

Interobserver Variability of Gleason Score and Completeness of Histopathology Report in Prostatic Adenocarcinoma in Prostate Needle Biopsy Specimens among General Pathologists in a Multi-institutional Setting

Anne Lizbeth Valdez and Jeffrey So

St. Luke's Medical Center, Quezon City, Philippines

ABSTRACT

Introduction. Gleason score, the most widely used grading system for prostatic adenocarcinoma, is the most powerful predictor of patient's clinical outcome and is used to customize treatment strategies. It possesses an inherent degree of subjectivity, as inter-observer and intra-observer variability does exist. Moreover, there are currently no structured histopathology report guidelines for prostate needle biopsies in our setting, making relevant information overlooked by pathologists and interpretation of report between laboratories challenging.

Objective. With these in mind, we sought to study the interobserver variability of Gleason score and completeness of histopathology report in prostate needle biopsy specimens.

Methodology. A set of 19 prostate needle biopsy slides was sent to 18 general pathologists from different institutions in the Philippines for histopathologic analysis of Gleason scores and completeness of reporting. The interobserver agreement of each pathologist will be evaluated using Spearman's rank correlation coefficient.

Results. Overall, there was moderate correlation between the interobserver's Gleason score and Gleason grade group. Low to moderate correlation was seen in primary grade while negligible correlation was seen in secondary grade. Best agreement was seen in poorly differentiated neoplasms. Undergrading was more common than overgrading. Most respondents gave an incomplete histopathology report.

Conclusion. There is an overall moderate correlation between Gleason score. A non-standardized histopathology report is currently used, leaving out relevant histopathologic findings.

Key words: prostate, prostate cancer, urology

ISSN 2507-8364 (Online)

Printed in the Philippines.

Copyright© 2018 by the PJP.

Received: 11 July 2018.

Accepted: 2 September 2018.

Published online first: 16 September 2018.

<https://doi.org/10.21141/PJP.2018.010>

Corresponding author: Anne Lizbeth B. Valdez, MD

E-mail: anne_lizbeth@yahoo.com

ORCID: <https://orcid.org/0000-0001-5863-5875>

INTRODUCTION

Gleason score is the most widely used grading system for prostatic adenocarcinoma. Inevitably, like all other grading systems, it is flawed by some degree of interobserver and intraobserver variability.¹ Although this grading system has undergone significant revisions for the past years, it still continues to have deficiencies that can potentially impact patient care.

Gleason score is the most powerful predictor of patient's clinical outcome and is a major determinant in customizing treatment strategies that is most appropriate for a patient. It is utilized to tailor-fit post biopsy treatment, plan for the type of radiation therapy and whether to administer hormonal therapy with radiation therapy. Patients with Gleason scores of <6 may benefit from watchful waiting and surveillance as initial management.¹ The presence of high-grade Gleason pattern (Gleason pattern 4 or 5) harbors the greatest risk for metastasis and treatment failure. Thus, discordance in Gleason scoring, albeit small, will have a dramatic effect on risk stratification and clinical management.



It has been observed that general pathologists more frequently underscore than over score, with a natural tendency to assign low Gleason pattern in such small core needle biopsies. In a study done by RV Singh et al.,² Gleason score 7 was identified as an area of difficulty as 14 of 63 readings (22%) were underscored. The differences centered on the assessment of small areas of fused and separate glands and fused small irregular glands. This has led to the inappropriate assignment of Gleason score 6 and probable suboptimal patient management as a consequence. In the same study, assignment of Gleason pattern 4 and 5 as distinction between few tiny poorly formed glands versus cords and nests of malignant cells were particularly challenging. As a result, sheets of cells with ill-defined lumina were inappropriately given as Gleason pattern 5 instead of pattern 4. These discrepancies suggest that misperceptions among each Gleason pattern in the scheme exist, especially for “borderline” cases, which exhibit features intermediate between 2 patterns. In another study by Coard,³ the greatest discordance is seen in distinguishing Gleason score 6 from 7 in biopsy specimens with less than 30% tumor volume. This has led to the conclusion that assignment of Gleason scores in core needle samples, in contrast to TURP and radical prostatectomy specimens, poses a diagnostic dilemma as these samples contain low tumor volume.^{4,5} Several data support that for needle biopsy grading, pathologist training and experience can influence the degree of interobserver agreement.^{6,7} In one study,⁷ 41 general pathologists exhibited moderate interobserver agreement with a kappa coefficient of 0.435, while substantial interobserver agreement with a kappa coefficient of 0.6-0.7 was seen among 9 of 10 urologic pathologists. Interest in urologic pathology, particularly in Gleason scoring, resulted in participation of general pathologists in educational courses and subspecialty training, which however is not readily available in our setting. Other sources of grading variation in core needle samples include difficulty in appreciation of infiltrative growth pattern, tissue sampling error and artifactual tissue distortion.

