EARLY ONLINE RELEASE

Note: This article was posted on the Archives Web site as an Early Online Release. Early Online Release articles have been peer reviewed, copyedited, and reviewed by the authors. Additional changes or corrections may appear in these articles when they appear in a future print issue of the Archives. Early Online Release articles are citable by using the Digital Object Identifier (DOI), a unique number given to every article. The DOI will typically appear at the end of the abstract.

The DOI for this manuscript is doi: 10.5858/arpa.2021-0279-OA

The final published version of this manuscript will replace the Early Online Release version at the above DOI once it is available.
Variance of Tumor Grade at Radical Prostatectomy With Assessment of Each Tumor Nodule Versus Global Grading

Oleksii A. Iakymenko, MD; Laurence M. Briski, MD; Sanoj Punnen, MD; Ivan Nemov, MD; Isabella Lugo, MPH; Merce Jorda, MD, PhD, MBA; Dipen J. Parekh, MD; Mark L. Gonzalgo, MD, PhD; Oleksandr N. Kryvenko, MD

Original Article

Context.—Multifocal prostate cancer at radical prostatectomy (RP) may be graded with assessment of each individual tumor nodule (TN) or global grading of all TNs in aggregate.

Objective.—To assess case-level grade variability between these 2 grading approaches.

Design.—We reviewed 776 RPs with multifocal prostate cancer with 2 or more separate TNs of different Grade Groups (GGs). Two separate grades were assigned to each RP: one based on the TN with the highest grade and a global grade based on the Gleason pattern volumes for all TNs. We then compared the results of these 2 methods.

Results.—The case-level grade changed by 1 or more GGs between the 2 grading methods in 35% (132 of 374) of GG3 through GG5 cases. Twelve percent (37 of 309) of GGs between the 2 grading methods in 35% (132 of 374) of TNs. We then compared the results of these 2 methods.

Conclusions.—The global grading approach resulted in a lower grade in 35% of GG3 through GG5 cases compared with grading based on the highest-grade TN. Two significant risk factors for this discrepancy with a global grading approach occur when the highest-grade TN has a relatively small tumor volume and with the higher number of TNs per RP. The observed grade variability between the 2 grading schemes most likely limits the interchangeability of post-RP multi-institutional databases if those institutions use different grading approaches.

(Arch Pathol Lab Med. doi: 10.5858/arpa.2021-0279-OA)

The Grade Group (GG)/Gleason score (GS) of prostate cancer assigned at radical prostatectomy (RP) predicts the likelihood of biochemical recurrence and prostate cancer–specific mortality.1–3 Currently, there are 2 major different approaches for grading RP specimens: one approach in which each tumor nodule (TN) is graded separately and the highest grade is designated as the overall grade, and another in which all TNs are analyzed simultaneously and given a single global grade as if they were a single tumor. As a result of these opposing methodologies, the 2 approaches can sometimes possibly yield different overall GGs/GSs for the same RP. This is important to point out because many of the current recommendations in reporting prostate cancer grade have been derived from studies using different grading approaches, and thus far there have been no uniform consensus recommendations on which grading method is better for pathologists to follow, and the prevalence of the methods used varies across the globe.1–3,4

Global grading of biopsy specimens was a major topic of discussion at the most recent consensus meetings of the Genitourinary Pathology Society and the International Society of Urological Pathology (ISUP). However, only the ISUP addressed the issue of grading multifocal prostate cancer in RP specimens.5,6 In the premeeting survey, 60% of respondents indicated that they graded each TN within an RP separately, whereas 38% of respondents indicated that they typically reported a global RP grade (combining all TNs within an RP together).5 The group consensus that resulted from this discussion is that (in most instances) a global grade is sufficient for patient management. Of note, the 2005 ISUP consensus conference suggested that the case grade should be assigned based on the dominant TN.7

1. For radical prostatectomy specimens, Gleason score should be assigned to the dominant nodule(s), if present.
2. Where more than one separate tumor is clearly identified, the Gleason scores of individual tumors can be recorded separately, or, at the very least, a Gleason score of the dominant or most significant lesion (highest Gleason score or pT category, if not the largest) should be recorded.

