiality; IRB approved protocol number; animal approved project number.

Supported by Blue Cross Blue Shield of Michigan, the Betz Family Endowment for Cancer Research and the Spectrum Health Foundation.

No direct or indirect commercial incentive associated with publishing this article.

* Financial interest and/or other relationship with the NIH (National Institutes of Health)/NIDDK (National Institute of Diabetes and Digestive and Kidney Diseases).

† Correspondence and requests for reprints: Urology Division, Spectrum Health Medical Group, 4069 Lake Dr., Suite 313, Grand Rapids, Michigan 49546 (telephone: 616-267-9333; FAX: 616-267-8040; e-mail: brian.lane@spectrumhealth.org).

Editor's Note: This article is the third of 5 published in this issue for which category 1 CME credits can be earned. Instructions for obtaining credits are given with the questions on pages 404 and 405.

0022-5347/19/2012-0278/0
THE JOURNAL OF UROLOGY®

https://doi.org/10.1016/j.juro.2018.08.081
Vol. 201, 278-283, February 2019
Printed in U.S.A.

Key Words: prostatic neoplasms, neoplasm grading, prostatectomy, mortality, risk

SINCE its inception during the 1960s, the GS system has been the most widely accepted grading tool in the evaluation of PCa.[1,2] The GS system has undergone numerous revisions to reduce confusion among patients and more accurately distinguish certain intermediate score tumors.[3–5] Most recently the GG system, based on the traditional GS system, was introduced. This newer system was not only validated but also endorsed by the ISUP after single institutional and multi-institutional validation studies showed its value. Along with its goal to more accurately distinguish differences between respective GS risk groups, it also aims to help in reducing patient anxiety and overtreatment of low grade PCa.[6–10]

In each grading system the grade is determined by needle core biopsy in patients with suspected PCa or by analysis of a RP specimen. The GS evaluates the predominant and the second most prevalent architectural patterns, which are each then assigned a score of 1 to 5. A score of 1 is given for most differentiated patterns and a score of 5 is given for the least differentiated patterns. The final GS is defined as the sum of the 2 most common grade patterns, eg $3 + 3 = 6$, $3 + 4 = 7$, etc. Patterns 1 and 2 are rarely, if ever, used in current practice due to the more recent evolution of pathological grading principles, such that there are almost no current diagnoses of GS less than 6.

The GS system was historically grouped using a 3-tier system (Gleason 6 or less vs 7 vs 8-10). The new GG system assigns 1 of 5 groups based on the core with the worst grade, including Gleason $3 + 3$ (GG 1) vs $3 + 4$ (GG 2) vs $4 + 3$ (GG 3) vs 8 (GG 4) vs 9-10 (GG 5). Additionally, the 3-tier GS system has been an integral part of risk stratification in the NCCN® (National Comprehensive Cancer Network®) and D'Amico classifications (see Appendix).[11–13] Data from academic centers suggest that GG better distinguishes between risk groups than GS risk groups but it is unclear whether these findings can be generalized to community practices.[5,14–16]

We investigated whether data prospectively collected in the MUSIC registry provides validation of the new GG system and what differences are observed for each GG group.

MATERIALS AND METHODS

Michigan Urological Surgery Improvement Collaborative

The MUSIC registry is a statewide, physician led quality improvement consortium. Patient data are entered prospectively by trained medical record data abstractors at respective sites throughout Michigan. Participating practices represent a broad spectrum of academic and community practices, including approximately 90% of the urologists in Michigan. Each MUSIC practice obtained an exemption or approval for collaborative participation from a local institutional review board.

Study Population

For the purpose of this study we included all the patients who underwent needle core biopsy and subsequent RP. Patients were excluded if they received neoadjuvant therapy prior to surgery, received adjuvant therapy (eg RT, ADT) after RP, underwent other definitive PCa treatment (eg RT, ADT or chemotherapy) instead of RP or were surveilled without definitive therapy of PCa. Adjuvant therapy was defined as RT or ADT administered within 1 year after RP with all post nadir PSA levels less than 0.1 ng/ml. Preoperative PSA levels were obtained in all patients and postoperative followup included PSA monitoring. For BCR postoperative PSA draws more than 30 days after surgery were considered. Patients without recorded PSA 30 days after surgery were excluded from study.

Exposure Variable

Patients were classified into 5 GG risk groups labeled GG 1 to 5 based on preoperative biopsy or surgical GG, including GGs 1, 2, 3, 4 and 5, and into low, intermediate and high GS risk groups, including GSs 6, 7 and 8-10.

