

# A second opinion pathology review improves the diagnostic concordance between prostate cancer biopsy and radical prostatectomy specimens

Takanori Maehara, Takuya Sadahira, Yuki Maruyama, Koichiro Wada, Motoo Araki, Masami Watanabe, Toyohiko Watanabe, Hiroyuki Yanai<sup>1</sup>, Yasutomo Nasu

Departments of Urology and <sup>1</sup>Pathology, Graduate School of Medicine Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, Japan

## Abstract

**Objectives:** The Gleason scoring system is an essential tool for determining the treatment strategy in prostate cancer (PCa). However, the Gleason grade group (GGG) often differs between needle-core biopsy (NCB) and radical prostatectomy (RP) specimens. We investigated the diagnostic value of a second opinion pathology review using NCB specimens in PCa.

**Materials and Methods:** We retrospectively evaluated 882 patients who underwent robot-assisted RP from January 2012 to September 2019. Of these, patients whose original biopsy specimens were obtained from another hospital and reviewed by the urological pathology expert at our institution were included in the study. Patients who received neoadjuvant hormonal therapy were excluded from the study. Weighted kappa (k) coefficients were used to evaluate the diagnostic accuracy of each review.

**Results:** A total of 497 patients were included in this study. Substantial agreement (weighted  $k = 0.783$ ) in the GGG between initial- and second-opinion diagnoses based on NCB specimens was observed in 310 cases (62.4%). Although diagnoses based on a single opinion showed moderate agreement with the GGG of RP specimens (initial: 35.2%, weighted  $k = 0.522$ ; second opinion; 38.8%, weighted  $k = 0.560$ ), matching initial and second opinion diagnoses improved the concordance (42.9%, 133/310 cases) to substantial agreement (weighted  $k = 0.626$ ).

**Conclusions:** A second opinion of PCa pathology helps to improve the diagnostic accuracy of NCB specimens. However, over half of diagnoses that matched between the initial and second opinions differed from the diagnosis of RP specimens.

**Keywords:** Gleason grade group, Gleason score, prostate biopsy, prostate cancer, prostatectomy

**Address for correspondence:** Dr. Takuya Sadahira, 2-5-1, Shikata-cho, Kita-ku, Okayama 700-8558, Japan.

E-mail: t.sadahira@gmail.com

**Received:** 26.05.2020, **Accepted:** 25.08.2020, **Published:** 04.03.2021

## INTRODUCTION

The Gleason score (GS) is the most important parameter used to predict tumor aggressiveness and select the

appropriate therapeutic management strategy for prostate cancer (PCa).<sup>[1]</sup> An accurate diagnosis is especially essential for PCa, as there are several treatments available for

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:** WKHLRPMedknow\_reprints@wolterskluwer.com

**How to cite this article:** Maehara T, Sadahira T, Maruyama Y, Wada K, Araki M, Watanabe M, *et al.* A second opinion pathology review improves the diagnostic concordance between prostate cancer biopsy and radical prostatectomy specimens. *Urol Ann* 2021;13:119-24.

Access this article online	
Quick Response Code:	Website: www.urologyannals.com
	DOI: 10.4103/UA.UA_81_20

this cancer.<sup>[2]</sup> However, previous studies have reported a significant discrepancy in the GS between needle core biopsy (NCB) and radical prostatectomy (RP) specimens (28%–76% of cases).<sup>[3–5]</sup> In many of these cases, the GS is underestimated in NCB specimens, which leads to an upgraded diagnosis in up to 43% of patients after RP.<sup>[5,6]</sup> These diagnostic errors are associated with multiple factors, including the number and length of the biopsy cores obtained, tumor location, pathologist misinterpretation, and interobserver variability.<sup>[3,4,7]</sup> To improve accuracy, a mandatory second opinion review by urological pathology experts has been recommended in several publications.<sup>[3,8,9]</sup> Such reviews can result in a change in GS from the original diagnosis in a significant number of specimens, leading to an alteration in therapeutic management.<sup>[7,8,10]</sup> At our institution, pathology slides of biopsy specimens and reports from previous institutions are routinely reviewed by a specialized urological pathologist before further clinical management, as a mandatory review program.<sup>[11]</sup>