3. For instance, if there is a large Gleason score 4 (2 + 2) transition zone tumor and a separate smaller Gleason score 8 (4 + 4) peripheral zone cancer, both scores should be reported, or, at the very least, the latter score should be reported rather than these scores being averaged.

Despite intuitively expected different grade outcomes between the individual TN and global grading approaches, there have been no studies to date that have examined the frequency of grade discrepancy within the same cohort between these 2 major RP grading methods. Herein, we compare these 2 major grading approaches in a large contemporary RP cohort from a single institution reviewed by a single urologic pathologist.

MATERIALS AND METHODS

All consecutive RP specimens (N = 1629) at one institution between the years 2014 and 2020 were histologically reviewed by one urologic pathologist (senior author). Only those cases with treated prostate multifocal prostate cancer consisting of 2 or more distinct TNs with different individual GGs were included. All cases that did not meet these criteria were subsequently excluded. Patient age and preoperative prostate-specific antigen (PSA) levels were obtained for each case from electronic medical records. We calculated PSA density for each case, defined as the ratio of preoperative PSA and prostate weight without the seminal vesicles.9,10

All RP specimens were weighed without seminal vesicles, fixed overnight in ambient 10% buffered formalin without formalin injection, serially sectioned on 0.3-cm intervals from apex to base, and entirely submitted with preserved orientation in regular-sized injection, serially sectioned at 0.3-cm intervals from apex to base, or 0.4 cm on consecutively submitted serial sections.10,12,13 We graded all TNs considered spatially separate if the distance between them exceeded 0.3 cm in the same plane of section (ie, on the same slide) or 0.4 cm on consecutively submitted serial sections.10,12,13 We graded all TNs separately in accordance with the most recent contemporary recommendations.3,6,14,15 For the purpose of this study, as we do in our routine clinical practice, we followed Genitourinary Pathology Society recommendations and therefore did not factor intraductal prostatic adenocarcinoma into the grading of individual TNs.6 We recorded percentage of Gleason patterns 4 and 5 (GP4 and GP5) using the following scale: less than 5, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100. Tumor volume (TV) was determined by measuring the surface of each TN in square millimeters and multiplying first by 3 mm (the thickness of each gross tissue section) and then by 1.12 (the fixation shrinkage factor).10,12,16 We documented the presence of extraprostatic tumor extension, seminal vesicle invasion, and surgical margin status for each individual TN, as well as regional lymph node status for each case. One genitourinary pathologist (senior author) reviewed all cases.

This study design allowed us to create a database whereby the volume of each Gleason pattern was recorded for each TN. For grading based on the individual TN, a case grade was assigned based on the TN with the highest-grade cancer. A global GS was generated for each RP by combining the Gleason pattern volumes from all distinct TNs within a RP specimen. For example, a case featuring only Gleason patterns 3 and 4 would be globally classified as GS 3 + 4 = 7 if GP4 ranged from 1% to less than 50%, GS 4 + 4 = 7 if GP4 ranged from 50% to 95%, and GS 4 + 4 = 8 if GP4 exceeded 95%. According to the College of American Pathologists recommendations, we restricted a minor higher-grade component (tertiary pattern 5) to only GG2 and GG3 TNs with 3 histologic grade patterns where the higher-grade component constituted less than 5% of TV.8 Otherwise, pattern 5 was included in the overall grade and GG (GS) composition. For each TN composed of 2 grade patterns, the higher-grade component was included in the overall grade even if less than 5%. Alternatively, if the higher-grade component constituted more than 95% of the TN, the lower-grade component was dropped and the GS was composed of only the higher-grade component. However, we recognize that the current prostate grading recommendation from the most recent ISUP consensus is to consider minor tertiary patterns in the setting of 2 Gleason patterns when the higher grade constitutes less than 5%.5 We therefore provide data on how many TNs graded as GG2 (GS 3 + 4 = 7) have a GP4 of less than 5% and would be considered GG1 (GS 3 + 3 + 6 = 6) with tertiary pattern 4 (<5%) according to ISUP guidelines.