Outcomes

Primary outcomes of interest included the presence/absence of EPE, SVI, N1 and positive surgical margins. The secondary outcome was time to BCR after surgery, which was defined as postoperative PSA 0.2 ng/ml or greater.[13,17] An additional analysis of time to BCR or to the initiation of adjuvant therapy was also performed to address potential biases from excluding these patients.

Statistical Analysis

Clinical characteristics of patients were compared by GG and GS groups using the chi-square test for categorical variables and the Wilcoxon rank sum test for continuous measures. The rate of each adverse pathology finding and recurrence was summarized by the 2 grouping systems and Kaplan-Meier curves were used to illustrate time to BCR following treatment. Multivariable logistic regression models for adverse pathology and Cox regression models for time to recurrence were constructed to compare the difference in outcomes between 2 adjacent grade groups (ie GG 2 vs 1, GG 3 vs 2, etc), with the Bonferroni correction used to adjust for multiple comparisons. Covariates accounted for during the analysis included patient age, race, preoperative PSA, clinical T stage and the Charlson comorbidity index. All analyses were done using SAS®, version 9.4 with statistical significance considered at $p = 0.05$.

RESULTS

We identified 8,052 men who underwent biopsy and RP in a MUSIC practice between March 2012 and June 2017. Median followup after RP was 19.6 months. Of the 8,052 patients included in our analysis 3,980 (49%) were treated at academic centers and 4,072 (51%) were treated at private practices. Median PSA was 5.8 ng/ml and age was 63.2 years in the entire cohort (supplementary table 1, http://jurology.com/). According to the classic GS system 1,879 patients (23.3%) were at low risk (GS 6), 4,847 (60.2%) were at intermediate risk (GS 7) and 1,326 (16.5%) were at high risk (GS 8 or greater). Supplementary table 1 (http://jurology.com/) lists patient characteristics according to GG and GS.

RP pathological outcomes according to biopsy GG and GS were examined. Significant variations in the rate of adverse pathology findings after RP (EPE, SVI, N1 and positive surgical margins) were observed across GG and GS risk groups (supplementary table 2, http://jurology.com/). For example, the rate of N1 was 0.3%, 2.7% and 12.4% in the GS low, intermediate and high risk groups, respectively ($p < 0.001$). It was 0.3%, 1.7%, 5.1%, 7.7% and 20.2% for GGs 1, 2, 3, 4 and 5, respectively ($p < 0.001$). On multivariable analysis significant differences were observed in the odds of most pathological outcomes when comparing adjacent GG groups. For example, GG 3 compared to GG 2 had significantly higher odds of EPE (OR 1.78), SVI (OR 2.32), N1 (OR 2.41) and predominant pattern 4/5 disease (OR 5.85) (supplementary table 3, http://jurology.com/). Similarly GG 5 compared to GG 4 was associated with a higher risk of EPE (OR 2.22), SVI (OR 2.45) and N1 (OR 2.59), and predominant pattern 4/5 disease (OR 2.95).

Table 1 shows the BCR rate after RP summarized by biopsy GG and GS groups. During the followup of this study 7.2%, 12.6% and 37.4% of patients in the biopsy GS low, intermediate and high groups, respectively, experienced recurrence after RP ($p < 0.001$). Across biopsy GG groups 7.2%, 9.1%, 20.9%, 31.0% and 49.0% of patients had GGs 1, 2, 3, 4 and 5 recurrence, respectively ($p < 0.001$). Figure 1 shows the Kaplan-Meier curve of time to recurrence after RP. When controlling for patient characteristics, all adjacent GG groups except GGs 1 and 2 showed a significant difference in the risk of recurrence (HR 2.06 for GG 3 vs 2, 1.53 for GG 4 vs 3 and 1.69 for GG 5 vs 4, table 2).

A similar pattern was observed for BCR when the evaluation was based on surgical GG and GS groups. Significant variation in the rate of recurrence was seen across GS and GG groups (table 1). On multivariable analysis each GG group showed a higher risk of recurrence compared to its adjacent lower GG group except GG 2 (table 2). Similar results were seen when the initiation of adjuvant therapy was considered along with BCR as a composite end point (supplementary table 4, http://jurology.com/). Significant differences in the risk of adjuvant therapy or BCR were seen in each GG group on these analyses, including biopsy GG 2 vs 1 (HR 1.37, $p = 0.006$) and surgical GG 2 vs 1 (HR 1.46, $p = 0.012$).

