

# A grading dilemma; Gleason scoring system: Are we sufficiently compatible? A multi center study

Yelda Dere, Özgür İlhan Çelik, Serkan Yasar Çelik, Sümeyye Ekmekçi<sup>1</sup>, Gözde Evcim<sup>2</sup>, Fatma Pehlivan<sup>3</sup>, Anıl Ağalar<sup>4</sup>, Hasan Deliktaş<sup>5</sup>, Nil Çulhacı<sup>6</sup>

Department of Pathology, Faculty of Medicine, Muğla Sıtkı Kocman University, <sup>1</sup>Department of Pathology, Tepecik Training and Research Hospital, <sup>2</sup>Department of Pathology, Çiğli Region Education Hospital, <sup>3</sup>Department of Pathology, Tinaztepe Special Hospital, <sup>4</sup>Department of Pathology, Faculty of Medicine, 9 Eylül University, Izmir, <sup>5</sup>Department of Urology, Faculty of Medicine, Muğla Sıtkı Kocman University, Muğla, <sup>6</sup>Department of Pathology, Faculty of Medicine, Adnan Menderes University, Aydın, Turkey

## Address for correspondence:

Dr. Yelda Dere, Department of Pathology, Muğla Sıtkı Kocman University, Faculty of Medicine, Muğla, Turkey.  
E-mail: yeldamorgul@gmail.com

## ABSTRACT


**Objective:** Gleason scoring is the grading system which strongly predicts the prognosis of prostate cancer. However, even being one of the most commonly used systems, the presence of different interobserver agreement rates push the uropathologists update the definitions of the Gleason patterns. In this study, we aimed to determine the interobserver agreement variability among 7 general pathologists, and one expert uropathologist from 6 different centers. **Methods:** A set of 50 Hematoxylin & Eosin stained slides from 41 patients diagnosed as prostate cancer were revised by 8 different pathologists. The pathologists were also grouped according to having their residency at the same institute or working at the same center. All pathologists' and the subgroups' Gleason scores were then compared for interobserver variability by Fleiss' and Cohen's kappa tests using R v3.2.4. **Results:** There were about 8 pathologists from 6 different centers revised all the slides. One of them was an expert uropathologist with experience of 18 years. Among 7 general pathologists 4 had surgical pathology experience for over 5 years whilst 3 had under 5 years. The Fleiss' kappa was found as 0.54 for primary Gleason pattern, and 0.44 for total Gleason score (moderate agreement). The Fleiss' kappa was 0.45 for grade grouping system. **Conclusion:** Assigning a Gleason score for a patient can be problematic because of different interobserver agreement rates among pathologists even though the patterns were accepted as well-defined.

**KEY WORDS:** Gleason score, interobserver variability, prostate cancer

## INTRODUCTION

Prostate cancer is the most common type of cancer in men in addition to being the second most common cause of death.<sup>[1]</sup> Gleason scoring is the widely recommended, and used system for grading prostate cancer.<sup>[2]</sup>

Gleason grading system is mainly being taught to all pathology residents by the first year of their education. And during their education all the residents came across many prostatic needle biopsies diagnosed as adenocarcinoma with a given Gleason score. In addition to that, after residency general pathologists also reviews many prostatic needle biopsies, diagnoses prostatic adenocarcinomas and gives Gleason scores to these tumors. However, the exact reason for interobserver variability is unclear and was thought to be associated with many different factors such as lack of education or lack of experience.<sup>[3,4]</sup>

<b>Access this article online</b>
<b>Website:</b> www.ijpmonline.org
<b>DOI:</b> 10.4103/IJPM.IJPM_288_18
<b>Quick Response Code:</b>


The most widely used diagnostic approach for patients with high prostate specific antigen (PSA) levels is the prostatic needle biopsies in our country as well as all over the world. The Gleason scores given to prostate cancer biopsies help urologists draw an optimal treatment strategy for their patients. Thus, as being one of the most important predictive factors of prostate cancer, Gleason score itself sometimes is the only factor guiding the treatment decisions. In addition to that the reproducibility of Gleason score becomes

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

**For reprints contact:** reprints@medknow.com

**How to cite this article:** Dere Y, Çelik Ö, Çelik SY, Ekmekçi S, Evcim G, Pehlivan F, et al. A grading dilemma; Gleason scoring system: Are we sufficiently compatible? A multi center study. Indian J Pathol Microbiol 2020;63(Special):S25-9

even more important when the patients decide to be treated in different centers. The reproducibility of Gleason score among pathologists is widely variable.