We analyzed grading discrepancies for each RP between the 2 grading approaches (grading based on the individual TN versus global grading). All computations were performed in Python 3 (package version 3.8.1). Statistical analysis was performed in RStudio (package version 4.0.2), including Welch t test, \( \chi^2 \) analysis, and univariable logistic regression. Results were considered statistically significant with \( P \) value < .05.

RESULTS

We identified 2438 distinct TNs (median = 3 TNs per RP) among the 1629 RPs that were reviewed. Within this cohort, only 776 RPs met our inclusion criteria in that each contained multifocals (ie, more than 1 TN) with different individual TN GGs. The remaining 853 RPs were excluded from our study for the following reasons: 326 RPs with unifocal cancer (ie, only a single TN [GG1–GG5]), 379 RPs with multifocal cancers of the same GG (212 GG1, 55 GG2, 17 GG3, and 95 GG5), 102 RPs with extensive bilateral disease where separate TNs could not be distinguished, 37 RPs with treated cancer, 6 RPs with vanishing cancer, 1 RP with only peripheral gland adenosis (and no cancer), 1 RP with carcinoid tumor, and 1 RP with small cell carcinoma. In 16 of the 776 RPs included in our study, the highest grade and stage and largest TV were not observed in the same TN.

The median age in our cohort was 63 years (range, 40–85 years). The average PSA density was 0.199 ng/mL (range, 0.01–1.83 ng/mL). Based on grading of individual TNs, the overall RP grade of our cohort included 402 GG2, 193 GG3, 31 GG4, and 150 GG5 cancers (Table). Among GG2 cancers, 23.1% (93 of 402) had GP4 less than 5%. With the global grading approach, the overall RP grade changed in some cases so that the same cohort included 470 GG2, 151 GG3, 58 GG4, and 97 GG5 cancers. We observed GP4 less than 5% in an additional 12% (37 of 309) of GG2 and 0.5% (1 of 193) of GG3 cases by 2 GGs, and 2% (3 of 151) of GG4 cases by 3 GGs. In one case that consisted of a...
Abbreviations: GP4, Gleason pattern 4; [5], minor higher pattern 5 component (tertiary pattern 5).

smaller GG4 TN (GS 3 + 5 = 8, GP5 = 30%, TV = 0.4 cm³) and a larger GG3 TN (GS 4 + 3 = 7, GP4 = 90%, TV = 1.0 cm³), the global grading scheme resulted in an overall grade being interpreted as GG5 (GS 4 + 5 = 9).

The risk of grade discrepancy between grading based on the dominant TN alone versus global grading was associated with older age (odds ratio = 1.05; CI, 1.01–1.08; P = .003), smaller TV of individual highest-grade TNs (odds ratio = 0.52; CI, 0.43–0.61; P < .001), and number of TNs (odds ratio = 1.4; CI, 1.19–1.66; P < .001). Prostate weight, PSA density, and total TV were not associated with risk of grade discrepancy between the 2 grading approaches.