A new PCa pathological grading system was recently proposed to reflect tumor biology better than the previous Gleason scoring system.<sup>[1,10]</sup> This system comprises five modified grading groups. Specifically, GS 3 + 3 disease is categorized as Gleason grade group (GGG) 1, GS 3 + 4 disease as GGG2, GS 4 + 3 disease as GGG3, GS 4 + 4, 3 + 5, and 5 + 3 disease as GGG4, and GS 4 + 4, 5 + 4, and 5 + 5 disease as GGG5. In a validation study that included <25,000 PCa patients, the GGG resulted in a more accurate prognostic prediction, increasing the C statistic by 0.02–0.05 compared with the three GS groups ( $\leq 6$ , 7, and 8–10).<sup>[12]</sup>

To date, only a few reports have evaluated the value of a second pathology review using this new grading system. In this study, we determined whether a second-opinion review is still mandatory for improving GS accuracy between NCB and RP specimens in the GGG era.

## MATERIALS AND METHODS

We retrospectively collected data from 882 PCa patients who underwent robot-assisted RP (RARP) at Okayama University Hospital between January 2012 and September 2019. NCB slides, which were prepared elsewhere, were routinely reviewed by a urological pathology expert at our hospital. Of these 882 patients, 497 fulfilled all inclusion criteria for this study: (a) initial NCB pathology evaluation conducted at a different hospital (patients who underwent the initial NCB at our hospital were excluded), (b) second opinion evaluation of the same specimen conducted at our hospital, (c) an available GGG (specimens from a different

hospital with an inadequate grade (e.g., GS 1 + 2) were excluded), and (d) no hormonal therapy received before RARP. We collected data on age, body mass index, prostate volume, initial prostate-specific antigen (PSA) level, PSA density, and numbers of biopsy cores and positive cores. Formal ethical approval for this study was obtained from the Okayama University Institutional Review Board (registration no. 1004) before study initiation. All patients provided written informed consent for the use of their clinical records.

NCB slides from elsewhere were reviewed, and evaluations of RP specimens were performed by the same urological pathology expert from our hospital, who was blinded to the previous NCB results. A GS was assigned to each lesion based on the sum of the primary and secondary patterns of the lesion; for patients with more than one positive lesion, the highest GS among all NCB and RP specimen evaluations was adopted. The GS assigned to each NCB specimen was compared with that assigned to the corresponding RP specimen, and a major GS discrepancy was defined as a difference in the GGG category between the two specimen types. The five GGG categories are GGG1 (GS 3 + 3), GGG2 (GS 3 + 4 = 7), GGG3 (GS 4 + 3 = 7), GGG4 (GS 8), and GGG5 (GS 9 or 10).<sup>[1,8]</sup>

To determine whether a second opinion pathology review improves the diagnostic accuracy, the agreement in the GGG between the NCB and RP specimens was calculated according to weighted Cohen's kappa ( $\kappa$ ) coefficients ( $5 \times 5$  tables; linear weights for categorical variables and quadratic weights for ordinal variables).  $\kappa = 1$  indicates that the raters are in complete agreement.  $\kappa = 0$  indicates no agreement among the raters other than what would be expected by chance. The agreement was categorized as almost perfect ( $\kappa \geq 0.81$ ), substantial ( $\kappa = 0.61$ – $0.80$ ), moderate ( $\kappa = 0.41$ – $0.60$ ), fair ( $\kappa = 0.21$ – $0.40$ ), or slight ( $\kappa \leq 0.20$ ). The statistical analyses were performed using (Saitama Medical Center, Jichi Medical University, Saitama, Japan), a graphical user interface for R.<sup>[13]</sup>

## RESULTS

From January 2012 to September 2019, 882 patients underwent RARP for PCa. Of these patients, 497 with available outside initial opinion, internal second opinion, and internal RP evaluations were included in the analysis. For the diagnosis of PCa, transrectal ultrasound-guided biopsy (TRUS-GB) was performed; no magnetic resonance imaging-guided biopsy (MRI-GB) or saturation template biopsy was performed. The characteristics of the patients and GGG distribution among the specimens are shown in Table 1. The rates of GGG1 and GGG4

were decreased in the RP compared with the NCB specimens, and that of GGG5 tended to be increased in the RP specimens. However, the median GGG was the same among the initial and second opinion NCB and RP specimens: three (interquartile range, 2–4). One case, which was diagnosed as GGG4 by the initial pathologist and as GGG2 by the second pathologist, was diagnosed with high-grade prostatic intraepithelial neoplasia after RARP.