Depending on these different interobserver agreement rates, we decided to assess interobserver reproducibility rates of Gleason grading of prostate cancer among 8 pathologists from 6 different centers, and put forth our approach to this problem.

## METHODS

A set of 50 Hematoxylin & Eosin stained glass slides diagnosed as prostatic adenocarcinoma was prepared for scoring. The cases were selected randomly from hospital records. All of the slides were prostatic needle biopsies done by the same urologist. The only exclusion criteria was the lack of slides of the patients due to patient's choice of being treated in a different center. In preparing the set all the cases were re-numbered from 1 to 50 to ensure the most simple way for scoring and statistical analysis. The same slides were revised by all the pathologists in order to prevent from diagnostic problems due to serial sectioning.

In addition a simple questionnaire targeting to assess the histopathological approach and personal experience was also prepared and the set and the questionnaire was sent to all the 8 participant pathologists. The main target of the questionnaire was to assess the pathologists' perspective for scoring such as tertiary pattern usage in needle biopsies, immunohistochemical antibodies used for scoring and the main additional parameters given in the pathology reports.

The set of glass slides and the questionnaire was then sent to the all participant pathologists one by one and asked for return in one month time. The scores were then exported to an excel file for statistical analysis.

The pathologists were also subgrouped according to being worked at the same center or being residents at the same center because these can be the criteria affecting agreement rates.

The total number of slides with exact agreement which is defined as the slides to which all the participants gave the same primary, secondary, and total Gleason scores is also noted. The patterns were given according to the 2014 ISUP revision.<sup>[5]</sup>

Kappa value for interobserver agreement was calculated for primary Gleason score, secondary Gleason score, total Gleason score in addition to K values of each subgroup and grade grouping. Kappa value was calculated by using the statistical programme of R v3.2.4 (x64) and the strength of K was accepted as follows; Kappa 0.00-0.20 very low agreement, 0.21-0.40 low agreement, 0.41-0.60 moderate, 0.61-0.80 high agreement, 0.81-1 perfect agreement.<sup>[6]</sup>

This study was supported by a Project from the Scientific Research Projects Management Unit of Muğla Sıtkı Koçman

University (Grant number: 15/086), approved by the ethics committee of the same center (19/2015) and a written informed consent was taken from all the patients included in the study.

## RESULTS

Among the participant pathologists; one is an expert uropathologist with 18 years of experience working at university hospital. One of the general pathologists was also interested in uropathology for two years at her center. Others are general pathologists with a range of experience from 1 to 8 years. Out of them, 3 of them are working at the same university hospital, 2 are working at different training and research hospitals, 1 is working in a state hospital, and the last one is working at a private hospital. Out of it 3 of the pathologists had surgical pathology experience of less than 5 years, whilst 5 of them had more than 5 years. About 3 pathologists had their residency at the same university hospital, another 3 at the same training and research hospital, and 2 at another training and research hospital.

According to the questionnaire; about 7 pathologists (87.5%) have attended a course focusing on this area, and 3 of them (37.5%) has taken a course at the last one year. But one of these 7 pathologists (12.5%) answered as still need to have a course. About 4 pathologists (50%) answered as experience is the most important factor affecting agreement whilst 1 pathologist (12.5%) reported as fresh knowledge is the most important one and other 3 pathologists (37.5%) reported both.

The statistical analysis showed Fleiss' kappa for primary Gleason pattern as 0.54 (moderate agreement), for secondary Gleason pattern as 0.34 (low agreement) and for total Gleason score as 0.44 (moderate agreement). When compared with grade grouping as proposed by Epstein *et al.*<sup>[5]</sup> the Fleiss' kappa was calculated as 0.45 (moderate agreement) between all the pathologists.

Among 50 slides; 24 slides were "consensus" cases in which 70% of the pathologists gave the same score. The three most common Gleason scores of these cases were "3 + 3 = 6" (10 cases) [Figure 1], 3 + 4 = 7 (9 cases) [Figure 2], 4 + 4 = 8 (3 cases).

After further subgrouping; kappa values among pathologists working at the same center was found as 0.59 for primary Gleason pattern and 0.48 for total Gleason score. Working at the same center was found as a factor increasing level of agreement.

Having residency at the same center was found as one of the most important factors for the level of agreement but the most important factor was found as having residency and working at the same center [Table 1].

