

Inter-Observer Variability in Gleason Scoring System For Histological Grading of Adenocarcinoma Prostate

Farah Ahsan, Ahmed Ahson Khan, Muhammad Asif, Maria Aslam, Saba Hamayun, Hafeez-Ud-Din

Armed Forces Institute of Pathology/National University of Medical Sciences (NUMS) Rawalpindi Pakistan

ABSTRACT

Objective: To compare the variability between observers while grading adeno carcinoma prostate when using the Gleason scoring system and checking its reproducibility between the observers.

Study Design: Prospective double-blind study.

Place and Duration of Study: Histopathology department, Armed Forces Institute of Pathology, Rawalpindi Pakistan, from Jan to Jul 2020.

Methodology: A total of 128 specimens of adeno carcinoma of prostate were collected and evaluated again by using modification by Epstein of the Gleason system done in 2005. Two pathologists performed the assessment of the specimens by the double-blind method. The level of similarity overlaps or concordance between both the pathologists was calculated using the kappa co-efficient.

Results: The level of concordance among the two pathologists for Gleason pattern sums stood at 67% (0.556, 95% CI 0.40-0.59), respectively. The highest level of concordance was observed at Gleason sum score of 7 in 30 samples (23.4%)

Conclusion: There was a moderate level of concordance in the assessment of the two pathologists by using the Gleason sum score system. It can be greatly improved if better trained uropathologists employ the Gleason scoring system resulting in less variability and more conformity.

Keywords: Gleason scoring, Inter-observer variability, Prostate cancer.

How to Cite This Article: Ahsan F, Khan AA, Asif M, Aslam M, Hamayun S, Din HU. Inter-Observer Variability in Gleason Scoring System for Histological Grading of Adenocarcinoma Prostate. Pak Armed Forces Med J 2022; 72(1): 327-330. Doi: <https://doi.org/10.51253/pafmj.v72i1.7776>

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Prostatic cancer is the second,¹ most common male cancer globally and the leading cause of death accounting for an estimated 9.6 million deaths; with a total ratio of one in six deaths worldwide.² According to most recent library records, prostatic cancer is the fourth most prevalent illness in Pakistani males of age more than 65 years of age.³ With this extensive burden of the disease, standardized diagnostic systems have been incorporated for diagnostic evaluation of this condition.⁴ The Gleason scoring system of prostate cancer is the recommended grading system by WHO which is consistently incorporated in synoptic reporting of prostate specimens.⁵ It is an independent prognostic indicator determining the biological behavior and deciding the management protocols of the patient.⁶

Gleason *et al* in 1966 made investigating structure in accordance with prostate cancer reliant upon the compositional illustration of the tumor.⁷ Gleason histological grading is performed by doing a low-power inspection of the tumor, followed by a more point-by-point architectural analysis.⁸ Gleason scoring system is

the summation of tumor morphology patterns, both the primary and secondary types (based on surface area of involvement). The most predominant pattern of involvement is the primary Gleason pattern and second most common morphology is referred to as secondary pattern.⁹ In TURP cases, where the result is multiple patterns or more than two patterns, the highest grade is to be selected to arrive at a concluding score summation.¹⁰

In 2014, Epstein proposed the concept of grade groups lumping the Gleason scores.¹¹ Later, the International Society of Urological Pathology (ISUP) approved and the modified system of grouping has since been incorporated in 2016 WHO classification.¹² A common problem is the variability subjected to inter-observer and intra-observer factors. Rationale of this study is to evaluate the interobserver variability and thus to increase the reproducibility and effectiveness of Gleason grading system.

METHODOLOGY

This observational double-blind study was conducted at the department of Histopathology, Armed Forces Institute of Pathology, Rawalpindi from May to December 2020. A total of 128 cases of prostate cancer were included in this study keeping confidence inter-

Correspondence: Dr Farah Ahsan, Resident Histopathologic, Armed Forces Institute of Pathology Rawalpindi-Pakistan

Received: 06 Dec 2021; revision received: 17 Feb 2022; accepted: 24 Feb 2022

val at 95% with a margin of error of 5% with population prevalence of carcinoma prostate at 1 (7.1%). Method of sampling was non-probability consecutive sampling. Tissue samples included in this study were either Transurethral resection of prostate (TURP) chips or Trans-rectal ultrasound guided needle-core incisional biopsies (TRUS). All these cases were histologically confirmed cases of prostate cancers. Biopsies that showed fixation artefacts, extensive necrosis or low tumor content were excluded from this study. The study was duly vetted and after final scrutinization, was accepted and allowed to proceed by the review board of the institution held at AFIP vide letter

There was no detection of Gleason pattern 1 or 2 in any of the cores studied by the two pathologists. Agreement was achieved when both the pathologists gave the same Gleason score for the slide under observation. Disagreement was reached when the Gleason score given was different between both pathologists for the same slide under observation. The most reported agreed pattern for both pathologists according to the Gleason scoring was 7 in 30 (23.4%) slide specimens, 8 in 23 (17.9%) slide specimens and 9 in 20 (15.6%) slide specimens in order of descending frequency (Table-I).

