Caspian Journal of Surgery
DOI: 10.22088/caspjs.2026.2087594.1058
(2026) vol 2: (2)

Original Article

Concordance Between Biopsy and Radical Prostatectomy Specimen Gleason Scores, in Patients Referred to Babol Shahid Beheshti Hospital and Referral Centers During the Years 2012-2024

Hossein Ghorbani Pashakolaei[1], Hamid Shafi[2], Hoda Shirafkan[2], Khadije Hosseinpor[2], Ghodsieh Kamrani[2*]

1. Student Research Committee, Babol University of Medical Sciences, Babol, Iran

2. Clinical Research Development Center, Shahid Beheshti Hospital, Babol University of Medical Sciences, Babol, Iran

Background: Prostate cancer is the most common malignancy of the genitourinary system in men, and the Gleason score (GS) remains the most validated prognostic indicator. Despite its critical role, discrepancies between GS values from biopsy and radical prostatectomy specimens have been reported. This study aimed to assess the concordance rate of GS and identify factors associated with upgrading or downgrading of tumour grade.

Methods: This retrospective observational cohort study reviewed the records of 133 patients diagnosed with prostate adenocarcinoma who underwent needle biopsy followed by radical prostatectomy at Shahid Beheshti Hospital, Babol, between 2012 and 2024. All specimens were re-evaluated by a single pathologist. Statistical analyses were performed using SPSS v27.

Results: The concordance rate of Gleason scores between biopsy and prostatectomy specimens was 60.9%, with a weighted kappa coefficient of 0.656 (95% CI 0.575–0.738), indicating good agreement. Upgrading was observed in 29.3% of cases and downgrading in 9.8%. Among the clinical and pathological parameters assessed, biopsy tumour involvement percentage differed significantly across concordance categories (Kruskal-Wallis, P = 0.006); age, PSA, prostate volume, time interval between biopsy and surgery, and perineural invasion showed no significant associations.

***Conclusion*:** Needle biopsy provides a moderately reliable estimate of the Gleason score, but biopsy tumour involvement percentage differed across concordance categories; this finding should not be interpreted as a confirmed directional association with upgrading. Integrating this histopathological parameter with imaging and emerging molecular and computational tools may help refine risk stratification and treatment planning.

Keywords: Prostatic Neoplasms; Neoplasm Grading; Prostatectomy; Biopsy, Needle; Neoplasm Staging

Received:
February 4, 2026
Revised:
April 26, 2026
Accepted:
May 10, 2026
Published on:
June 30, 2026

Corresponding author:
Ghodsieh Kamrani
Address**:** Clinical Research Development Center, Shahid Beheshti Hospital, Babol University of Medical Sciences, Babol, Iran
E-mail:
gh.kamrani@gmail.com

© The Author(s).
Publisher: Babol University of Medical Sciences
This work is published as an open access article distributed under the terms of the Creative Commons Attribution 4.0 License (http://creativecommons.org/licenses/by-nc/4). Non-commercial uses of the work are permitted, provided the original work is properly cited.

Introduction

Prostate cancer (PCa) remains a significant public health concern globally, constituting one of the most frequently diagnosed malignancies among men (1). Accurate pathological grading is paramount for risk stratification, prognosis prediction, and determining the optimal treatment strategy, which ranges from active surveillance to radical treatments such as radical prostatectomy (RP) or radiotherapy.

The Gleason grading system, initially introduced in 1966, serves as the cornerstone for pathological assessment of prostate adenocarcinoma based on the architecture of glandular patterns (2,3). The current standard, incorporating the 2014 International Society of Urological Pathology (ISUP) modifications, categorizes PCa into five Grade Groups (1–5), which correspond directly to the traditional Gleason Score (GS) summation. A higher GS indicates a more aggressive tumor phenotype with poorer prognosis, necessitating more intensive treatment.

The standard initial diagnostic procedure for localized PCa is transrectal ultrasound (TRUS)-guided or multipara metric MRI (mpMRI)-guided systematic needle biopsy (NB). The pathological findings from this limited tissue sampling form the basis for initial clinical decisions. However, because the prostate gland is relatively large and NB samples only a small fraction of the total prostatic tissue, inherent sampling error frequently results in discrepancies between the tumor grade identified on the NB specimen and the definitive grade ascertained from whole-mount sectioning of the RP specimen (4).