**DISCUSSION**

The grading of RP specimens varies between different laboratories, with some assigning overall grade based on the grade of the highest-grade TN and others assigning a global grade based on combining all TNs within an RP with respect to TV and different Gleason pattern percentages. Although neither system is necessarily better than the other, the lack of a universally agreed-upon approach may create confusion in some cases between institutions and within multi-institutional databases. Although the latest recommendations by the College of American Pathologists are rather straightforward and advocate against a global grading approach, it is still prevalent in contemporary literature.5 To highlight the diversity of approaches, we have identified several significant manuscripts that influenced contemporary prostate cancer understanding but either used different grading approaches or did not describe which method was used. For example, early works on the GG system were performed on RP cohorts with grade assigned based on the highest-grade TN.1,2 Several validation studies assessing the prognostic significance of GG in large databases of RP specimens provide no description of which grading system was followed.7,8,20 Other studies (such as one by Sauter et al9) demonstrate the significance of the proportion of the higher-grade component in each GG when the global grading approach is used. In our study, we show that many cases (with one anecdotal exception) graded using the global approach result in a lower overall grade compared with grading each TN individually and assigning an overall grade based on the highest-grade TN. In fact, 35% of GG3 through GG5 cases in our cohort were downgraded when the global grading approach was applied. Furthermore, in some GG5 cases based on grading of the dominant TN, the GP5 decreased to less than 5%, resulting in 5 GG3 and 3 GG2 cases with a minor higher-grade component (tertiary pattern 5) when graded using the global approach.

The recommendations of the Genitourinary Pathology Society and the ISUP also notably differ in their approaches regarding minor high-grade Gleason patterns and how to incorporate them into the overall RP grade.5,6,21 Take, for example, the presence of GP4 less than 5% in what is otherwise a GS 3 + 3 = 6 prostate cancer. The ISUP recommendation5 would be to designate such a tumor as GG1 (GS 3 + 3 = 6) with a tertiary pattern 4. In contrast, the Genitourinary Pathology Society recommendation would be to designate such a tumor as GG2 (GS 3 + 4 = 7) and specify the GP4 percentage based on the recognition that GS 3 + 3 = 6 prostate cancers with a tertiary pattern 4 (GP4 < 5%) have been shown to be associated with worse clinical outcomes compared with pure GG1 (GS 3 + 3 = 6) prostate cancers.16,22–24

It is not within the scope of our study to pass judgment on either of these 2 approaches, but rather to highlight the implications and heretofore unforeseen consequences of having 2 competing sets of recommendations for RP grading and question if the data generated by such are interchangeable. Although neither approach may be superior to the other, significant grade discrepancies between the 2 approaches may result in major changes in the approach to patient counseling and management. In our cohort of 402 cases with GG2 cancer, 23.1% (93 of 402) of cases had GP4 less than 5% based on individual TN grading. When the global grading approach was applied to this same cohort, an additional 9.2% (37 of 402) of cases decreased their GP4 to less than 5%. This poses a potential challenge for drawing inferences from large multi-institutional databases that include institutions using different grading schemes if those institutions do not report the grading methodology used or the fractions of Gleason patterns present in a given RP.25,26

---

**Outcomes in Each Grade Group (GG) With Individual Tumor Nodule (TN) and Corresponding Global Grading**

<table>
<thead>
<tr>
<th>Highest-Grade TN GG</th>
<th>Highest-Grade TN Gleason Score</th>
<th>Global GG</th>
<th>Global Gleason Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG2 (n = 402)</td>
<td>3 + 4 = 7, GP4 &gt;5% (n = 309)</td>
<td>GG2 (n = 402)</td>
<td>3 + 4 = 7, GP4 &gt;5% (n = 272)</td>
</tr>
<tr>
<td></td>
<td>3 + 4 = 7, GP4 &lt;5% (n = 93)</td>
<td>GG2 (n = 402)</td>
<td>3 + 4 = 7, GP4 &lt;5% (n = 37)</td>
</tr>
<tr>
<td>GG3 (n = 193)</td>
<td>4 + 3 = 7 (n = 193)</td>
<td>GG3 (n = 134)</td>
<td>4 + 3 = 7 (n = 134)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GG2 (n = 59)</td>
<td>3 + 4 = 7, GP4 &gt;5% (n = 58)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GG5 (n = 1)</td>
<td>3 + 4 = 7, GP4 &lt;5% (n = 1)</td>
</tr>
<tr>
<td>GG4 (n = 31)</td>
<td>4 + 4 = 8 (n = 20)</td>
<td>GG4 (n = 5)</td>
<td>4 + 4 = 8 (n = 5)</td>
</tr>
<tr>
<td></td>
<td>3 + 5 = 8/5 + 3 = 8 (n = 11)</td>
<td>GG3 (n = 10)</td>
<td>4 + 3 = 7 (n = 10)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GG2 (n = 5)</td>
<td>3 + 4 = 7, GP4 &gt;5% (n = 5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GG5 (n = 1)</td>
<td>4 + 5 = 9 (n = 1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GG4 (n = 7)</td>
<td>4 + 5 = 8 (n = 7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GG3 (n = 2)</td>
<td>4 + 3 = 7 [5] (n = 2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GG2 (n = 1)</td>
<td>3 + 4 = 7 [5] (n = 1)</td>
</tr>
<tr>
<td>GG5 (n = 150)</td>
<td>9–10 (n = 150)</td>
<td>GG5 (n = 96)</td>
<td>9–10 (n = 96)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GG4 (n = 46)</td>
<td>3 + 5 = 8/5 + 3 = 8 (n = 46)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GG3 (n = 5)</td>
<td>4 + 3 = 7 [5] (n = 5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GG2 (n = 3)</td>
<td>3 + 4 = 7 [5] (n = 3)</td>
</tr>
</tbody>
</table>