Discrepancies in the GGG of the NCB specimens between the initial and second opinions are listed in Table 2. Of the 497 cases, the agreement was observed in 310 (62.3%). Among the 187 cases with a major discrepancy, the diagnosis was upgraded in 119 (63.6%) and downgraded in 68 (36.4%) after review. Both the GGG1 and GGG5 assignments had a higher rate of agreement (both 69%) between the initial and second opinions, whereas agreement

between the initial and second opinions for GGG3 was observed in only half of the cases.

Discrepancies in the GGG of the initial and second opinion NCB specimens compared with the RP specimens are shown in Tables 3 and 4, respectively. GGG agreement was observed in 35.2% (175/497) of cases between the initial opinion NCB and RP specimens and in 38.8% (193/497) of cases between the second opinion NCB and RP specimens. The rate of GGG upgrade after RARP was higher for the initial opinion evaluation (41.0%; 204/497) than second opinion evaluation (33.4%; 166/497).

The GGG concordance rates between the RP specimens and the NCB specimens in which the first and second pathologists' assignments were the same ( $n = 310$ ) are summarized in Table 5. In this analysis, the GGG concordance rate between RP and NCB specimens improved to 42.9% (133/310), compared with the comparisons based on the NCB initial or second opinions alone. Of the cases without concordance, the GGG was upgraded in 35.2% (109/310) and downgraded in 21.9% (68/310) of cases after RARP. Specifically, 50% (42/84) and 20% (17/84) of the GGG1 cases were upgraded to GGG2 and an even higher-grade group, respectively.

To determine the diagnostic improvement potential of second opinion pathology review, weighted Cohen's kappa coefficients were used [Table 6]. Initial and second

**Table 1: Demographic and clinical characteristics of the overall cohort (n=497)**

	n (%)		
Age, median (IQR)	68 (64-72)		
BMI, median (IQR)	23.7 (22.0-25.5)		
Prostate volume, median (IQR)	28 (21-37)		
Initial PSA, median (IQR)	7.4 (5.3-10.9)		
PSA density, median (IQR)	0.28 (0.19-0.44)		
Transrectal/transperineal biopsy, n (%)	352 (70.8)/145 (29.2)		
Clinical T-stage, n (%)			
T1	100 (20)		
T2	365 (73)		
T3	32 (7)		
Number of biopsy cores, median (IQR)	11 (10-14)		
Positive cores on initial opinion, median (IQR)	3 (2-5)		
Positive cores on second opinion, median (IQR)	3 (2-5)		
Gleason grade group	Initial, n (%)	Second, n (%)	Pathology, n (%)
Atypical	-	-	1 (1)
1	123 (25)	105 (21)	38 (8)
2	122 (25)	106 (21)	183 (36)
3	75 (15)	82 (16)	122 (25)
4	121 (24)	140 (29)	54 (11)
5	56 (11)	64 (13)	99 (20)

BMI: Body mass index, IQR: Interquartile range, PSA: Prostate specific antigen

**Table 2: Agreement in the Gleason grade group between the initial and second evaluations of needle core biopsy specimens**

Initial GGG	Second GGG					Total
	GGG 1	GGG 2	GGG 3	GGG 4	GGG 5	
GGG1 (≤3+3)	84 (69)	23 (19)	8 (6)	8 (6)	0 (0)	123
GGG2 (3+4)	17 (14)	70 (57)	17 (14)	17 (14)	1 (1)	122
GGG3 (4+3)	1 (1)	8 (11)	38 (51)	21 (28)	7 (9)	75
GGG4 (8)	3 (2)	5 (4)	17 (14)	79 (66)	17 (14)	121
GGG5 (9, 10)	0 (0)	0 (0)	2 (4)	15 (27)	39 (69)	56
Total	105	106	82	140	64	497