To compare GG with GS for the prediction of pathological and oncologic outcomes we examined c-statistics in univariable and multivariable analyses adjusted for age, race, preoperative PSA, clinical T stage and comorbidity. Biopsy GG alone improved prediction compared to GS alone on univariable and

Table 1. *Biochemical recurrence rate by biopsy and surgical GS/GG group*

Group	No. Pts	Median Mos Post-RP Followup (IQR)	p Value	No. Recurrence (%)	p Value
		Biopsy			
GS:					
6	1,801	20.0 (10.4–32.7)	0.61	129 (7.2)	<0.001
7	4,526	19.3 (10.3–32.3)		570 (12.6)	
8-10	1,152	20.1 (10.1–32.4)		431 (37.4)	
GG:					
1	1,801	20.0 (10.4–32.7)	0.90	129 (7.2)	<0.001
2	3,185	19.5 (10.3–32.3)		290 (9.1)	
3	1,341	18.8 (10.3–32.2)		280 (20.9)	
4	742	20.4 (10.3–32.1)		230 (31.0)	
5	410	19.4 (9.8–33.2)		201 (49.0)	
		Surgical			
GS:					
6	1,188	21.3 (11.2–33.3)	0.030	72 (6.1)	<0.001
7	5,405	19.2 (10.2–32.4)		655 (12.1)	
8-10	886	19.3 (10.4–32.2)		403 (45.5)	
GG:					
1	1,188	21.3 (11.2–33.3)	0.097	72 (6.1)	<0.001
2	3,841	19.2 (10.2–32.5)		308 (8.0)	
3	1,564	19.4 (10.2–31.9)		347 (22.2)	
4	387	19.3 (10.4–30.7)		138 (35.7)	
5	499	19.3 (10.4–34.0)		265 (53.1)	

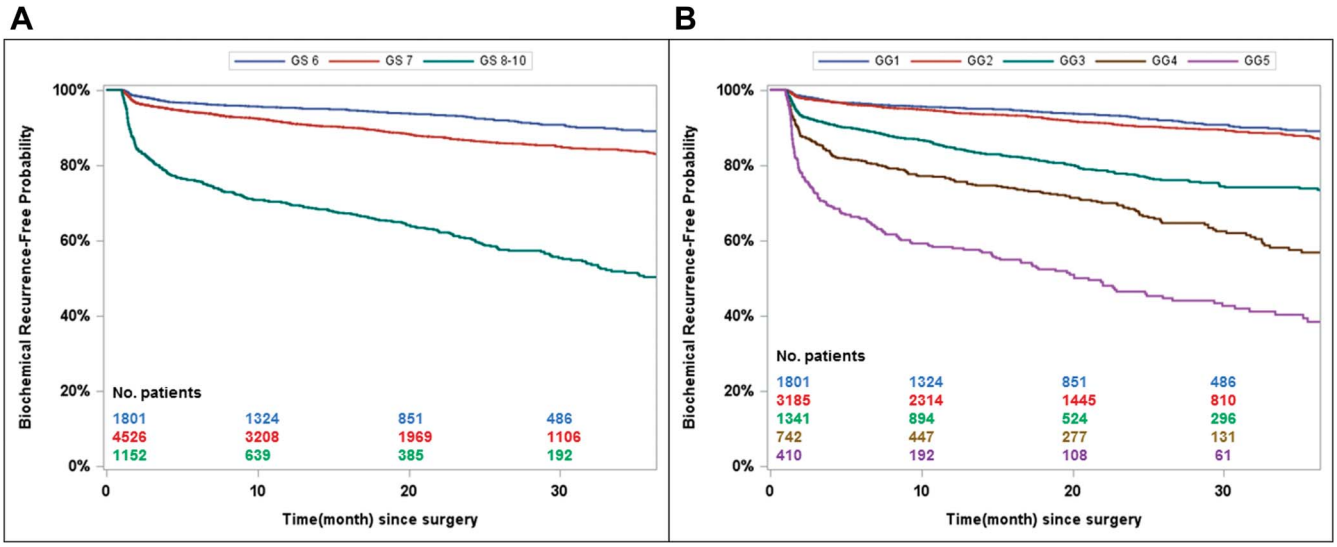

Figure 1. Kaplan-Meier curve of biochemical recurrence stratified by biopsy GS (*A*) and GG (*B*) groups

multivariable analyses (supplementary table 5, http://jurology.com/).