A structured histopathology report for prostate needle biopsies has an essential role in conveying the result to clinicians. The report should be uniform and formatted to provide compete, clear and unambiguous data. The inclusion of tumor volume and presence of extraprostatic extension, perineural and lymphovascular invasion, prostatic intra-epithelial neoplasia and intraductal carcinoma in prostate needle biopsy reports are equally essential as the Gleason score, and must be reported when present since these are associated with adverse clinical outcome.⁸ Moreover, these pathologic findings are being utilized in common nomograms used to guide clinical decision making and therefore must be reported when present. In one study by Kryvenko et al.,¹ analysis of needle biopsy cores showed that the number of positive cores, tumor volume and perineural invasion predicts presence of extraprostatic extension, seminal vesicle invasion and positive surgical margins in radical prostatectomy specimens. In the same study, they concluded that biopsy specimens with perineural invasion is significantly associated biochemical recurrence.

With these in mind, our study intends to 1) determine the interobserver agreement of the respondent pathologists in Gleason grading of prostatic adenocarcinoma in terms of: primary grade, secondary grade, Gleason score and Gleason Grade Group; and 2) describe the completeness of reporting of histopathology results by respondent pathologists in terms of inclusion of tumor volume and mention of presence of extraprostatic extension, perineural and lymphovascular invasion, prostatic intra-epithelial neoplasia and intraductal carcinoma.

METHODOLOGY

Board certified fellows or diplomates in anatomic pathology by the Philippine Society of Pathologists who acquired no formal training in uropathology and practicing as a general pathologist were recruited for this study. Information on respondents' age, number of years in practice, current affiliation/s and other demographic profiles were not collected. They were invited to take part in the study via phone calls, letters and emails. Our study welcomed 18 pathologists from all over the Philippines, including areas outside Metro Manila such as Ilocos Norte, Cagayan, Isabela, Zamboanga, Cebu and Davao. A set of 19 slides diagnosed by a uropathologist with prostatic adenocarcinoma at St. Luke's Medical Center Quezon City was sent to the respondent pathologists. These cases were seen by a second pathologist from the same institution who concurred with the diagnosis. The slides were selected by the original sign-out pathologist to roughly represent the spectrum of Gleason scores based on the 2015 Modified Gleason Grading System and no effort was made to select particularly difficult cases. The slides, in hematoxylin and eosin preparation, was of uniform and adequate quality and was assessed prior to shipping to ensure proper and easeful examination. Also sent along with the slides was a copy of the questionnaire and endorsement letter.

The questionnaire had assigned codes (P1-P18) to maintain the respondent's anonymity while the endorsement letter contained a brief description of the study. Each slide was given a code number (1-19) to maintain patient's anonymity and to ensure that these could not be identified by the respondent pathologists. Each respondent was instructed to give a complete diagnosis as they normally would with their own cases. He/she reviewed the slides without the knowledge of the previous Gleason scores. The interobserver agreement was evaluated using Spearman's rank correlation coefficient. Agreement was calculated for primary grade, secondary grade, Gleason score and Gleason grade group (based on 2015 ISUP and 2016 WHO grading system). The completeness of reporting of each pathologist was evaluated by the mention or failure to mention of tumor volume, extraprostatic extension, perineural and lymphovascular invasion, prostatic intra-epithelial neoplasia and intraductal carcinoma. Institutional Review and Ethics Research Committee approval was secured prior to the commencement of this study.

RESULTS

To assess for interobserver agreement, a mathematical consensus was first calculated (Table 1). The overall percentage of Gleason score agreement for all respondents is 43.0% (10.5% to 68.4%) (Table 2). The maximum number of readings were in the Gleason score 7 (33.9%; n=78/342) and least in Gleason score 2-4 (3.2%; n=11/342).

The distribution of percentage agreement for Gleason score with consensus score was computed (Table 3). 43%

(n=147/342) of all assigned Gleason scores were in exact agreement with the consensus score. 72.8% and 83.6% of the assigned Gleason score were within ±1 and ±2 of the consensus score, respectively. Agreement was best in Gleason 9 (75%; n=25/33) and worst with Gleason 3 (0%; n=0/18) and Gleason 8 (30%; n=30/100). Overall, undergrading was seen in 30.4% while overgrading was seen in 26.9% of the readings. Most commonly undergraded is Gleason score 8 (46/100; 46%) while Gleason score 6 is most commonly overgraded (43/114; 38%).