Table-I: Gleason scoring pattern frequency and percentage agreement and disagreement between pathologist 1 and pathologist-2.

Pathologist 1 Tumor Scoring in Slides n (%)	Pathologist 2 Tumor Scoring In Slides n (%)					
	Gleason Score	6	7	8	9	10
6		7 (5.4%)	-	-	-	-
7		3 (2.3%)	30 (23.4%)	3 (2.3%)	1 (0.7%)	-
8		-	6 (4.6%)	23 (17.9%)	2 (1.5%)	-
9		-	-	13 (10.1%)	20 (15.6%)	8 (6.2%)
10		-	-	-	7 (5.4%)	5 (3.9%)

number FC-HSP18-11/READ-IRB/20/358 dated 20 April 2020.

The tissue samples were fixed in 10% buffered formal saline. After adequate fixation, they were processed into paraffin embedded blocks. Subsequently sections were prepared, and slides were anonymously numbered. A panel of two histopathologists who had previous expertise in the interpretation of uropathology specimens was selected. The panel of pathologists, unaware to each other for the specimen assessment and to the report furnished earlier, made the observation and subsequent assessment. Sections were independently reported, and results were recorded on pre-approved data proforma.

Data Analysis

Data were statistically described in terms of mean ± SD, frequencies, and percentages when appropriate. To assess concordance between the observers, the kappa (K) co-efficient test was used (CI 95%). The level of agreement was accepted as poor for kappa values between 0-0.19, fair for 0.20–0.39, moderate for 0.40–0.59, good for 0.60–0.79, and very good for values ≥0.80. All statistical calculations were completed using SPSS-26

RESULTS

A total of 128 core slides were assessed by the independent observers. Pathologist 1 and pathologist 2 spotted a tumor in all 128 (100%) with $p < 0.039$.

The rates of concordance or agreement among the observers for Gleason pattern sums were seen in 85 (67%) (K=0.556) of the total 128 slide specimens (Table-II). The greatest agreement for tumor score was seen at Gleason score 7.

Table-II: Level of agreement of gleason sums between both pathologists.

Gleason Sum	Agreement in Total Slides for Gleason Sum n (%)
Gleason Sum 6	7 (5.4%)
Gleason Sum 7	30 (23.4%)
Gleason Sum 8	23 (17.9%)
Gleason Sum 9	20 (15.6%)
Gleason Sum 10	5 (3.9%)
Level of Agreement	85 (66.2%) (K=0.556)

DISCUSSION

Histological grading and scoring systems are subjective assessments owing to inter-observer and intra-observer variability.¹³ There has been a great deal of inconsistency with respect to individual grades of the tumors.¹⁴ Historically there has been a predilection to diagnose potentially aggressive tumors and even a small proportion of poorly formed glands were recognized and graded as highest pattern. This has the subsequent implication of patients getting overtreated.¹⁵ Conversely, the Gleason score 3+3=6 was difficult to be distinguished from 3+4=7. The new ISUP grade groups system has suggested a conservative grading approach

for difficult borderline cases to be downgraded to avoid such overtreatment.¹⁶ Hence, adopting the grade group system has significantly reduced the discordance rates.¹²

Our study aimed to assess the inter-observer variability among the two histopathologists in various grade groups of prostate adenocarcinomas. It was observed that among the 128 selected cases, total agreement with high concordance was found in 56 (44%) cases.

Various factors have been reviewed leading to these difference in scoring systems such as lack of specialized training or education, unawareness of recent guidelines, lack of confidence especially when reporting difficult cases.

Literature review has consistently shown that training workshops and years of experience decreases the degree of dissimilarity in Gleason scoring and thus reduces variability between observers. In a study by Mulay *et al*, an agreement (0.36-0.64) was stated but the ratio of this was profoundly increased and improved by a simple online training on the internet, thus highlighting the value of simple one to one trainings to improve agreement in effectively interpreting prostate biopsies.¹⁷ This level of disagreement was observed more in general his topathologists as compared to Uro-pathologists. In a similar study from Iran; Abdollahi *et al*, reported a kappa value of 0.25 before and 0.52 after a web-based further enforcing the effectivity of pre-trainings and the modality of web-based in general because of their ease of access and cost effectivity.¹⁸

Results of our study are comparable to published literature worldwide. Various steps have been introduced to minimize these interobserver variations like dual reporting and specialized training programs and webinars.

LIMITATION OF STUDY

Limitations of our current study are the lack of specialized Uro-pathologists, as the interobserver variability is expected to be more among general histopathologists than uro-pathologists and inability to evaluate the effects after subsequent educational training of the his topathologists.

CONCLUSION

The agreement between the two pathologists on the Gleason sum was moderate. It can be greatly reduced if better trained uro-pathologists employ the Gleason scoring system resulting in less variability and more conformity.