DISCUSSION

Data on diverse community and academic based MUSIC cohorts confirm better discrimination of pathological outcomes at RP with the GG system compared with GS biopsy risk groups. In addition, we found that biopsy GG and RP GG improved the prediction of biochemical recurrence compared with biopsy GS and surgical GS. Our study also revealed that biopsy and surgical GG 1 and 2 are more similar than dissimilar with respect to BCR after RP, a finding which to our knowledge has not been published previously but is consistent with the concept of a lesser risk for relapse associated with smaller amounts of Gleason pattern 4 cancer. The current study provides several additional justifications for using the 5-tier GG system for PCa risk classification.

Heterogeneity is substantial in GS risk groups. Prior data from academic centers indicate that the GG system appears to provide prognostic differences between each of the 5 groups.[16] For all practical purposes the lowest GS on needle biopsy is 6, which is associated with an extremely favorable prognosis. This feature is difficult for patients to understand (6 of 10 is the best).

To recalibrate, GS 6 cancers are GG 1, which is more indicative of the favorable prognosis. In addition, the GS 7 group is separated into GGs 2 and 3, which portend significantly different prognoses (figs. 1 and 2). The nearly overlapping curves of GG 1 and 2 show that these 2 groups performed at the same level with respect to BCR, indicating that the 2 groups not only benefit from treatment but more importantly do equally well following RP. Given this finding, our study may suggest that certain patients with GG 2 at biopsy may benefit from nonsurgical treatment such as surveillance or radiation, similar to those with GG 1 disease. Moreover, GS 8 cancers fare appreciably better than GS 9-10 disease and, thus, they become a separate grade group.[6] A groundbreaking article by Epstein et al in 2016 validated the GG system in 20,845 men after RP.[5] However, the performance of GG in a population based cohort representing academic and community practices has been unknown until the present.

Gleason score 7 tumors have been shown to be significantly heterogeneous in biological behavior and in oncologic outcome.[18] While prior studies had divergent findings, the majority demonstrated worse pathological stage and BCR rates in men with GS 4 + 3 compared with GS 3 + 4 cancers in surgery and biopsy series.[18–20] We observed a similar relationship with significant differences between GGs 2 and 3 with respect to BCR, EPE, SVI and N1 status.

Spratt et al examined the impact on biochemical recurrence-free survival in 3,694 men and noted

Table 2. *Cox regression adjusted HR of biochemical recurrence of adjacent biopsy and surgical GG group adjusted for age, race, preoperative PSA, clinical T stage and comorbidity*

Grade Group	HR	p Value*
Biopsy:		
2 vs 1	1.20	0.398
3 vs 2	2.06	<0.001
4 vs 3	1.53	<0.001
5 vs 4	1.69	<0.001
Surgery:		
2 vs 1	1.17	0.996
3 vs 2	2.70	<0.001
4 vs 3	1.72	<0.001
5 vs 4	1.65	<0.001

* Adjusted for multiple comparisons by Bonferroni correction.