GGG: Gleason grade group

**Table 3: Agreement in the Gleason grade group between radical prostatectomy specimens and the initial evaluation of needle core biopsy specimens**

Initial GGG	Atypical	Radical prostatectomy GGG					Total
		GGG 1	GGG 2	GGG 3	GGG 4	GGG 5	
GGG1 (≤3+3)	0 (0)	29 (24)	63 (50)	18 (15)	6 (5)	7 (6)	123
GGG2 (3+4)	0 (0)	6 (5)	71 (58)	29 (24)	6 (5)	10 (8)	122
GGG3 (4+3)	0 (0)	2 (3)	25 (33)	20 (27)	16 (21)	12 (16)	75
GGG4 (8)	1 (1)	1 (1)	15 (12)	45 (36)	22 (18)	37 (31)	121
GGG5 (9, 10)	0 (0)	0 (0)	9 (16)	10 (18)	4 (7)	33 (59)	56
Total	1	38	183	122	54	99	497

GGG: Gleason grade group

**Table 4: Agreement in the Gleason grade group between radical prostatectomy specimens and second-opinion review of needle core biopsy specimens**

Second GGG	Atypical	Radical prostatectomy GGG					Total
		GGG 1	GGG 2	GGG 3	GGG 4	GGG 5	
GGG 1 (≤3+3)	0 (0)	29 (28)	52 (49)	13 (12)	5 (5)	6 (6)	105
GGG 2 (3+4)	1 (1)	6 (6)	67 (63)	22 (20)	4 (4)	6 (6)	106
GGG 3 (4+3)	0 (0)	1 (1)	32 (39)	28 (35)	11 (13)	10 (12)	82
GGG 4 (8)	0 (0)	2 (1)	24 (17)	48 (34)	29 (21)	37 (27)	140
GGG 5 (9, 10)	0 (0)	0 (0)	8 (13)	11 (17)	5 (8)	40 (62)	64
Total	1	38	183	122	54	99	497

GGG: Gleason grade group

**Table 5: Agreement in the Gleason grade group between needle core biopsy (with matching initial- and second-opinion Gleason grade group) and radical prostatectomy specimens**

1 <sup>st</sup> =2 <sup>nd</sup> GGG	Radical prostatectomy GGG					Total
	GGG 1	GGG 2	GGG 3	GGG 4	GGG 5	
GGG 1 (≤3+3)	25 (30)	42 (50)	9 (10)	4 (5)	4 (5)	84
GGG 2 (3+4)	2 (3)	50 (71)	13 (19)	1 (1)	4 (6)	70
GGG 3 (4+3)	0 (0)	15 (39)	13 (34)	6 (16)	4 (11)	38
GGG 4 (8)	1 (1)	6 (8)	32 (41)	18 (23)	22 (27)	79
GGG 5 (9, 10)	0 (0)	5 (13)	5 (13)	2 (5)	27 (69)	39
Total	28	118	72	31	61	310

GGG: Gleason grade group

**Table 6: The diagnostic concordance between specimens and the weighted kappa coefficients**

	Concordance rate	Weighted kappa coefficient (95% CI)
Initial×second opinion	62.3% (310/497)	0.783 (0.743-0.822)
Initial×RP specimen	35.2% (175/497)	0.522 (0.453-0.592)
Second opinion×RP specimen	38.8% (193/497)	0.560 (0.495-0.627)
1 <sup>st</sup> =2 <sup>nd</sup> ×RP specimen	42.9% (133/310)	0.626 (0.547-0.706)

CI: Confidence interval, RP: Radical prostatectomy

opinion GGGs showed substantial agreement (weighted  $k = 0.783$ ). The weighted  $k$  value improved from 0.522 when based on the initial opinion to 0.560 when based on the second opinion, both values indicating “moderate” agreement. In addition, when only the NCB specimens exhibiting GGG agreement between the initial and second opinions were compared with the RP specimens, the agreement ( $k = 0.626$ ) improved to “substantial.”

## DISCUSSION

The present study demonstrates the efficacy of second opinion pathology review for improving the GGG concordance between NCB and RP specimens. The accuracy of the initial opinion GGG was improved by the second opinion review (from  $k = 0.52$  to  $k = 0.56$ ). The GGG concordance further improved to “substantial” ( $k = 0.63$ ) when the RP specimens were compared exclusively with the NCB specimens exhibiting agreement between the initial and second opinions. However, it should be noted that the GGG in over half of those NCB specimens was discordant with that of the RP specimens.