These discrepancies - specifically GS upgrading or downgrading - significantly impact treatment planning. Upgrading often means that a patient initially deemed low-risk may in fact harbor more aggressive disease requiring escalation of therapy, whereas downgrading may lead to unnecessary overtreatment (5). Previous literature has explored the magnitude and predictors of this discordance, often citing factors such as biopsy technique, number of cores sampled, tumor burden (percentage of positive cores), and the time interval between diagnosis and surgery (6).

This study was conducted at Shahid Beheshti Hospital, Babol, drawing on a cohort spanning over a decade (2012–2024). The primary objective was to quantify the concordance rate of the Gleason score between initial NB and definitive RP specimens in this regional Iranian population. We further aimed to analyze clinical and pathological parameters to identify factors associated with GS discordance, with particular attention to upgrading, which carries greater clinical implications for under-treatment.

Materials and Methods

This retrospective observational cohort study reviewed the institutional electronic medical records and archived pathology reports of male patients diagnosed with prostate adenocarcinoma who underwent an initial prostate needle biopsy followed by radical prostatectomy at Shahid Beheshti Hospital, Babol, , between 2012 and 2024. To minimize inter-observer variability, all biopsy and radical prostatectomy pathology specimens were re-evaluated by a single dedicated pathologist for the purposes of this study.

Inclusion criteria were: (1) histologically confirmed prostate adenocarcinoma on needle biopsy; (2) subsequent definitive treatment via robot-assisted or open radical prostatectomy (RP) within the study period; and (3) availability of complete pathological data for both biopsy and RP specimens. Exclusion criteria were patients who received neoadjuvant hormonal or chemotherapeutic treatment prior to RP, as these interventions can alter tumor morphology and grading, and cases with an insufficient number of biopsy cores for reliable grading or incomplete pathological records.

The final cohort comprised 133 patients who met all inclusion criteria. Data collection adhered to the ethical guidelines of the Institutional Review Board of Babol University of Medical Sciences (ethics code IR.MUBABOL.HRI.REC.1403.265).

For each eligible patient, the following variables were extracted from the electronic medical records and pathology databases:

Biopsy (NB) parameters:

age at diagnosis; preoperative serum prostate-specific antigen (PSA, ng/mL); biopsy Gleason Score and Grade Group; tumor involvement percentage; presence of perineural invasion (PNI).

Pathological (RP) parameters

final Gleason Score and Grade Group.

Perioperative parameters:

total prostate volume derived from imaging; time interval between the date of biopsy and the date of RP (days).

Tumor involvement percentage in the biopsy specimen was defined as the percentage of tissue area occupied by adenocarcinoma on microscopic examination relative to the total tissue present; an analogous measure (occasionally supplemented by gross/macroscopic evaluation) is recorded for the radical prostatectomy specimen at final pathological reporting, but it was the biopsy-assessed value that was used as a predictor of Gleason score discordance in the analyses below. Perineural invasion (PNI) was likewise assessed on the biopsy specimen.

Statistical Analysis

All statistical analyses were conducted using SPSS version 27. Continuous variables were summarized as median [interquartile range (IQR)] because of skewed distributions and small subgroup sizes, and categorical variables as frequencies and percentages.

Continuous variables (age, PSA, and prostate volume) were compared across the five categories of Gleason score change - two-grade downgrading, one-grade downgrading, agreement, one-grade upgrading, and two-grade upgrading - using the non-parametric Kruskal-Wallis test due to skewed distributions, unequal subgroup sizes, and small numbers in some categories. Tumor involvement percentage and the time interval between biopsy and surgery, which showed skewed distributions, were compared using the non-parametric Kruskal-Wallis test and the Jonckheere-Terpstra test, respectively.

The overall agreement between biopsy and radical prostatectomy Gleason scores, based on ISUP Grade Groups, was quantified using the weighted Cohen's kappa coefficient, with linear weights reflecting the ordinal nature of the Grade Group categories, and reported with its 95% confidence interval.

The association between perineural invasion and Gleason score concordance outcome was assessed using the chi-square test. A p-value less than 0.05 was considered statistically significant for all tests.

Results

A total of 133 patients who underwent RP following NB between 2012 and 2024 were included in the final analysis. The mean age of the cohort was 65.6 ± 7.7 years (range 50–88). The mean preoperative PSA was 14.7 ± 13.5 ng/mL (range 2.4–125.0). The mean total prostate volume was 55.3 ± 18.9 cc (range 20–110). The average time interval between biopsy and surgery was 84.6 ± 47.8 days (range 12–366). The mean tumor involvement percentage was 42.4% ± 18.6% (range 10–85%). Perineural invasion (PNI) was present in 81 of 133 specimens (61%).