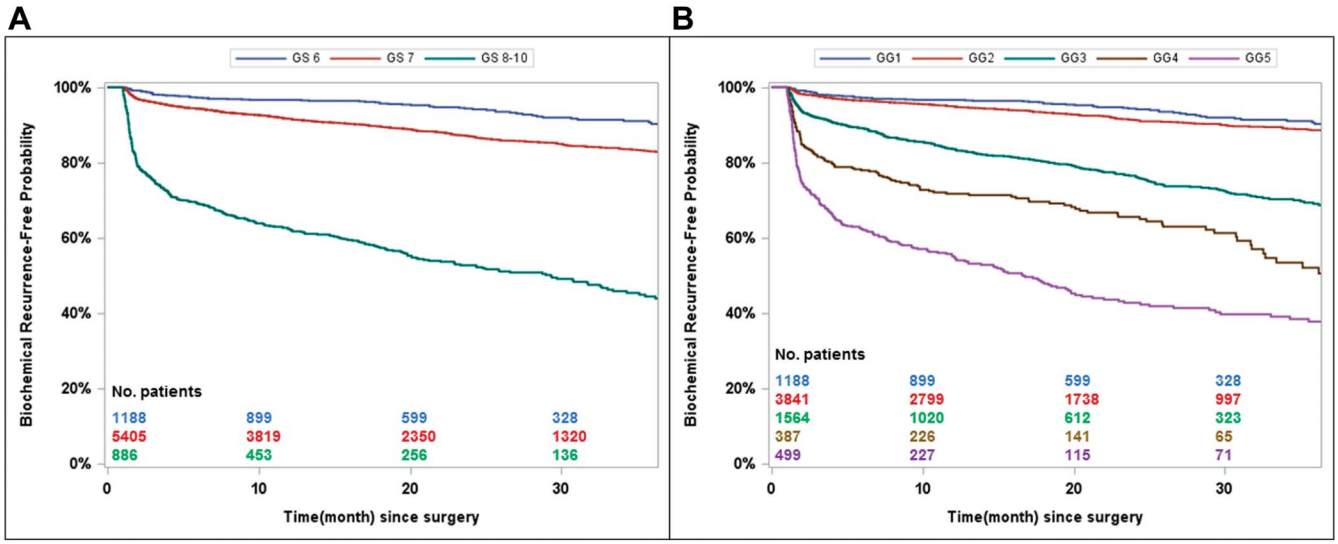

Figure 2. Kaplan-Meier curve of biochemical recurrence stratified by surgical GS (*A*) and GG (*B*) groups

significant differences in patients with GG 2 vs 3 disease at biopsy (89.2% vs 73.1% at 5 years, p <0.001) and at RP (93.0% vs 74.0% at 5 years, p <0.001).[16] We found similar differences in biochemical recurrence-free survival in patients with GG 2 vs 3 at biopsy (89% vs 79% at 3 years, p <0.001) and at RP (90% vs 74% at 3 years, p <0.001). We also found statistically different outcomes with respect to prognosis for the highest risk tumors, GGs 4 and 5. Pierorazio et al also reported that GS 9-10 cancers carry almost twice the risk of progression compared with GS 8 cancer.[6] Our study confirms differences in 3-year biochemical recurrence-free survival in patients with GG 4 PCa (63% and 56% for biopsy and RP) and GG 5 PCa (46% and 43% for biopsy and RP).

A particular strength of our work is the inclusion of biopsies and RP performed at a consortium of urology practices across Michigan, overcoming some of the limitations of prior single institution academic studies.[5,16] Furthermore, respective biopsies and RP in all of these cases were not performed at the same institution and/or graded at the same laboratory, which is another strength as it reflects real world practice. The represented patients and cancers are, therefore, likely more representative of the general population of the United States.

Limitations of this study include the relatively short followup with respect to disease-free survival and the few metastatic events or cancer specific mortalities in the cohort. Pathological processing and evaluation of prostate biopsy and prostatectomy tissue varied across the state. However, this bias must be weighed against the value of greater generalizability by using data from a diverse group of practices.

CONCLUSIONS

The new GG system allows for better discrimination between individual groups than GS using biopsy tissue and RP specimens, specifically because of the separation of GS 3 + 4 (GG 2) from GS 4 + 3 (GG 3) and GS 8 (GG 4) from GS 9-10 (GG 5). Data prospectively collected in the MUSIC registry demonstrate a stepwise increased risk of high risk features at prostatectomy with each biopsy GG tier.

APPENDIX

Prostate Cancer Risk Groups

D'Amico criteria[11,12]	
Low risk	PSA less than 10.0 and Gleason 6 or less and/or cT1-2a
Intermediate risk	PSA greater than 10 and less than 20.0 ng/ml and/or Gleason 7 and/or cT2b
High risk	PSA greater than 20.0 ng/ml and/or Gleason 8-10 and/or cT2c-3a
NCCN Guidelines[13]	
Very low risk	PSA less than 10.0 ng/ml and Gleason 6 or less and/or cT1c.
Low risk	PSA less than 10.0 ng/ml and Gleason 6 or less and/or cT1-2a
Intermediate risk	PSA 10 to 20.0 ng/ml or Gleason 7 or cT2b or cT2c
High risk	PSA greater than 20.0 ng/ml or Gleason 8-10 or cT3a
Very high risk	PSA greater than 20.0 ng/ml or primary Gleason 5 or cT3b-T4a