Interobserver Spearman’s rank correlation coefficient for primary grade, secondary grade Gleason score and Gleason grade group were computed (Table 4). Majority had moderate to low correlation (64.7%; n=198/306) in the primary grade while majority had negligible correlation (61.4%; n=188/306) for secondary grade. Likewise, moderate correlation (35.9%; n=110/306) was seen in the majority of the Gleason scores and moderate correlation (39.2%; n=120/306) with the Gleason grade group.

A total of 8 respondents (44.4%; n=8/18) mentioned at least 1 other histopathologic finding (Table 5).

DISCUSSION

Agreement was best seen in Gleason score 9. This is may be due to the straightforward identification of sheets, cords and solid nests of infiltrative neoplastic cells and necrosis and the large tumor volume of such poorly differentiated neoplasms.

Table 1. Mathematical consensus score per slide

	Median of primary score	Median of secondary score	Mathematical consensus score
Slide 1	3	3.5	7
Slide 2	3	3	6
Slide 3	3	3	6
Slide 4	3	3	6
Slide 5	4	4	8
Slide 6	2	1	3
Slide 7	3	4	7
Slide 8	4	4	8
Slide 9	3	4	7
Slide 10	3	3	6
Slide 11	4	4	8
Slide 12	3	3	6
Slide 13	5	4	9
Slide 14	4	4	8
Slide 15	4	4	8
Slide 16	3	3	6
Slide 17	5	4	9
Slide 18	4	4	8
Slide 19	4	3	7

Table 2. Percent agreement with Gleason score

Respondents	Gleason scores					Total number of readings	Percent agreement with consensus
	0-1	2-4	5-6	7	8-10		
1	0	0	3	9	7	19	47.4
2	0	0	6	7	6	19	57.9
3	0	4	4	8	3	19	21.1
4	3	1	9	6	0	19	26.3
5	0	0	1	9	9	19	36.8
6	3	0	4	8	4	19	47.4
7	2	0	4	2	11	19	47.4
8	0	0	3	4	12	19	36.8
9	0	0	0	13	6	19	47.4
10	2	0	6	5	6	19	68.4
11	4	0	8	3	4	19	31.6
12	3	0	0	5	11	19	36.8
13	3	0	0	7	9	19	42.1
14	1	0	4	6	8	19	52.6
15	1	0	4	6	8	19	52.6
16	2	0	7	7	3	19	52.6
17	0	6	9	3	1	19	10.5
18	1	0	6	8	4	19	57.9
Total	25	11	78	116	112	342	43.0

Table 3. Distribution of percentage of agreement of Gleason scores

Consensus Gleason score	Number of reading with							Total number of reading
	<-3	-2	-1	Exact	+1	+2	>+3	
3	9	0	0	0	0	0	9	18
6	17	2	4	48	28	10	5	114
7	6	3	8	44	8	4	4	77
8	4	6	36	30	16	8	0	100
9	2	4	2	25	0	0	0	33
Total	38	15	50	147	52	22	18	342
Percentage	11.1	4.4	14.6	43	15.2	6.4	5.3	

Table 4. Spearman’s rank correlation coefficient for primary score, secondary score, Gleason score and Gleason grade group

Correlation	Value	Primary Grade		Secondary grade		Gleason score		Gleason grade group	
		Number of readings	%	Number of readings	%	Number of readings	%	Number of readings	%
Very high	0.9-1.00	2	0.7	2	0.7	4	1.3	6	2
High	0.7-0.89	30	9.8	4	1.3	26	8.5	42	13.7
Moderate	0.5-0.69	96	31.4	28	9.2	110	35.9	120	39.2
Low	0.3-0.49	102	33.3	84	27.5	96	31.4	86	28.1
Negligible	0.0-0.29	76	24.8	188	61.4	70	22.9	52	17
		306	100	306	100	306	100	306	100

Table 5. Mention or failure to mention of other pertinent histopathologic findings*

Respondent	Mention of other histopathologic findings*
1	Mention
2	No mention
3	No mention
4	No mention
5	No mention
6	Mention
7	Mention
8	Mention
9	Mention
10	No mention
11	Mention
12	No mention
13	No mention
14	No mention
15	No mention
16	Mention
17	No mention
18	Mention
Total	8 (n=8/18; 44.4%)

*Tumor volume, extraprostatic extension, perineural and lymphovascular invasion, prostatic intra-epithelial neoplasia and/or intraductal carcinoma

Predictably, underscoring is seen more often than overscoring. Literature has supported the fact that there is a natural tendency to underscore in such small specimens, most especially for low tumor volume cores and is may be due to the difficulty in appreciating the infiltrative nature of the tumor.

In contrast, overscoring of consensus score 7 was seen and is may be due to the challenging distinction between subtle differences in poorly formed glands and well-formed glands and/or the loss of acinar spaces caused by compression artifact. There is moderate to low correlation between the primary grades and negligible correlation between the secondary grades. This is because of the problems faced in determining the predominant pattern present in one core.