The comparison between biopsy and radical prostatectomy Gleason Grade Group is summarized in Table 1.

Table 1. Concordance between biopsy and radical prostatectomy Gleason Grade Group

Outcome Category	Frequency, n	Percentage, %
Agreement	81	60.9
Upgrading	39	29.3
Downgrading	13	9.8
Total	133	100.0

The overall concordance rate, defined as exact agreement, was 60.9% (81/133). The weighted Cohen's kappa coefficient, using linear weighting and based on

ISUP Grade Groups, was 0.656 (SE=0.042, z=11.278, p<0.001, 95% CI 0.575–0.738), indicating good (substantial) agreement between biopsy and RP grading.

Upgrading occurred in 39 patients (29.3%) and downgrading in 13 patients (9.8%). The most frequent shift was upgrading from biopsy Grade Group 1 (GS 6) to RP Grade Group 2 (GS 3+4=7), observed in 22 patients (16.5% of the cohort), followed by upgrading from Grade Group 2 to Grade Group 3, and downgrading from Grade Group 3 to Grade Group 2.

A detailed cross-tabulation of biopsy against RP Grade Group is presented in Table 2, which allows visualization of the exact pattern of Gleason score change across the full grading spectrum.).

Table 2. Cross-tabulation of biopsy and radical prostatectomy Grade Group (observed counts)

Biopsy Grade Group	RP GG 1	RP GG 2	RP GG 3	RP GG 4	RP GG 5	Total
GG 1	40	22	1	0	0	63
GG 2	1	20	6	1	0	28
GG 3	1	4	9	0	6	20
GG 4	0	0	5	5	3	13
GG 5	0	0	2	0	7	9
Total	**42**	**46**	**23**	**6**	**16**	**133**

Table 3. Comparison of continuous variables across the five Gleason score concordance categories

Concordance Category	n	Age (years) Median [IQR]	PSA (ng/mL) Median [IQR]	Prostate Volume (cc) Median [IQR]	Time Interval (days) Median [IQR]
Two-grade downgrade	3	64.7 [6.0]	13.6 [—]	55.0 [—]	70.0 [—]
One-grade downgrade	10	65.6 [35.0]	16.4 [14.6]	50.0 [28.3]	89.0 [85.3]
Agreement	81	64.8 [28.0]	10.4 [10.2]	55.0 [20.5]	80.0 [35.0]
One-grade upgrade	31	66.8 [38.0]	11.2 [9.8]	60.0 [27.0]	80.0 [41.0]
Two-grade upgrade	8	69.8 [20.0]	9.9 [12.7]	45.0 [31.5]	74.0 [34.0]
Total	133	65.6 [38.0]	11.5 [13.0]	55.0 [38.0]	80.0 [92.0]
p-value		0.349*	0.531*	0.154*	0.616†

*** One-way Kruskal-Wallis test. † Jonckheere-Terpstra test.**

Because of sparse expected cell counts in the 5x5 cross-tabulation, this table is presented descriptively. Agreement between biopsy and radical prostatectomy Grade Groups was primarily assessed using weighted Cohens kappa.

Analysis of Factors Associated with Discordance

We assessed clinical and pathological factors for their association with Gleason score concordance outcome across the five categories of grade change.

Continuous variables. Age, PSA, and prostate volume were compared across the five concordance categories using the Kruskal-Wallis test; none showed

astatistically significant difference (all p>0.05; Table3).

Biopsy tumor involvement percentage differed significantly across Gleason score concordance categories. However, because the Kruskal-Wallis test is an omnibus comparison, this finding indicates group differences rather than a confirmed directional relationship with upgrading. Further studies using multivariable modeling are required to determine whether tumor involvement percentage independently predicts Gleason score discordance.

Perineural invasion

Perineural invasion (PNI) was present in 81 of 133 specimens (61%). No statistically significant association was found between PNI and Gleason score concordance outcome (chi-square test, p=0.073).

Discussion

Our overall exact agreement rate of 60.9% aligns reasonably well with the broader literature, which typically reports exact agreement ranging from 55% to 75% (5,7). The weighted Cohen's kappa ($\kappa = 0.656$, 95% CI 0.575–0.738) indicates good (substantial) agreement. This suggests that, for the majority of patients, the biopsy grade serves as a sound foundation for immediate clinical management.