REFERENCES

1. Gleason DF: Classification of prostatic carcinomas. Cancer Chemother Rep 1966; **50:** 125.
2. Epstein JI: An update of the Gleason grading system. J Urol 2010; **183:** 433.
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:** 1228.
4. 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:** 244.
5. 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:** 428.
6. Pierorazio PM, Walsh PC, Partin AW et al: Prognostic Gleason grade grouping: data based on the modified Gleason scoring system. BJU Int 2013; **111:** 753.
7. Leapman MS, Cowan JE, Simko J et al: Application of a prognostic Gleason grade grouping system to assess distant prostate cancer outcomes. Eur Urol 2017; **71:** 750.
8. Djaladat H, Amini E, Xu W et al: Oncological outcomes after radical prostatectomy for high-risk prostate cancer based on new Gleason grouping system: a validation study from University of Southern California with 3,755 cases. Prostate 2017; **77:** 743.
9. Yeong J, Sultana R, Teo J et al: Gleason grade grouping of prostate cancer is of prognostic value in Asian men. J Clin Pathol 2017; **70:** 745.
10. 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:** 292.
11. D'Amico AV, Desjardin A, Chen MH et al: Analyzing outcome-based staging for clinically localized adenocarcinoma of the prostate. Cancer 1998; **83:** 2172.
12. D'Amico AV, Whittington R, Malkowicz SB et al: Predicting prostate specific antigen outcome preoperatively in the prostate specific antigen era. J Urol 2001; **166:** 2185.
13. National Comprehensive Cancer Network®: NCCN Guidelines® and Clinical Resources. Clinical Practice Guidelines in Oncology: Prostate Cancer (Version 1.2018). Available at https://www.nccn.org/professionals/physician_gls/default.aspx. Accessed August 24, 2018.
14. Rubin MA, Girelli G and Demichelis F: Genomic correlates to the newly proposed grading prognostic groups for prostate cancer. Eur Urol 2016; **69:** 557.
15. He J, Albertsen PC, Moore D et al: Validation of a contemporary five-tiered Gleason grade grouping using population-based data. Eur Urol 2017; **71:** 760.
16. Spratt DE, Cole AI, Palapattu GS et al: Independent surgical validation of the new prostate cancer grade-grouping system. BJU Int 2016; **118:** 763.
17. Sanda MG, Chen RC, Crispino T et al: Clinically Localized Prostate Cancer: AUA/ASTRO/SUO Guideline. Available at https://www.auanet.org/guidelines/clinically-localized-prostate-cancer-new-(aua/astro/suo-guideline-2017). Accessed August 24, 2018.
18. Chan TY, Partin AW, Walsh PC et al: Prognostic significance of Gleason score 3+4 versus Gleason score 4+3 tumor at radical prostatectomy. Urology 2000; **56:** 823.
19. Koontz BF, Tsivian M, Mouraviev V et al: Impact of primary Gleason grade on risk stratification for Gleason score 7 prostate cancers. Int J Radiat Oncol Biol Phys 2012; **82:** 200.
20. Makarov DV, Sanderson H, Partin AW et al: Gleason score 7 prostate cancer on needle biopsy: is the prognostic difference in Gleason scores 4 + 3 and 3 + 4 independent of the number of involved cores? J Urol 2002; **167:** 2440.

EDITORIAL COMMENT

Since its introduction in 1961 (reference 1 in article), the GS system for prostate cancer has remained in wide use with only minor updates in the 1970s and in 2005.[1] Despite its ubiquity GS is not without issue. From the patient perspective GS can be confusing and distressing. For physicians predicting the oncologic risk associated with each GS group can be a similarly inscrutable task. The ISUP GG system, introduced by Pierorazio et al in 2013 (reference 6 in article), organizes GS into 5 simple predictive groups. This study demonstrates the improved prognostic value of GG over GS risk groups and argues that the former more accurately characterizes tumors.

Although the findings of this study are not novel (GG has been independently validated at a number of academic centers (references 5 and 7 in article), it does offer greater generalizability than previously available. This large cohort samples patients treated at community and academic centers at close to a 50:50 ratio and it includes biopsy and radical prostatectomy specimens analyzed in the most up-to-date GS era (prior studies included specimens scored before the 2005 revision). Although the strength of inertia is in favor of the GS nomenclature, this study offers further evidence that the GG system is valid, predictive and applicable across a diverse range of practices.

Molly E. DeWitt-Foy and Robert Abouassaly
Department of Urology
Cleveland Clinic
Cleveland, Ohio

REFERENCE

1. Delhunt B, Miller RJ, Sringley JR et al: Gleason grading: past, present and future. Histopathology 2012; **60:** 75.