acta medica

ORIGINAL ARTICLE

A tale of two uropathologists: concordance of Gleason Grade Groups in prostatic adenocarcinoma over needle biopsies and radical prostatectomies

Güneş Güner[1]
ORCID: 0000-0002-7338-1524

Kemal Kösemehmetoğlu[1]
ORCID: 0000-0002-7747-0460

[1] Hacettepe University, Faculty of Medicine, Department of Pathology, Ankara, Türkiye

Corresponding Author: Güneş Güner
E-mail: gunesguner@gmail.com

Received: 6 August 2024, Accepted: 19 December 2024, Published online: 30 December 2024

ABSTRACT

Objective: The standard 12-core transrectal prostate needle biopsies don't reflect the tumor entirely. Approximately 35-36% of needle biopsy diagnoses of Gleason grade group (GG) 1 are upgraded upon radical prostatectomy (RP). Pathologists are not in perfect concordance in Gleason scoring. Two uropathologists in a university hospital aimed to determine how concordant needle biopsy GGs were with RP GGs. Moreover, they also assessed how frequently they up-/down-graded the needle biopsy GGs each other gave when they graded RPs.

Material and Methods: In-house prostate needle biopsies and RPs from 31/01/2020 to 10/09/2022 were retrieved from the hospital database. Patients who had both a needle biopsy and an RP were included. Whether each case was down-/up-graded upon RP, GGs and the pathologist that reported the cases were tabulated.

Results: One hundred and thirty cases were assessed. Needle biopsy and RP GGs were identical in 63,1% (n=82). Pathologist1 (P1) assessed both the needle biopsy and RPs of 41 patients, 8 of which they downgraded and 8 upgraded (19,5%). Pathologist2 (P2) assessed both the needle biopsy and RP samples of 23 patients; downgrading 13% (n=3) and upgrading 17,4% (n=4) of cases. Where the needle biopsy was reported by P2 and RP was reported by P1 (n=48), 10 (20,8%) were downgraded and 8 (16,7%) were upgraded. The reverse scenario was noted in 18 patients; 2 (11,1%) of which were downgraded, 5 (27,8%) were upgraded. While P2 showed a tendency to upgrade more frequently, this was not statistically significant (p=0,2774, Pearson chi-square).

Discussion: The two uropathologists' up- and down-grading rates seemed concordant. Routine practice doesn't allow time for one pathologist to re-score the other's cases, nor is every case consulted. Looking back at pathologist-specific tendencies to up/downgrade one's own or a colleague's scores may help direct ourselves and others to curb tendencies to over/undergrade.

Keywords: prostatic adenocarcinoma, gleason grade group, concordance, upgrading, downgrading, needle biopsy, radical prostatectomy

INTRODUCTION

Accurate grading of prostatic adenocarcinoma is of paramount importance for treatment decisions and prognostication. Gleason grading is decisive for treatment in about 40% of all prostate cancer patients [1]. Laboratory grading practices may impact treatment decisions [2].

Needle biopsy Gleason scores tend to differ from that of the radical prostatectomy Gleason scores in a considerable portion of cases. The routine 12-core biopsy scheme showed a biopsy-radical Gleason score concordance rate as high as 85% with a possibility of upgrading at 17% of cases and downgrading of 14% [3]. UK data indicate rates of 25.5% and 15.6% for upgrade and downgrading, respectively; their concordance rate is 58.9% [4]. In another study, about one third of Gleason score 6 cases were upgraded at radical prostatectomy; about one third of Gleason score 8 cases got upgraded and another third got downgraded [5]. More than a third (38.2%) of cases with Gleason score 6 at biopsy got upgraded at radical prostatectomy [6]. Several factors affect down- and up-grading tendencies such as needle biopsy procedures themselves [7], cancer volume/extent at biopsy [8, 9], PSA levels [5, 8, 9], prostate volume [5, 8], the presence of extraprostatic extension and surgical margin positivity [9], risk group of disease [4, 10], and the presence of a tertiary Gleason pattern [5]. Lower D'Amico risk groups tend to have a higher rate of upgrading [4]; however, the opposite is also claimed [10]. Magnetic resonance imaging (MRI) - guided biopsies are more sensitive and less specific than transrectal US-guided biopsies to detect clinically relevant prostate cancer [7], MRI-targeted biopsies and systematic (12-core) biopsies have a similar downgrading risk while MRI-targeted sampling has a lower risk of upgrading [11]. Properties of the population under study (low vs high clinical risk / mixed), Gleason scoring systems used (modified vs previous), tumor and organ properties (high vs low extent tumor, high vs low prostate volume), sampling types (targeted vs systematic needle sampling, transrectal vs transperineal [12]) all vary among studies; yet overall it is safe to comment that needle biopsy Gleason scores tend to differ from that of the radical prostatectomy Gleason score in about one fourth to one third of all cases. This has prognostic and therapeutic implications; upgrading on radical prostatectomy is associated with increased risk of biochemical recurrence and adverse pathological parameters [10]. Undergraded cases may be undertreated, overgraded cases can be overtreated [4].