Therefore, considering the data from literature review and findings of this study it is suggested that in order to

reduce the inter observer variability among the general histopathologists, specialized training programs should be incorporated to maximize the diagnostic utility of Gleason scoring system and thus its implications on patient management.

Conflict of Interest: None.

Authors' Contribution

FA: Primary author, collected data, AAK: Initial concept, collected data, MA: Critically reviewed article, MA: Helped in statistical analysis, SH: Helped in statistical analysis, HUD: Reviewed the final draft for submission.

REFERENCES

1. Rawla P. Epidemiology of prostate cancer. *World J Oncol* 2019; 10(2): 63.
2. Culp MB, Soerjomataram I, Efstathiou JA, Bray F, Jemal A. Recent global patterns in prostate cancer incidence and mortality rates. *European Urol* 2020; 77(1): 38-52.
3. Idrees R, Fatima S, Abdul-Ghafar J, Raheem A, Ahmad Z. Cancer prevalence in Pakistan: meta-analysis of various published studies to determine variation in cancer figures resulting from marked population heterogeneity in different parts of the country. *World J Sur Oncol* 2018; 16(1): 1-11.
4. Boesen L, Nørgaard N, Løgager V, Balslev I, Bisbjerg R, Thestrup K-C, et al. Assessment of the diagnostic accuracy of biparametric magnetic resonance imaging for prostate cancer in biopsy-naive men: the Biparametric MRI for Detection of Prostate Cancer (BIDOC) study. *JAMA Net Open* 2018; 1(2): e180219.
5. Gupta S, Dubey I, Agarwal V, Agarwal S. New perspectives in modified Gleason's grading for prostatic cancer and its comparison with original Gleason's. 2019.
6. Nagpal K, Foote D, Liu Y, Chen PHC. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *NPJ Digital Med* 2019; 2(1): 1-10.
7. Niu Y, Liu CC, Zhang B-L, Song ZG, Chen H, Liu PP, et al. Clinically applicable Gleason grading (GD) system for prostate cancer based on deep learning. *Chinese Med J* 2021; 134(7): 859.
8. Karimi D, Nir G, Fazli L, Black PC, Goldenberg L, Salcudean SE. Deep Learning-Based Gleason grading of prostate cancer from histopathology Images Role of multiscale decision aggregation and data augmentation. *IEEE J Biomed and Health Inform* 2019; 24(5): 1413-26.
9. Faruqi A, Rous B. Standards and datasets for reporting cancers Dataset for histopathological reporting of vulvar carcinomas January 2018. *Royal College Pathol Lond* 2018; 4(G070): 1-37.
10. Kongsanon S, Lerdpraiwan W. The study of incidence of prostate cancer in transurethral resection of the prostate (TURP) specimens. *The Insight Urol* 2019; 40(1): 1-8.
11. Kryvenko ON, Williamson SR, Schwartz LE, Epstein JI. Gleason score 5 + 3 = 8 (grade group 4) prostate cancer-a rare occurrence with contemporary grading. *Human Pathol* 2020; 97(1): 40-51.
12. De Nunzio C, Pastore AL, Lombardo R, Simone G, Leonardo C, Mastroianni R, et al. The new Epstein gleason score classification significantly reduces upgrading in prostate cancer patients. *European J Surg Oncol* 2018; 44(6): 835-839.
13. Nir G, Hor S, Karimi D, Fazli L. Automatic grading of prostate cancer in digitized histopathology images: Learning from multiple experts. *Med Image Analysis* 2018; 50(2): 167-80.
14. Timofte A, Giucă S-E, Manole Mb, Girlescu N, Profire B, Prutianu I, et al. Consistencies and inconsistencies of gleason score and grade group system in prostate biopsy and radical prostatectomy. *Medical-Surgical J* 2021; 125(2): 280-290.

Gleason Scoring System

15. Hanna B, Ranasinghe W, Lawrentschuk N. Risk stratification and avoiding overtreatment in localized prostate cancer. *Current Opinion Urol* 2019; 29(6): 612-619.
 16. Ferro M, Lucarelli G, de Cobelli O, Vartolomei MD, Damiano R, Cantiello F, et al. Circulating preoperative testosterone level predicts unfavourable disease at radical prostatectomy in men with International Society of Urological Pathology Grade Group 1 prostate cancer diagnosed with systematic biopsies. *World J Urol* 2021; 39(6): 1861-1867.
 17. Mulay K, Swain M, Jaiman S, Gowrishankar S. Gleason scoring of prostatic carcinoma: impact of a web-based tutorial on inter- and intra-observer variability. *Ind J Pathol Microbiol* 2008; 51(1): 22.
 18. Abdollahi A, Sheikhabaei S, Meysamie A, Bakhshandeh M, Hosseinzadeh H. Inter-observer reproducibility before and after web-based education in the Gleason grading of the prostate adenocarcinoma among the Iranian pathologists. *Acta Med Iran* 2014; 52(5): 370-374.
-