Original Article

Interobserver Reproducibility of Quantifying Gleason Pattern 4 Cancer in Prostate Biopsy: Implications for Clinical Practice

Jianhong Li[1], Mark Ettel[2], Ali Amin[3], Ritu Bhalla[4], Kasturi Das[5], Fang-Ming Deng[6], Peng Lee[6], Andres Matoso[7], Jonathan Melamed[6], Savvas Mendrinos[8], Wei Tian[9], Oksana Yaskiv[5], Rajal B. Shah[10] and Ming Zhou[11*]

[1]Geisinger Medical Laboratories, Geisinger Medical Center, Danville, PA, USA; [2]Department of Pathology and Laboratory Medicine, University of Rochester, Rochester, NY, USA; [3]Department of Pathology, Brown University, Providence, RI, USA; [4]Department of Pathology, Louisiana State University, New Orleans, LA, USA; [5]Department of Pathology and Laboratory Medicine, Zucker School of Medicine at Hofstra/Northwell, Lake Success, NY, USA; [6]Department of Pathology, New York University Langone Medical Center, New York, NY, USA; [7]Department of Pathology, Johns Hopkins University Hospital, Baltimore, MD, USA; [8]Dianon Systems, Shelton, CT, USA; [9]Inform Diagnostics, Irving, TX, USA; [10]Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX, USA; [11]Department of Pathology and Laboratory Medicine, Tufts Medical Center, Boston, MA, USA

Abstract

Background and objectives: The quantity of Gleason pattern (GP) 4 in Grade Group 2 prostate cancer (PCa) is an important prognostic factor and may influence treatment decisions. To be impactful for patient prognosis and management, GP4 quantification must be done in a uniform and reproducible way. This study investigated the interobserver reproducibility of GP4 quantification, and whether it may be affected by any histological features, including GP4 sub-patterns and tumor size. ***Methods:*** Glass slides containing 55 biopsy cores of various amounts of GP4 were distributed to 12 pathologists who quantified GP4 and determined the most common GP4 sub-patterns (poorly formed glands [P], fused glands [F], cribriform [C], and glomeruloid [G]) in each core. ***Results:*** The interobserver reproducibility for 12 pathologists to quantify GP4 in 55 biopsy cores was moderate (κ = 0.57). The κ value for cores with a PCa length ≤2 mm, 2.1–5 mm, and >5 mm was 0.51, 0.50, and 0.66, respectively. The κ value for cores with the most common sub-pattern (P, F, [C and G], no consensus) was 0.43. 0.57, 0.74, and 0.57, respectively. When the consensus of the percentage of GP4 was 41–50% and 51–60%, 35% and 41.7% of the individual measurements were significantly different from the consensus measurements. ***Conclusions:*** This study shows the reproducibility of quantifying GP4 PCa is moderate and significantly lower for small sized cancer with a predominant P pattern. There is significant variability in quantifying GP4 when it is 40–60%. These findings highlight the significant limitations of GP4 quantification that pathologists and clinicians must be aware of, and argue for more training for pathologists and standardization of methodology to improve the GP4 quantification.

Introduction

The quantity of Gleason pattern 4 (GP4) in Grade Group 2 (Gleason score 7) prostate cancer (PCa) is an important prognostic factor and may influence treatment decisions. Studies have shown that the quantity of GP4 in Grade Groups 2 and 3 PCa in prostate biopsies, measured in a percentage or linear length, correlates with adverse pathological findings in radical prostatectomies (RP), including extraprostatic extension, seminal vesicle invasion, lymph node metastases, positive surgical margins, and oncological outcomes (postsurgical biochemical progression).[1–9] For these reasons, reporting the quantity of GP4 in a needle biopsy has been recommended by the World Health Organization (WHO),[10] Genitourinary Pathology Society (GUPS),[11] International Society of Urological Pathology (ISUP),[12] and College of American Pathologists (CAP).[13]

For the GP4 quantity to be impactful for patient prognosis and management, quantification should be done in a uniform and reproducible way. However, the interobserver reproducibility of quantifying GP4 by pathologists has been rarely studied. It is also not known if the methodology of GP4

Table 1. Gleason pattern 4 quantification reproducibility in prostate cores stratified by cancer length

	N	icc/kappa value	95% CIL	95% CIU	P value
Group 1	7	0.51	0.07	0.73	n.a.
Group 2	23	0.50	0.30	0.60	Group 1 vs 2: >0.05
Group 3	25	0.66	0.50	0.76	Group 2 vs 3: <0.05

Cancer length in Group 1, 2 and 3 was ≤2, 2.1–5 and >5 mm. Icc, intraclass correlation; 95% CIL, 95% confidence intervals lower limit; 95% CIU, confidence intervals upper limit.

quantification and histological features, including GP4 sub-patterns, may affect the interobserver reproducibility. This study was designed to answer these questions and provide guidance to pathologists for reproducible quantification of GP4 in prostate biopsies.