Gleason grading scheme is arguably the best described grading system in all surgical pathology; yet interobserver variability remains. In the early 2000s, North American general pathologists showed barely moderate agreement (kappa=0,435) while uropathologists were in moderate to substantial agreement (kappa=0,56-0,70) [13]. British pathologists in 2006 followed with moderate overall interobserver agreement (kappa=0,54); their intra-observer agreement was good (kappa=0,66) [14]. Turkish general pathologists showed barely moderate agreement both before and after the Gleason grade modification [15] (kappa=0,45 [16], kappa=0,43 [17]). Others report similar extents of interobserver agreement (Kappa=0,482 [18], kappa=0,55 [19]) with outliers (kappa= 0,753 [20]). Poorly differentiated/high grade tumors may have better overall agreement (kappa=0,65) than well differentiated/low grade tumors (kappa=0,15) [21]. Low tumor volumes in needle biopsies [22, 23] and distinguishing Gleason score 6 and 7 with low pattern 4 percents tend to be the most challenging aspects [23].

Interobserver agreement studies typically include re-assessment of previously reported samples; this is not easily achievable and can't be repeated frequently in busy pathology practices. Needle biopsy - radical prostatectomy Gleason score variation per patient is also a variable in routine pathology work that needs consideration. The present study aims to combine these two sources of inconsistency/variability in an easily evaluated manner and assess 1) the needle-to-radical Gleason scoring changes of two uropathologists in the same academic institution in Turkey, and 2) attempt to evaluate how concordant these two pathologists are without reassessment of previous samples/ slides and complicated statistical calculations.

METHODS

All in-house prostate needle biopsies and in-house radical prostatectomies between 31st January 2020 and 10th September 2022 of one university hospital were listed using the digital hospital database. Patients who underwent both a needle biopsy and an RP at our institution were included in the study. The highest GG detected in the needle biopsy and the GG of the dominant nodule in RP were noted for each patient. Minor/tertiary Gleason patterns in RPs were not evaluated. Whether each case was down- or upgraded upon radical prostatectomy and the pathologist that signed out the cases were assessed. The percentages and numbers of cases each pathologist up- or downgraded were calculated. Cases where one pathologist signed out the needle biopsy and the other reported the radical prostatectomy were separately assessed to see whether there was a tendency of up or downgrading in any of the two scenarios. Numbers of cases were tabulated whenever possible, and Pearson chi-square test was applied to detect statistically significant associations for 2x2 tables (http://vassarstats.net/, accessed 08/29/23). Fisher's Exact Test was used for larger tables (https://astatsa.com/FisherTest/, accessed 07/31/2024). A p value <0.05 was considered significant.

No patient identifier was included in the data; the pathology reports along with any associated human tissue/blocks remained unchanged and intact. The study was approved by the Health Sciences Research Ethics Committee of the Hacettepe University Hospital (Protocol no. SBA 23/239).

RESULTS

One hundred and thirty patients had both their needle biopsies and prostatectomies reported in our department between 31st January 2020 and 10th September 2022. The distribution of cases between Pathologist 1 (P1) and Pathologist 2 (P2) are given in Table 1. The breakdown of discordant (up- or downgraded) cases are given in Table 2.

Overall concordance between needle biopsy and radical prostatectomy GGs was 63,1%. P2 assessed both the needle biopsy and radical prostatectomy samples of 23 patients; downgraded three cases (13%) and upgraded 4 (17,4%). P1 assessed both the needle biopsy and radical prostatectomy samples of 41 patients, downgraded and upgraded 19,5% each. As is seen in Table 1, P1 up/downgraded their own needle biopsy GGs in 16 cases (39%); P2 up/downgraded their own needle biopsy GGs in 7 cases (30,4%). Of the 59 patients whose needle biopsies were reported by P1, 10 (16,9%) were downgraded and 13 (22%) were upgraded upon radical prostatectomy. Seventy-one needle biopsies were reported by P2; 13 (18,3%) were downgraded and 12 (16,9%) were upgraded upon radical prostatectomy. Overall, the possibility of P1's needle biopsy GG getting changed at radical prostatectomy was 38,98% (23/59) and for P2 the same possibility was 35,21% (25/71).