## DISCUSSION

Interobserver variability is slightly more common in scoring systems in general pathology practice. Gleason grading is also one

of these scoring systems. However, it almost always cause many problems, and one of the most frequent one is confused patients and clinicians. The confusion of clinicians generally occurs when patients want to get a second opinion for the diagnosis or the scores. The importance of this variability lies beneath the difference in the treatment options or prognostic prediction suggested by the clinicians.

Thus, reducing interobserver variability should be a target for every uropathology association worldwide.<sup>[3,4,7,8]</sup> The difference of interobserver variability may depend on many factors such as lack of experience or knowledge, lack of the chance of getting a second opinion, being unaware of new revisions of grading systems, untracking of the current literature, unattending to courses because of money or time problems. The most dangerous reason is self confidency generally seen in new residents especially in problematic cases. In addition to that undergrading of prostatic carcinoma is reported as one of the most common problem for Gleason grading.<sup>[9,10]</sup>

In a study from Iran; Abdollahi *et al.* reported a kappa value of 0.25 before and 0.52 after a web-based education where web-based

educations were found attractive because of the low costs.<sup>[11]</sup> Web-based educations are relatively uncommon in our country however, many pathologists tend to attend interactive courses and educational seminars arranged by different pathology societies. Also Griffiths *et al.* reported low kappa value (0.33) before and moderate agreement ( $k = 0.41$ ) after a teaching session among United Kingdom (UK) pathologists.<sup>[12]</sup> Allsbrook *et al.* worked on “consensus” cases and found moderate agreement among 41 general pathologists ( $k = 0.43$ ), and pointed out that one of the most important factors affecting agreement is learning Gleason score at a course.<sup>[3]</sup> These studies may reflect the strength of education on pathologists’ diagnostic approach.

Working together at the same center for a long time can also be a factor for better agreement results since pathologist is a human-being and a second opinion supporting him or coming up with a new idea is almost always needed and welcomed. But, the rotten apple injures its neighbours. If you work at the same center with a colleague over years you may realize that your choices generally start to be similar. Because of this, professors or expert pathologists should be accessible and modest that every general or uropathologist can get a second opinion from them without

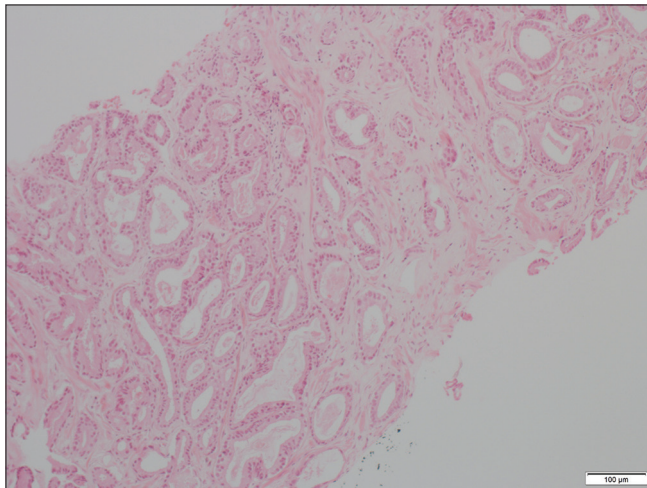


Figure 1: A histopathological appearance of an example of consensus cases for Gleason score 3 + 3 = 6

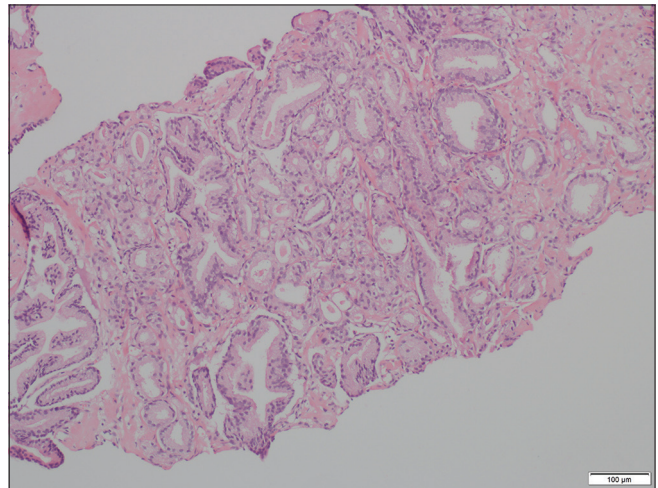


Figure 2: A histopathological appearance of an example of consensus cases for Gleason score 3 + 4 = 7