Material and methods

Case selection

The study was performed in accordance with the ethical standards of the contributing authors' institutions and with the Declaration of Helsinki (as revised in 2013), and was approved by the lead author's Institutional Review Boards. The patients' consents were waived, as the slides were anonymized. Forty-seven glass slides containing 55 biopsy cores with PCa of various amounts of GP4 were selected for this study. Of the 12 participating pathologists, nine were practicing in university teaching hospitals and three in specialized urological pathology laboratories.

Review and quantification of Gleason pattern 4

De-identified glass slides were distributed to 12 participants who then reviewed the biopsy cores to confirm the cancer diagnosis. When GP4 cancer was present, the participants quantified the percentage of the GP4 of the entire cancer focus in the biopsy core (ranging from 1–100%), and documented the presence of GP4 sub-patterns (poorly formed glands [P], fused glands [F], cribriform [C], and glomeruloid [G]) and the most common sub-pattern in each biopsy core. Specific diagnostic criteria for P, F, C and G were not provided to the participants, who graded and quantified the biopsy cores based on their own experience. Eleven of the 12 participants quantified the GP4 percentage based on the areas of GP4 vs total cancer area, and one participant based the quantification on the length of GP4 vs total cancer length. For the final analysis, the mean percentage of GP4 measured by 12 participants was calculated as the consensus of GP4 quantity in each biopsy core. A GP4 sub-pattern was considered the most common pattern by the group consensus when identified as the most common pattern by ≥9 (75%) participants. Since the G sub-pattern was identified as the most common pattern in only one case, it was combined with the C sub-pattern for analysis. The length of the tumor in the biopsy cores was measured by the lead author of this study (MZ) in millimeters from one end to the other without excluding intervening benign glands.

Statistical analysis

Statistical analyses were performed using R software for statistical computing (R Foundation for Statistical Computing, Vienna, Austria). Interobserver reproducibility of quantifying GP4 by 12 participating pathologists was assessed by Fleiss k using the package "psy." The reproducibility of quantifying GP4 was also calculated for biopsy cores grouped based on the tumor length (≤2 mm, 2.1–5 mm and >5 mm) and the most common sub-pattern in each core (P, F, C/G, and no consensus). A *p* value <0.05 was considered statistically significant.

Results

The mean percentage of GP4 in 55 biopsy cores ranged from 6% to 82%. Thirty-eight cores contained Grade Group 2 PCa with GP4 ≤50%, while 17 contained Grade Group 3 PCa with GP4 >50%. The interobserver reproducibility for 12 pathologists to quantify GP4 in 55 biopsy cores was moderate (κ = 0.57).

The mean PCa length was 5.4 (range of 0.6–13) mm. None of the cores had discontinuous cancer involvement defined as cancer foci >3 mm apart.[14] These 55 cores were categorized into three groups based on the PCa length: ≤2 mm (Group 1), 2.1–5 mm (Group 2), and >5 mm (Group 3). There were seven cores in Group 1, 23 cores in Group 2, and 25 cores in Group 3. The quantification reproducibility of GP4 in these three groups is shown in Table 1 and Figure 1. The k value was 0.51, 0.50, and 0.66, respectively, for these three groups. The reproducibility was significantly higher for Group 3 than that for Groups 1 and 2 (*p* value was >0.05 for Group 1 vs 2, and <0.05 for Group 2 vs 3).

The P, F, C, and G glands (Fig. 2) were the most common sub-pattern in 18 (33%), seven (13%), nine (16%), and one (2%) of the cores, respectively. The remaining 20 (36%) had no consensus in terms of the most common sub-pattern. These 55 cores were categorized into four groups according to their most common sub-pattern (P, F, C/G, and no consensus). The κ value for these four groups was 0.43, 0.57, 0.74, and 0.57, respectively (*p* value <0.05 for P vs F and F vs [C/G]) (Table 2). The interobserver reproducibility for the C/G sub-patterns was significantly better than that for P and F.