Where the needle biopsy was reported by P2 and the radical prostatectomy was reported by P1 (n=48), 10 (20,8%) were downgraded and 8 (16,7%)

Table 1. Distribution of cases

Needle biopsy reported by	Radical prostatectomy reported by	Number of cases downgraded (n, %)	Number of cases upgraded (n, %)	Number of cases with no change in GG (n, %)	Total
P1	P1	8 (19,5%)	8 (19,5%)	25 (61%)	41 (100%)
P1	P2	2 (11,1%)	5 (27,8%)	11 (61,1%)	18 (100%)
P2	P1	10 (20,8%)	8 (16,7%)	30 (62,5%)	48 (100%)
P2	P2	3 (13%)	4 (17,4%)	16 (69,6%)	23 (100%)
	Total	23 (17,7%)	25 (19,2%)	82 (63,1%)	130

p= 0.9251 Fisher's Exact Test, 2-sided

Table 2. The tendency of each pathologist to down/upgrade a needle biopsy grade upon radical prostatectomy

Cases down/upgraded by	Downgraded (n, %)	Upgraded (n, %)	Total
P1	18 (78,3%)	16 (64%)	34
P2	5 (21,7%)	9 (36%)	14
Total	23 (100%)	25 (100%)	48

p=0.2774 Pearson chi square

were upgraded. The reverse scenario was noted in 18 patients; 2 (11,1%) of which were downgraded, 5 (27,8%) were upgraded. While P2 showed a tendency to upgrade more frequently, this was not statistically significant (p=0.2774, Pearson chi-square).

DISCUSSION

Gleason grading system is one of the most important prognostic factors for prostatic adenocarcinomas. It is known that about one third or one fourth of cases undergo a Gleason grading change between their needle biopsies and radical prostatectomies; noting a sampling issue in Gleason grading. Another source of variation is interobserver variability, in which most reports indicate moderate concordance between pathologists. These two major sources of variability in prostate carcinoma Gleason grading (interobserver and sampling variabilities) are evaluated and reported separately using labor-intensive studies. The present study aimed to demonstrate the cumulative effect of these two factors in the routine workflow of one uropathology practice of two pathologists without microscopic re-reviewing of past slides or statistical calculations of kappa values.

The total number of cases whose needle biopsy samples and radical prostatectomies were assessed at the same institution (university hospital) in a span of 2 years was 130. About two thirds of these cases have not undergone up- or downgrading (82; 63,1%); which seems to be concordant with relevant Turkish literature [24, 25]. As the up- and downgrading rates are very much alike (25 - 19,2% and 23 - 17,7%, respectively); institution-wise, the present set of pathologists do not seem to be over-upgrading or over-downgrading. Some laboratories may tend to grade lower or higher than average [2]. The up and downgrading rates calculated in the present study are similar to those reported by Bjurlin et al (17% up and 14% downgrading [3]); and lower than that of most of the pertinent literature [4-6].

P2 tended to have lower rates of downgrading than P1 (5 vs 18 cases, Table 2), yet this tendency did not reach statistical significance. This may mean that individual pathologist variability may not substantially impact overall grading consistency. Evaluating a larger patient group and a longer period might help support or refute this tendency.

The present study has several drawbacks beyond its retrospective design. Relatively low numbers of patients in the present study precludes detailed analyses of other parameters claimed to affect rates of upgrading such as PSA levels and prostate volume [26] along with Gleason grade groups themselves. Prognostic data, such as biochemical recurrence, metastasis, dead of disease, etc. was not assessed. Discordant (up/downgraded) needle biopsy - radical prostatectomy pairs were not re-assessed by both pathologists to turn this endeavor to an educational opportunity for minimizing interobserver variability. However, the tables in the results section can be adapted to any lab and reproduced by almost any healthcare worker (nurse, medical secretary, medical student, etc), creating an ongoing and temporally evolving database of up and downgrading tendencies that can eventually be used for discussions and create teaching opportunities. Future studies that have larger sample sizes and include prognostic outcomes may help combine the two factors of inconsistency in Gleason grading (interobserver variability and variations inherent to sampling) to validate or refute these findings.

Author contribution

Study conception and design: GG, KK; data collection: GG; analysis and interpretation of results: GG, KK; draft manuscript preparation: GG, KK. All authors reviewed the results and approved the final version of the manuscript.

Ethical approval

The study was approved by the Health Sciences Research Ethics Committee of the Hacettepe University Hospital (Protocol no. SBA 23/239).

Funding

The authors declare that the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

REFERENCES

[1] NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) Prostate Cancer Version 3.2023 National Comprehensive Cancer Network®; 2023 [updated 08/07/2023. Version 3.2023. Available from: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf.

[2] Flach RN, van Dooijeweert C, Aben KKH, et al. Interlaboratory Gleason grading variation affects treatment: a Dutch historic cohort study in 30 509 patients with prostate cancer. J Clin Pathol. 2023; 76(10):690-697.