The GGG has been identified as the most important independent risk factor for PSA recurrence-free survival and strongly influences PCa treatment strategies.<sup>[12,14]</sup> Thus, an accurate determination of the GGG in NCB specimens is critical for the management of PCa patients. Interpretation bias and sampling error are two major factors potentially affecting the GS interpretation in NCB specimens.<sup>[15,16]</sup>

Regarding interpretation bias, numerous studies have investigated the reproducibility of this classification

tool across pathologists, with GS agreement values ( $k$ ) ranging from 0.41–0.64 in NCB specimens.<sup>[7,17]</sup> Such bias is caused by the different interpretations of the gray areas between adjacent grades of the GS system, particularly between GS 3 and GS 4.<sup>[18]</sup> A previous study revealed that variations in the interpretation of criteria result in a lower rate of agreement: Only 9.9% of 71 specimens exhibited total agreement among three pathologists, with a total disagreement rate of 26.8%.<sup>[18]</sup> In addition, it was reported that at least 17% of cases of GS overestimation in NCB specimens, compared with prostatectomy specimens, are due to the misinterpretation of the pathologist.<sup>[4]</sup> To resolve these problems, a consensus diagnosis among two or more pathologists has been recommended.<sup>[19]</sup> Numerous publications showed that mandatory second opinion pathology review by specialized urological pathologists leads to alteration of the management strategy and improved care in some cases.<sup>[3,7,8,11,20]</sup> The accuracy of the GS in NCB specimens depends on the experience in uropathology and workload of the reviewing pathologist.<sup>[3]</sup> A major discrepancy in the GS assigned by general versus specialized pathologists occurs in 15%–41% of NCB specimens.<sup>[8,20]</sup> The adoption of second opinion pathology review improves the GS agreement between NCB and RP specimens, and thus GS accuracy, and leads to a change in the treatment recommendation in 9%–26% of cases.<sup>[21]</sup> In our study, the GGG agreement in NCB specimens between the initial and second opinions was relatively high (weighted  $k = 0.78$ ) compared with previous reports (0.41–0.64).<sup>[7,17]</sup> Hence, an improvement in GGG accuracy was observed in a small number of cases (35%–39%). This may be explained by the improvements in GS accuracy of general pathologists with increasing experience; one study reported that the concordance of GSs assigned by general pathologists was significantly higher during the second half of their 13-year training period.<sup>[21]</sup> Our concordance rate between NCB and RP specimens was similar to others (35% and 39%), and a tendency for a higher GGG in the second opinion review was also observed.

Even with mandatory second-opinion pathology reviews, it is difficult to control interpretation bias. In this study, 57% of NCB cases with a matching GGG between the initial and second opinions were discordant with the GGG of RP specimens. Similarly, Gleason himself in 1992 reported an exact GS agreement in only 50% of cases after a second review of the same specimens used for his original classification.<sup>[22]</sup> To overcome this bias, the diagnoses of the second-opinion NCB specimen and RP specimen were determined by the same urological pathology expert using the same criteria; furthermore, the diagnosis of the NCB specimens was performed before

that of the RP specimens, and no retrospective change in the NCB diagnosis was conducted in the current study. Thus, the inaccuracy might have been caused by sampling error rather than interpretation bias.

Sampling error is caused by the extensive, multifocal, and heterogeneous characteristics of PCa, which render proper sampling of the prostate gland difficult.<sup>[23]</sup> Significant sampling variation occurs with the use of a systematic number of biopsy cores from prostate glands that fluctuate in volume. Thus, increasing the number of biopsy cores improves both sampling and accuracy.<sup>[11]</sup> Conversely, a small number of biopsy cores with a shorter length might lead to the overestimation of the GGG in NCB specimens. Another way to reduce sampling error is to improve tumor visualization and targeting during the biopsy. To this end, MRI-GB has been rapidly introduced worldwide.<sup>[5,23]</sup> Several studies showed a similar cancer detection rate between MRI-GB and TRUS-GB; however, the GS of RP specimens has a higher concordance with that of MRI-GB (57%–90%) than that of TRUS-GB specimens (28%–76%).<sup>[3-5,23]</sup> MRI-GB also leads to a shift in the GGG distribution of newly diagnosed PC patients toward a diagnosis of higher-risk disease, which allows detection of 30% more high-risk cancers and 17% fewer low-risk cancers than those detected by systemic biopsy.<sup>[24]</sup> Similarly, Xu *et al.* reported that 32% of TRUS-GB cases were upgraded, compared with only 21% of MRI-GB cases, following analysis of RP specimens.<sup>[5]</sup>