Abbreviations: GP4, Gleason pattern 4; [5], minor higher pattern 5 component (tertiary pattern 5).
Similar challenges exist for clinical assessment of minor tertiary pattern 5 (GS $<5\%$) recognized by both urologic pathology organizations. Some studies have shown that there is intermediate biochemical recurrence-free survival associated with GG2 and GG3 cancers with tertiary pattern 5 compared with the same GS without a minor tertiary pattern and the next higher GS when each TN was graded separately at RP.22,23 Jang et al27 observed similar results for GS 7 with a GS$5 less than 5\%$ cutoff when a global grading approach was used. However, other studies using a global grading approach have suggested that there are significant prognostic implications for a minor tertiary pattern 5 only when its fraction exceeds $5\%$.28,29 Such conflicting results might be explained by the different fractions of included multifocal prostate cancers and the volume of the highest-grade TN, which is inversely correlated with the risk of grade change.

There is also a higher incidence of rare GS combinations seen with the global RP grading approach compared with grading each TN individually. In our cohort, grading of individual TNs resulted in only 1 case graded as GS 5 + 3 = 8. When we applied the global grading method to the same cohort, an additional 3 cases (4 cases total) were graded as GS 5 + 3 = 8. Each of these 3 cases was initially graded as GS 5 + 4 = 9 using the individual TN grading approach based on the highest-grade TN. However, because each of these 3 RP's contained separate sufficiently large nondominant GS 3 + 3 = 6 TNs, the global grading method resulted in a downgrading to this seldom-encountered GS combination. In a recent study, we reviewed 24 RP's from different institutions that were reported in the past as GS 5 + 3 = 8.30,31 When the most recent contemporary grading recommendations were applied, none of these 24 cases were assigned a grade of GS 5 + 3 = 8. In this series, common grade assignment to different TNs in multifocal disease was a major reason for reclassification upon rereview with a case grade assessment on the individual TN.