Rate of Upgrading and Downgrading

The observation that upgrading occurred in 29.3% of cases is clinically important and comparable to international series (4). Gleason score upgrading signifies that initial risk stratification underestimated the true aggressiveness of the disease.

In the context of modern risk-stratification models such as the NCCN guidelines, a patient managed conservatively based on a biopsy Grade Group of 2 (Gleason 3+4) might be inappropriately managed if the true Grade Group is 3 or higher. This highlights the inherent limitations of tissue sampling.

Conversely, the lower rate of downgrading (9.8%) is also expected. Downgrading often occurs due to sampling bias, where the biopsy may have inadvertently sampled poorly differentiated areas adjacent to higher-grade cancer, or due to interpretive variance between the two specimens (8).

Predictors of Discordance

Biopsy tumour involvement percentage differed significantly across Gleason score concordance categories. However, because the Kruskal–Wallis test is an omnibus comparison, this finding indicates group differences rather than a confirmed directional relationship with upgrading. Further studies using multivariable modelling are required to determine whether tumour involvement percentage independently predicts Gleason score discordance.

None of the remaining clinical and pathological parameters examined - age, PSA, prostate volume, the time interval between biopsy and surgery, or the presence of perineural invasion (PNI) - showed a statistically significant association with concordance outcome (all P > 0.05). This pattern is consistent with several prior series: Abedi et al. and Mansouri et al. similarly found no significant relationship between age, PSA, or prostate volume and Gleason score discordance (5,6), and Barsky et al. found no independent association between PNI on biopsy and Gleason upgrading at prostatectomy in a multi-institutional analysis (11). Taken together, these findings suggest that histological tumour burden, rather than broader clinical or perioperative parameters, is the most relevant correlate of discordance in this cohort.

Clinical Implications

The observed association suggests that biopsy tumour involvement percentage may be considered for further investigation in future risk prediction models; however, its independent clinical utility remains to be established.

Emerging Directions

Beyond conventional histopathological parameters, recent work suggests that artificial intelligence-assisted digital pathology platforms can grade Gleason patterns from whole-slide biopsy images with accuracy approaching that of expert pathologists, offering a potential means of reducing the sampling- and

interpretation-related variability that underlies much of the discordance observed in this and prior studies (12,13); a 2025 external validation further demonstrated the feasibility of such algorithms specifically on radical prostatectomy specimens (14). In parallel, biopsy-based genomic classifiers have shown independent prognostic value in large randomised trial cohorts and may help refine risk stratification beyond grade alone, particularly in intermediate-risk disease where the clinical consequences of grade discordance are greatest (15). Prospective integration of histopathological tumour-burden metrics, such as those examined here, with these emerging imaging, computational, and genomic tools represents a promising direction for improving the accuracy of pre-treatment risk assessment.

Limitations

This study has several limitations. First, it was conducted at a single primary referral center, which may limit generalizability to other populations and practice settings. Second, the cohort size (n=133), while adequate for the primary concordance analysis, limited statistical power for some subgroup comparisons, particularly the smaller downgrading categories (n=3 for two-grade downgrading). Third, information on the precise anatomical location of biopsy cores within the prostate was not available, precluding analysis of concordance by sampling site. Fourth, biopsies were performed using conventional systematic transrectal technique; MRI-fusion targeted biopsy, which may reduce sampling error, was not used during the study period and was therefore not available for comparison. Fifth, tumor involvement percentage was derived from the pathologist's visual estimation in a subset of cases rather than standardized digital morphometry, which may introduce measurement variability. Finally, as a retrospective cohort restricted to patients who proceeded to radical prostatectomy, the results may not generalize to patients managed with active surveillance or other treatment modalities, and evolving biopsy and grading practices over the 2012–2024 study period could not be fully accounted for.

Conclusion

Needle biopsy provides a moderately reliable estimate of the Gleason score, evidenced by a weighted kappa of 0.656. However, nearly one-third of patients experienced clinically important pathological discordance, predominantly characterized by Gleason score upgrading in the radical prostatectomy specimen. Among the clinical and pathological parameters examined, only tumor involvement percentage was significantly associated with discordance in this cohort. The observed association suggests that biopsy tumor involvement percentage may be considered for further investigation in future risk prediction models; however, its independent clinical utility remains to be established.