The presence of 2 distinct patterns in seemingly equal proportions and/or the discontinuous arrangement of neoplastic cells complicate the assignment of a primary grade. The most striking observation for consensus score, however, is the presence of Gleason score <6, which is traditionally not assigned to needle biopsy specimens using the upgraded Gleason grading system. This ascertains that some pathologists are indeed still using the outdated Gleason scoring system.

Majority of the histopathology reports were incomplete. This indicates that a non-standardized histopathology report is still currently being used which makes interpretation of report between institutions challenging.

CONCLUSION

Overall, tumor heterogeneity giving rise to various patterns/mimickers and the presence of morphologically borderline tumors complicates Gleason scoring. We strongly believe that subjectivity will always be present in any grading system and that a good agreement can only achieved by understanding the definition of each pattern in the scheme, as well as the pitfalls, in the updated Gleason grading system. In addition, our study puts emphasis that a complete histopathologic report is an important contributor to the success of patient management. The need to identify relevant histopathologic findings, which are often, overlooked greatly impact patient management.

ACKNOWLEDGMENTS

The authors extend their gratitude to the consultants of St. Luke’s Medical Center Quezon City Institute of Pathology, all the respondent pathologists of this study for sharing their knowledge, and to the staff of the Histopathology Section of St. Luke’s Medical Center Quezon City for their kind assistance during the collection of test materials.

STATEMENT OF AUTHORSHIP

All authors certified fulfillment of ICMJE authorship criteria.

AUTHOR DISCLOSURE

The authors declared no conflict of interest.

FUNDING SOURCE

None.

REFERENCES

1. Kryvenko ON, Epstein JI. Prostate cancer grading: a decade after the 2005 modified Gleason grading system. Arch Patho Lab Med. 2016;140(10):1140-52. PMID: 26756649. <https://doi.org/10.5858/arpa.2015-0487-SA>.
2. Singh RV, Agashe SR, Gosavi AV, Sukhyan KR. Interobserver reproducibility of Gleason grading of prostatic adenocarcinoma among general pathologists. Indian J Cancer. 2011;48(4):488-95. PMID: 22293266. <https://doi.org/10.4103/0019-509X.92277>.
3. Coard KC, Freeman VL. Gleason grading of prostate cancer: level of concordance between pathologists at the University Hospital of the West Indies. Am J Clin Pathol. 2004;122(3):373-6. PMID: 15362366. <https://doi.org/10.1309/3A32-DTVM-H640-M2QA>.

4. Barqawi AB, Turcanu R, Gamito EJ, et al. The value of second-opinion pathology diagnoses on prostate biopsies from patients referred for management of prostate cancer. *Int J Clin Exp Pathol.* 2011;4(5):468-75. PMID: 21738818. PMCID: PMC3127068.
5. Majoros A, Szász AM, Nyirády, et al. The influence of expertise of the surgical pathologist to undergrading, upgrading and understaging of prostate cancer in patients undergoing subsequent radical prostatectomy. *Int Urol Nephrol.* 2014;46(2):371-7. PMID: 23990495. <https://doi.org/10.1007/s11255-013-0538-4>.
6. Allsbrook WC Jr., Mangold KA, Johnson MH, et al. Interobserver reproducibility of Gleason grading of prostatic carcinoma: urologic pathologists. *Hum Pathol.* 2001;32(1):74-80. PMID: 11172298. <https://doi.org/10.1053/hupa.2001.21134>.
7. Allsbrook, WC Jr., Mangold, KA, Johnson MH, Lane RB, Lane CG, Epstein JI. Interobserver reproducibility of Gleason grading of prostatic carcinoma: general pathologist. *Hum Pathol.* 2001;32(1):81-8. PMID: 11172299. <https://doi.org/10.1053/hupa.2001.21135>.
8. Prostate cancer (core/needle biopsy) structured reporting protocol, 1st ed. NSW, Australia: RCPA, 2014. Available at: <https://authorzilla.com/d6eZa/microsoft-word-v0-24-prostate-core-biopsy-post-oc.html>.

Disclaimer: This journal is **OPEN ACCESS**, providing immediate access to its content on the principle that making research freely available to the public supports a greater global exchange of knowledge. As a requirement for submission to the PJP, all authors have accomplished an **AUTHOR FORM**, which declares that the ICMJE criteria for authorship have been met by each author listed, that the article represents original material, has not been published, accepted for publication in other journals, or concurrently submitted to other journals, and that all funding and conflicts of interest have been declared. Consent forms have been secured for the publication of information about patients or cases; otherwise, authors have declared that all means have been exhausted for securing consent.