[3] Bjurlin MA, Carter HB, Schellhammer P, et al. Optimization of initial prostate biopsy in clinical practice: sampling, labeling and specimen processing. J Urol. 2013;189(6):2039-46.

[4] Bullock N, Simpkin A, Fowler S, et al. Pathological upgrading in prostate cancer treated with surgery in the United Kingdom: trends and risk factors from the British Association of Urological Surgeons Radical Prostatectomy Registry. BMC Urol. 2019;19(1):94.

[5] Epstein JI, Feng Z, Trock BJ, et al. Upgrading and downgrading of prostate cancer from biopsy to radical prostatectomy: incidence and predictive factors using the modified Gleason grading system and factoring in tertiary grades. Eur Urol. 2012;61(5):1019-24.

[6] Milonas D, Grybas A, Auskalnis S, et al. Factors predicting Gleason score 6 upgrading after radical prostatectomy. Cent European J Urol. 2011;64(4):205-8.

[7] Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet. 2017;389(10071):815-22.

[8] Dong F, Jones JS, Stephenson AJ, et al. Prostate cancer volume at biopsy predicts clinically significant upgrading. J Urol. 2008;179(3):896-900.

[9] Hong SK, Han BK, Lee ST, et al. Prediction of Gleason score upgrading in low-risk prostate cancers diagnosed via multi (> or = 12)-core prostate biopsy. World J Urol. 2009;27(2):271-6.

[10] Freedland SJ, Kane CJ, Amling CL, et al. Upgrading and downgrading of prostate needle biopsy specimens: risk factors and clinical implications. Urology. 2007;69(3):495-9.

[11] Goel S, Shoag JE, Gross MD, et al. Concordance Between Biopsy and Radical Prostatectomy Pathology in the Era of Targeted Biopsy: A Systematic Review and Meta-analysis. Eur Urol Oncol. 2020;3(1):10-20.

[12] Marra G, Eldred-Evans D, Challacombe B, et al. Pathological Concordance between Prostate Biopsies and Radical Prostatectomy Using Transperineal Sector Mapping Biopsies: Validation and Comparison with Transrectal Biopsies. Urol Int. 2017;99(2):168-76.

[13] Allsbrook WC, Jr., Mangold KA, Johnson MH, et al. Interobserver reproducibility of Gleason grading of prostatic carcinoma: general pathologist. Hum Pathol. 2001;32(1):81-8.

[14] Melia J, Moseley R, Ball RY, et al. A UK-based investigation of inter- and intra-observer reproducibility of Gleason grading of prostatic biopsies. Histopathology. 2006;48(6):644-54.

[15] 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(3):428-35.

[16] Dere Y, Celik OI, Celik SY, et al. A grading dilemma; Gleason scoring system: Are we sufficiently compatible? A multi center study. Indian J Pathol Microbiol. 2020;63(Supplement):S25-S9.

[17] Ozkan TA, Eruyar AT, Cebeci OO, et al. Interobserver variability in Gleason histological grading of prostate cancer. Scand J Urol. 2016;50(6):420-4.

[18] Qureshi A, Lakhtakia R, Al Bahri M, et al. Gleason's Grading of Prostatic Adenocarcinoma: Inter-Observer Variation Among Seven Pathologists at a Tertiary Care Center in Oman. Asian Pac J Cancer Prev. 2016;17(11):4867-8.

[19] Singh RV, Agashe SR, Gosavi AV, et al. Interobserver reproducibility of Gleason grading of prostatic adenocarcinoma among general pathologists. Indian J Cancer. 2011;48(4):488-95.

[20] Al Nemer AM, Elsharkawy T, Elshawarby M, et al. The updated grading system of prostate carcinoma: an inter-observer agreement study among general pathologists in an academic practice. APMIS. 2017;125(11):957-61.

[21] Bori R, Salamon F, Moczar C, et al. Interobserver reproducibility of Gleason grading in prostate biopsy samples. Orv Hetil. 2013;154(31):1219-25.

[22] Sadimin ET, Khani F, Diolombi M, et al. Interobserver Reproducibility of Percent Gleason Pattern 4 in Prostatic Adenocarcinoma on Prostate Biopsies. Am J Surg Pathol. 2016;40(12):1686-92.

[23] 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.

[24] Akarken İ, Dere Y, Tarhan H, et al. Assessment of gleason score concordance between prostate biopsy and radical prostatectomy. J Cukurova Anesth Surg. 2022;5(2):274-9.

[25] Öztürk EY, Yıkılmaz TN. Gleason score correlation between prostate biopsy and radical prostatectomy specimens. Bulletin of Urooncology. 2018;17(1):1-4.

[26] Tilki D, Schlenker B, John M, et al. Clinical and pathologic predictors of Gleason sum upgrading in patients after radical prostatectomy: results from a single institution series. Urol Oncol. 2011;29(5):508-14.