Defining the key element(s) that constitute(s) a dominant TN—TV, pathologic stage, histologic grade, etc—has been a controversial topic and is something that is not universally agreed upon when the most adverse findings do not occur in the same TN. Consequently, grading of multifocal prostate cancer may pose a challenge for pathologists if the TN with the highest pathologic stage (≥pT2+ or greater) TV is associated with a lower grade than a lower-pathologic-stage and/or smaller-TV TN.31 Although risk of adverse surgical outcomes following RP related to TV in contemporarily graded cohorts is a subject to be explored in future inquiries, one might suspect to find an increased incidence of extraprostatic tumor extension, seminal vesicle invasion, and surgical margin positivity in larger tumors, whereas regional (pelvic) lymph node involvement and distant metastasis would more likely be associated with tumors of higher grade and possibly volume. In a recent study including RP specimens only with GG2 and GG3 cancers, TV was a significantly stronger predictor of extraprostatic tumor extension, seminal vesicle invasion, and surgical margin positivity than the percentage of GP4.32 Therefore, it is reasonable to separately grade and assign adverse surgical outcomes to each corresponding TN as opposed to separating TNs into those that are dominant and those that are secondary in cases where the most adverse finding are not observed in the same TN, as proposed by Huang et al,33 and such an approach is similar to the current College of American Pathologists recommendations.4 In our cohort, only 2.1% (16 of 776) of cases had an organ-confined higher-grade TN, whereas extraprostatic tumor extension, seminal vesicle invasion, and/or SM+ were observed in a lower-grade TN and none of these specimens showed lymph node metastasis.

The strongest risk factors for grade discrepancy between the 2 grading methods were a smaller TV of individual highest-grade TN and higher number of separate TNs. It is intuitive to believe that the 2 variables can be interrelated, as cases with multiple TNs would have smaller individual TN volumes. Thus, the total cumulative volume of all lower-grade TNs can overcome that of the highest-grade TN, resulting in grade discrepancy between the 2 methods. Older patient age was marginally associated with grade discrepancy between the 2 methods. We could not identify a definitive explanation for the association between patient age and the likelihood of grade discrepancy.

Finally, the assignment of one grade to multifocal prostate cancer by grouping together all TNs at RP seems to be a practice contrasting with guidelines in other organ systems where multifocal disease can be observed. Donald Gleason, MD, developed the first reproducible grading system for prostate cancer on cohorts with symptomatic advanced disease where distinguishing between separate TNs was technically impossible and clinically irrelevant. However, the subsequently adopted widespread screening and early detection of prostate cancer introduced the pathologic concept of insignificant prostate cancer that is based on assessment of individual TNs.3,34,35

Our study is the first centralized review and detailed analysis to illustrate differences between grading of individual TNs and global grading in a large contemporary RP cohort. Although we did not assess risks of discrepant grading for GG3 through GG5 cases, data presented herein for GG2 prostate cancer that had GP4 less than 5% (when applying the global grading method) are very likely to be indicative of the trends that would be seen in higher-grade cancers. Our principal objective was to assess if the 2 major RP grading systems will generate comparable grade results for the same RP cohort because of differences in methodology. However, the utmost question to be answered is if this difference in grading has an impact on biochemical recurrence–free survival and cancer-specific mortality. As our cohort includes more recent cases, we could not yet assess this because of a limited follow-up time. Finally, we acknowledge that although partial gland submission (which has been deemed acceptable for current practice31) may introduce another variable leading to grading discrepancies, this would have been difficult to compute and analyze within the framework of our study, as all RP's included in our study were entirely submitted for histologic examination. However, we previously demonstrated the inferiority of partial submission relative to complete submission in a minor subset of cases with respect to grading and staging of prostate cancer.36

**CONCLUSIONS**

Individual TN and global grading methods of multifocal prostate cancer in RP specimens have significantly different pathologic results. Downgrading of up to 35% of GG3 through GG5 cases with global grading compared with grading of individual TNs per RP may be observed. Two significant risk factors for this discrepancy with the global grading approach occur when the highest-grade TN has a
relatively small TV and with the higher number of TNs per RP. This has significant implications for multi-institutional databases as well as for comparing different databases if the involved institutions/databases are using different grading approaches. Clinical observations generated on large-cohort data can possibly be incomparable if different grading modalities are used. This can be even further complicated with different reporting practices of a minor pattern 4 component (<5%; inclusion in GS sum or reporting as a tertiary pattern) and/or if partial RP submission is used. With such a significant grade variation between the 2 methods, it seems important for clinical reports and scientific works to specify which grading approach is used in order to be able to stratify risk based on comparatively graded cases, facilitate cross-study comparisons, and make consensus recommendations at large.

References