With the growing availability of prostate MRI, different functional imaging modalities have increased the role of prostate MRI in detecting, locating, and staging PCa.<sup>[25]</sup> The adoption of multi-parametric MRI can improve the detection of PCa, with a specificity of 0.88 and a sensitivity of 0.74.<sup>[26]</sup> The Prostate Imaging Reporting and Data System, version 2, (PI-RADS) score was proposed to detect PCa with the aim of increasing multiparametric MRI efficacy.<sup>[25]</sup> A high PI-RADS score can predict >80% of cases with clinically significant PCa.<sup>[25,26]</sup> The present results suggest that the adoption of multiparametric MRI is important, in addition to a second opinion pathology review.

The present study had several limitations. First, only cases treated with RARP were included. Cases diagnosed as benign or deemed unsuitable for RARP are not referred to our institution. Such selection bias may have increased the potential for overestimating the GGG. Second, the biopsy method was not unified, as the NCB specimens were obtained from different institutions; thus, the number of cores and the biopsy location differed among the

initial institutions. Third, the same urological pathologist from our hospital determined the GS for both the NCB and RP specimens. Therefore, interpretation bias might have influenced the agreement between the initial and subsequent diagnoses.

## CONCLUSIONS

Second opinion pathology review can improve the GGG concordance between NCB and RP specimens. However, even in the comparison limited to the NCB cases with a matching GGG between the initial and second opinions, the GGG rate was different after RARP in over half of the cases.

## Acknowledgments

The authors would like to thank the clinical laboratory technicians of Okayama University Hospital for their technical support.

## Financial support and sponsorship

Nil.

## Conflicts of interest

There are no conflicts of interest.

## REFERENCES

1. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA, *et al.* 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.
2. Sadahira T, Mitsui Y, Araki M, Maruyama Y, Wada K, Edamura K, *et al.* Pelvic magnetic resonance imaging parameters predict urinary incontinence after robot-assisted radical prostatectomy. *Low Urin Tract Symptoms* 2019;11:122-6.
3. Barqawi AB, Turcanu R, Gamito EJ, Lucia SM, O'Donnell CI, Crawford ED, *et al.* The value of second-opinion pathology diagnoses on prostate biopsies from patients referred for management of PCa. *Int J Clin Exp Pathol* 2011;20:468-75.
4. Ruijter E, van Leenders G, Miler G, Debruyne F, van de Kaa C. Errors in histological grading by prostatic needle biopsy specimens: Frequency and predisposing factors. *J Pathol* 2000;192:229-33.
5. Xu N, Wu YP, Li XD, Lin MY, Zheng QS, Chen SH, *et al.* Risk of upgrading from prostate biopsy to radical prostatectomy pathology: Is magnetic resonance imaging-guided biopsy more accurate? *J Cancer* 2018;9:3634-9.
6. Borkowetz A, Platzek I, Toma M, Renner T, Herout R, Baunacke M, *et al.* Direct comparison of multiparametric magnetic resonance imaging (MRI) results with final histopathology in patients with proven prostate cancer in MRI/ultrasonography-fusion biopsy. *BJU Int* 2016;118:213-20.
7. Truesdale MD, Cheetham PJ, Turk AT, Sartori S, Hruby GW, Dinneen EP, *et al.* Gleason score concordance on biopsy-confirmed prostate cancer: Is pathological re-evaluation necessary prior to radical prostatectomy? *BJU Int* 2011;107:749-54.
8. Brimo F, Schultz L, Epstein JI. The value of mandatory second opinion pathology review of prostate needle biopsy interpretation before radical prostatectomy. *J Urol* 2010;184:126-30.