Declaration

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available because of restrictions related to patient privacy but are available from the corresponding author on reasonable request and subject to institutional approval.

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding

The authors received no specific funding for this study.

Authors' contribution

HGP, GK, and HS contributed to the study conception and design. HGP, HS, HSh, KH, and GK contributed to data collection and data curation. HSh and GK contributed to statistical analysis and interpretation of the findings. HGP and GK drafted the manuscript. All authors critically revised the manuscript, approved the final version, and agreed to be accountable for all aspects of the work.

Acknowledgments

The authors would like to thank the Clinical Research Development Unit (CRDU) of Shahid Beheshti Hospital and Mrs Sakineh Kamali Ahangar for their cooperation and assistance throughout the study period. The authors also thank the pathology department staff at Shahid

Beheshti Hospital for their assistance with data retrieval and verification.

Ethics approval and consent to participate

The study was approved by the Institutional Review Board of Babol University of Medical Sciences (ethics code IR.MUBABOL. HRI.REC. 1403. 265). The requirement for informed consent was waived by the Ethics Committee because of the retrospective nature of the study.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018; 68(1): 7-30.

2. Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. J Urol. 1974; 111(1): 58-64.

3. Gleason DF, Mellinger GT. Prediction of Prognosis for Prostatic Adenocarcinoma by Combined Histological Grading and Clinical Staging. J Urol. 2017; 197(2S): S134-S139.

4. Schreiber D, Wong AT, Rineer J, Weedon J, Schwartz D. Prostate biopsy concordance in a large population-based sample: a Surveillance, Epidemiology and End Results study. J Clin Pathol. 2015; 68(6): 453-7.

5. Mansouri N, Msakni I, Gargouri F, Khiari R, Bouziani A, Laabidi B. Evaluation of concordance of Gleason score between prostate biopsy and radical prostatectomy. Tunis Med. 2018; 96(7): 430-6.

6. Abedi AR, Basiri A, Shakhssalim N, Sadri G, Ahadi M, Hojjati SA, et al. The discrepancy between needle biopsy and radical prostatectomy Gleason score in patients with prostate cancer. Urol J. 2021; 18(04): 395-9.

7. Epstein JI, Amin MB, Reuter VE, Humphrey PA. Contemporary Gleason grading of prostatic carcinoma: an update with discussion on practical issues to implement the 2014 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. Am J Surg Pathol. 2017; 41(4):e1-e7.

8. Montironi R, Mazzuccheli R, Scarpelli M, Lopez -Beltran A, Fellegara G, Algaba F. Gleason grading of prostate cancer in needle biopsies or radical prostatectomy specimens: contemporary approach, current clinical significance and sources of pathology discrepancies. BJU Int. 2005; 95(8): 1146-52.

9. Ariafar A, Rezaeian A, Zare A, Zeighami S, Hosseini SH, Nikbakht H-A, et al. Concordance between Gleason score of prostate biopsies and radical prostatectomy specimens and its predictive factors. Urologia. 2023; 90(2): 236-43.

10. Soenens C, Dekuyper P, De Coster G, Van Damme N, Van Eycken E, Quackels T, et al. Concordance Between Biopsy and Radical Prostatectomy Gleason Scores: Evaluation of Determinants in a Large-Scale Study of Patients Undergoing RARP in Belgium. Pathol Oncol Res. 2020; 26(4): 2605-12.

11. Barsky AR, Kraus RD, Carmona R, Santos PM, Li C, Schwartz LE, et al. Investigating association of perineural invasion on prostate biopsy with Gleason score upgrading at prostatectomy: a multi-institutional analysis. Cancer Med. 2020; 9(10): 3383-9.

12. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. Lancet Oncol. 2020; 2 1(2): 233-41.

13. Ström P, Kartasalo K, Olsson H, Solorzano L, Delahunt B, Berney DM, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. Lancet Oncol. 2020; 21(2): 222-32.

14. Schmidt B, Soerensen SJC, Bhambhvani HP, et al. External validation of an artificial intelligence model for Gleason grading of prostate cancer on prostatectomy specimens. BJU Int. 2025; 135(1): 133-9.

15. Spratt DE, Liu VYT, Michalski J, et al. Genomic Classifier performance in intermediate-risk prostate cancer: results from NRG Oncology/RTOG 0126 randomized phase 3 trial. Int J Radiat Oncol Biol Phys. 2023; 117(2): 370-7