Table 1: Kappa values for primary Gleason pattern and total Gleason score for every

	Kappa value for primary Gleason pattern	Level of agreement (primary Gleason pattern)	Kappa value for total Gleason score	Level of agreement (total Gleason score)
Overall agreement	0.54	moderate	0.44	moderate
Residency at the same center	0.74	high	0.56	Moderate
	0.95	perfect	0.95	perfect
Working at the same center	0.41	moderate	0.39	low
Residency and working at the same center	0.59	moderate	0.48	moderate
Attending a course	0.95	Perfect	0.95	Perfect
Attending a course at the last one year	0.51	moderate	0.43	moderate
Experience <5 years	0.51	moderate	0.37	Low
Experience >5 years	0.71	High	0.45	moderate
>1000 prostate biopsies for a year observed by the same pathologist	0.48	moderate	0.46	moderate
	0.63	high	0.43	moderate

**Table 2: Other studies focusing interobserver agreement of Gleason score among needle biopsies**

Reference	Pathologist (general/uropathologist)	Number of Pathologist (n)	Number of cases (n)	Percentage of agreement/kappa value
Lessels <i>et al.</i> <sup>[16]</sup>	General pathologists	12	46	0.45
Allsbrook <i>et al.</i> <sup>[3]</sup>	General pathologists	41	38	0.43
Allsbrook <i>et al.</i> <sup>[4]</sup>	Uropathologists	10	46	0.56-0.70
Rodriguez <i>et al.</i> <sup>[7]</sup>	Uropathologists	4	50	0.56
Melia <i>et al.</i> <sup>[17]</sup>	Uropathologists	9	81	0.54
Oyama <i>et al.</i> <sup>[18]</sup>	Mix	14	37	0.49 ve 0.68
Egevad <i>et al.</i> <sup>[8]</sup>	Uropathologist and members of the ENUP*	15 and 337	15 cases and 85 microscopic images	62.4% among the two groups
Veloso <i>et al.</i> <sup>[19]</sup>	General pathologists	3	110	0.35
Özdamar <i>et al.</i> <sup>[15]</sup>	Unknown	2	96	70.8%
Bori <i>et al.</i> <sup>[20]</sup>	General pathologists	23	37	0.49
Current study	Mix	8	50	0.44

\*P=0.009; ENUP: European Network of Uropathologists

being offended. Luckily, getting a second opinion or consultation is readily acceptable in our country.

The new Gleason grading system has already started being used worldwide and also in our country.<sup>[5]</sup> When Fleiss' kappa was calculated according to this grading group system, the agreement status was found as moderate (K = 0.45). This result was slightly better than other studies in the literature.<sup>[13]</sup>

Muezzinoglu *et al.* reported the evaluation of prostatic biopsy samples, the use of tertiary pattern may be different in different centers.<sup>[14]</sup> The same study also showed that reporting of Gleason score is even different among uropathology working group members of Turkey such as 46.4% of them report individual Gleason score for each biopsy core. In another study from Turkey, Ozdamar *et al.* reported 70.8% of interobserver variability whereas different authors reported different results ranging from 0.43-0.70<sup>[15]</sup> [Table 2].

Rodrigues-Urredo *et al.* studied all these variabilities according to microscopic and digital assessment of the biopsies and reported similar interobserver agreement results by the two methods.<sup>[7]</sup> This can increase the use of telepathology and digital consultation which allows you to access experts easily. However in standards of our country, digital slide scanning can be used in only senior pathology centers. This inability can create opportunity when experts use this method to detect the approach of the general pathologist and educate them via web-based seminars.

## CONCLUSION

The interobserver variability of Gleason scoring is still a common problem among general and uropathologists as well as the urologists. This problem may be overcome by web-based educations, free access of articles specialized upon this subject worldwide, disseminate new revisions free by internet or journals and continue to arrange more on-site meetings among pathologists. Therefore, more and more courses targeting strictly newly announced scoring revisions or just scoring biopsies in which multiple attendants have chance of observing the same

slide sections at the same time are needed. For the management of prostate cancer, being aware of this problem may enlighten the way of treatment in cases having different scores from different pathologists.

## Financial support and sponsorship

This study was supported by a Project from the Scientific Research Projects Management Unit of Muğla Sıtkı Koçman University (Grant number: 15/086).

## Conflicts of interest

There are no conflicts of interest.