9. Weir MM, Jan E, Colgan TJ. Interinstitutional pathology consultations. A reassessment. *Am J Clin Pathol* 2003;120:405-12.
10. Epstein JI, Walsh PC, Sanfilippo F. Clinical and cost impact of second-opinion pathology. Review of prostate biopsies prior to radical prostatectomy. *Am J Surg Pathol* 1996;20:851-7.
11. Maruyama Y, Sadahira T, Araki M, Mitsui Y, Wada K, Rodrigo AG, *et al.* Factors predicting pathological upgrading after prostatectomy in patients with Gleason grade group 1 prostate cancer based on opinion-matched biopsy specimens. *Mol Clin Oncol* 2019;12:384-9.
12. Epstein JI, Zelefsky MJ, Sjoberg DD, Nelson JB, Egevad L, Magi-Galluzzi C, *et al.* A contemporary prostate cancer grading system: A validated alternative to the gleason score. *Eur Urol* 2016;69:428-35.
13. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplant* 2013;48:452-8.
14. Loeb S, Folkvaljon Y, Robinson D, Lissbrant IF, Egevad L, Stattin P. Evaluation of the 2015 Gleason grade groups in a nationwide population-based cohort. *Eur Urol* 2016;69:1135-41.
15. Steinberg DM, Sauvageot J, Piantadosi S, Epstein JI. Correlation of prostate needle biopsy and radical prostatectomy Gleason grade in academic and community settings. *Am J Surg Pathol* 1997;21:566-76.
16. Ooi K, Samali R. Discrepancies in Gleason scoring of prostate biopsies and radical prostatectomy specimens and the effects of multiple needle biopsies on scoring accuracy. A regional experience in Tamworth, Australia. *ANZ J Surg* 2007;77:336-8.
17. Melia J, Moseley R, Ball RY, Griffiths DF, 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. McLean M, Srigley J, Banerjee D, Warde P, Hao Y. Interobserver variation in prostate cancer Gleason scoring: Are there implications for the design of clinical trials and treatment strategies? *Clin Oncol (R Coll Radiol)* 1997;9:222-5.
19. Stav K, Judith S, Merald H, Leibovici D, Lindner A, Zisman A. Does prostate biopsy Gleason score accurately express the biologic features of prostate cancer? *Urol Oncol* 2007;25:383-6.
20. Montironi R, Lopez-Beltran A, Cheng L, Montorsi F, Scarpelli M. Central prostate pathology review: Should it be mandatory? *Eur Urol* 2013;64:199-201.
21. Renshaw AA, Schultz D, Cote K, Loffredo M, Ziemba DE, D'Amico AV. Accurate Gleason grading of prostatic adenocarcinoma in prostate needle biopsies by general pathologists. *Arch Pathol Lab Med* 2003;127:1007-8.
22. Gleason DF. Histologic grading of prostate cancer: A perspective. *Hum Pathol* 1992;23:273-9.
23. Quentin M, Blondin D, Arsov C, Schimmoller L, Hiester A, Godehardt E, *et al.* Prospective evaluation of magnetic resonance imaging guided in-bore prostate biopsy versus systematic transrectal ultrasound guided prostate biopsy in biopsy naïve men with elevated prostate specific antigen. *J Urol* 2014;192:1374-9.
24. Siddiqui MM, Rais-Bahrami S, Turkbey B, George AK, Rothwax J, Shakir N, *et al.* Comparison of MR/ultrasound fusion-guided biopsy with ultrasound-guided biopsy for the diagnosis of prostate cancer. *JAMA* 2015;313:390-7.
25. Washino S, Okochi T, Saito K, Konishi T, Hirai M, Kobayashi Y, *et al.* Combination of prostate imaging reporting and data system (PI-RADS) score and prostate-specific antigen (PSA) density predicts biopsy outcome in prostate biopsy naïve patients. *BJU Int* 2017;119:225-33.
26. de Rooij M, Hamoen EH, Fütterer JJ, Barentsz JO, Rovers MM. Accuracy of multiparametric MRI for prostate cancer detection: A meta-analysis. *AJR Am J Roentgenol* 2014;202:343-51.