## REFERENCES

1. Billis A, Guimaraes MS, Freitas LL, Meirelles L, Magna LA, Ferreira U. The impact of the 2005 international society of urological pathology consensus conference on standard Gleason grading of prostatic carcinoma in needle biopsies. *J Urol* 2008;180:548-52.
2. The Royal College of Pathologists. Standards and Minimum datasets for Reporting Common Cancers. Minimum Dataset for Prostate Cancer Histopathology Reports. London: The Royal College of Pathologists; 2000.
3. Allsbrook WC, Mangold KA, Johnson MH, Lane RB, Lane CG, Epstein JI. Interobserver reproducibility of Gleason grading of prostatic carcinoma: General pathologists. *Hum Pathol* 2001;32:81-8.
4. Allsbrook WC Jr, Mangold KA, Johnson MH, Lane RB, Lane CG, Amin MB, *et al.* Interobserver reproducibility of Gleason grading of prostatic carcinoma: Urologic pathologists. *Hum Pathol* 2001;32:74-80.
5. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA; Grading Committee. 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;40:244-52.
6. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971;76:378-82.
7. Rodriguez-Urrego PA, Cronin AM, Al-Ahmadie HA, Gopalan A, Tickoo SK, Reuter VE, *et al.* Interobserver and intraobserver reproducibility in digital and routine microscopic assessment of prostate needle biopsies. *Hum Pathol* 2011;42:68-74.
8. Egevad L, Ahmad AS, Algaba F, Berney DM, Boccon-Gibod L, Compérat E, *et al.* Standardization of Gleason grading among 337 European pathologists. *Histopathology* 2013;62:247-56.
9. Rubin MA, Dunn R, Kambham PA, Misick CP, O'Toole KM. Should a

- Gleason score be assigned to a minute focus of carcinoma on prostate biopsy? *Am J Surg Pathol* 2000;24:1634-40.
10. Steinberg DM, Sauvageot J, Piantadosi S, Epstein JI. Correlation of prostate needle biopsy and radical prostatectomy Gleason grade in academic and community setting. *Am J Surg Pathol* 1997;21:566-76.
  11. Abdollahi A, Sheikhabaehi 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;52:370-4.
  12. Griffiths DF, Melia J, McWilliam LJ, Ball RY, Grigor K, Harnden P, *et al.* A study of Gleason score interpretation in different groups of UK pathologists; techniques for improving reproducibility. *Histopathology* 2006;48:655-62.
  13. Ozkan TA, Eruyar AT, Cebeci OO, Memik O, Ozcan L, Kuskonmaz I. Interobserver variability in Gleason histological grading of prostate cancer. *Scand J Urol* 2016;50:420-4.
  14. Muezzinoglu B, Yorukoglu K. Current practice in handling and reporting prostate needle biopsies: Results of a Turkish survey. *Pathol Res Practice* 2015;211:374-80.
  15. Ozdamar SO, Sarikaya S, Yildiz L, Atilla MK, Kandemir B, Yildiz S. Intraobserver and interobserver reproducibility of WHO and Gleason histologic grading systems in prostatic adenocarcinomas. *Int Urol Nephrol* 1996;28:73-7.
  16. Lessels AM, Burnett RA, Howatson SR, Lang S, Lee FD, McLaren KM, *et al.* Observer variability in the histopathological reporting of needle biopsy specimens of the prostate. *Hum Pathol* 1997;28:646-9.
  17. Melia J, Moseley R, Ball RY, Griffiths DFR, Grigor K, Harnden P, *et al.* A UK-based investigation of inter- and intra-observer reproducibility of Gleason grading of prostatic biopsies. *Histopathology* 2006;48:644-54.
  18. Oyama T, Allsbrook WC Jr, Kurokawa K, Matsuda H, Segawa A, Sano T, *et al.* A comparison of interobserver reproducibility of Gleason grading of prostatic carcinoma in Japan and the United States. *Arch Pathol Lab Med* 2005;129:1004-10.
  19. Veloso SG, Lima MF, Salles PG, Berenstein CK, Scalon JD, Bambirra EA. Interobserver agreement of Gleason score and modified Gleason score in needle biopsy and in surgical specimen of prostate cancer. *Int Braz J Urol* 2007;33:639-51.
  20. Bori R, Salamon F, Móczár C, Cserni G. Interobserver reproducibility of Gleason grading in prostate biopsy samples. *Orv Hetil* 2013;154:1219-25.

Copyright of Indian Journal of Pathology & Microbiology is the property of Wolters Kluwer India Pvt Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.