We also investigated how often individual participant's quantification would fall in the same clinically meaningful GP4 range as the consensus measurements (0% [Grade Group 1], 1–10% and 11–50% [both would be graded as Grade Group 2], and ≥51% [Grade Group 3]) (Table 3). Quantification of a biopsy core by one pathologist was considered as one event. The number of classification events for any one core was 12. When the consensus of GP4 was 1–10% (Grade Group 2), there were 4/24 (17%) individual measurements as 0% (Grade Group 1), and 2/24 (8.3%) individual measurements as 11–50% (Grade Group 1); the latter was considered as misclassified. When the consensus of GP4 was 11–40% (Grade Group 2), there were 12/372 (3.2%) individual measurements as 0% (Grade Group 1), and 28 (7.5%) individual measurements as ≥51% (Grade Group 3); therefore, 10.8% (40/372) were considered as misclassified. When the consensus of GP4 was 41–50% (Grade Group 2), there were 21/60 (35%) individual measurements as >50% (Grade Group 3) that were considered misclassified. When the consensus of GP4 was 51–60% (Grade Group 3), there were 2/60 (3.3%) individual measurements as 0% (Grade
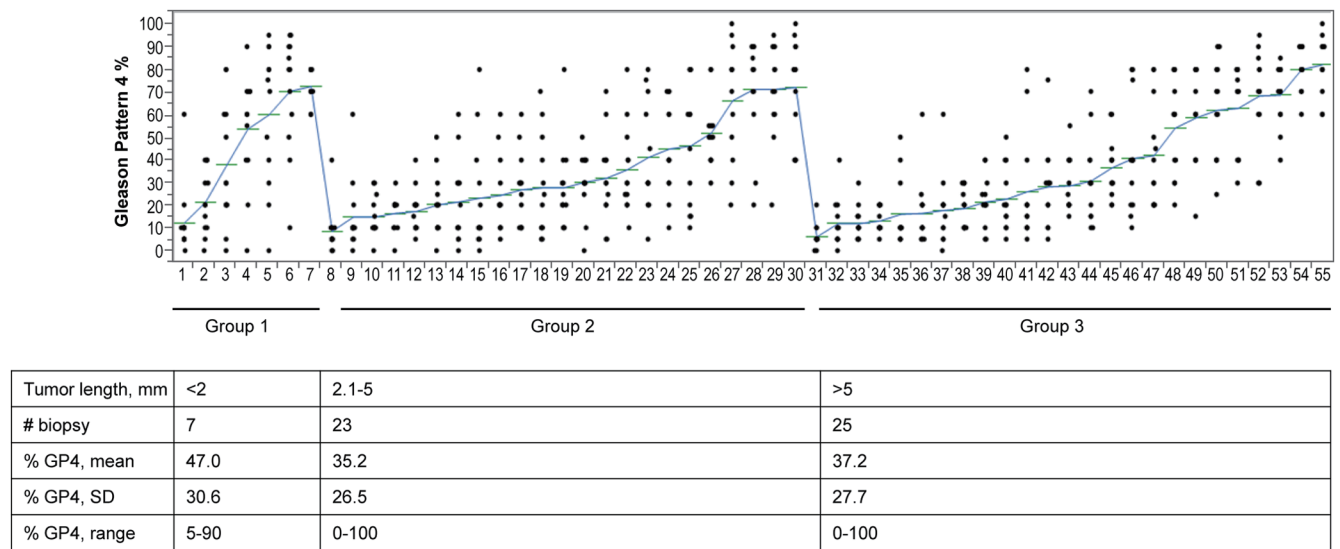
Tumor length, mm	<2	2.1-5	>5
# biopsy	7	23	25
% GP4, mean	47.0	35.2	37.2
% GP4, SD	30.6	26.5	27.7
% GP4, range	5-90	0-100	0-100

Fig. 1. The percentage of Gleason pattern 4 in the 55 biopsy cores. The 55 cores are displayed on the x-axis and are categorized into three groups based on the cancer length: Group 1, <2 mm; Group 2, 2.1–5 mm; Group 3, >5 mm. The y-axis represents the percentage of GP4. The dots represent the quantification by the 12 pathologists. The larger dots represent the quantification by more than one pathologist. The horizontal bars represent the mean GP4 percentage by the 12 participants.
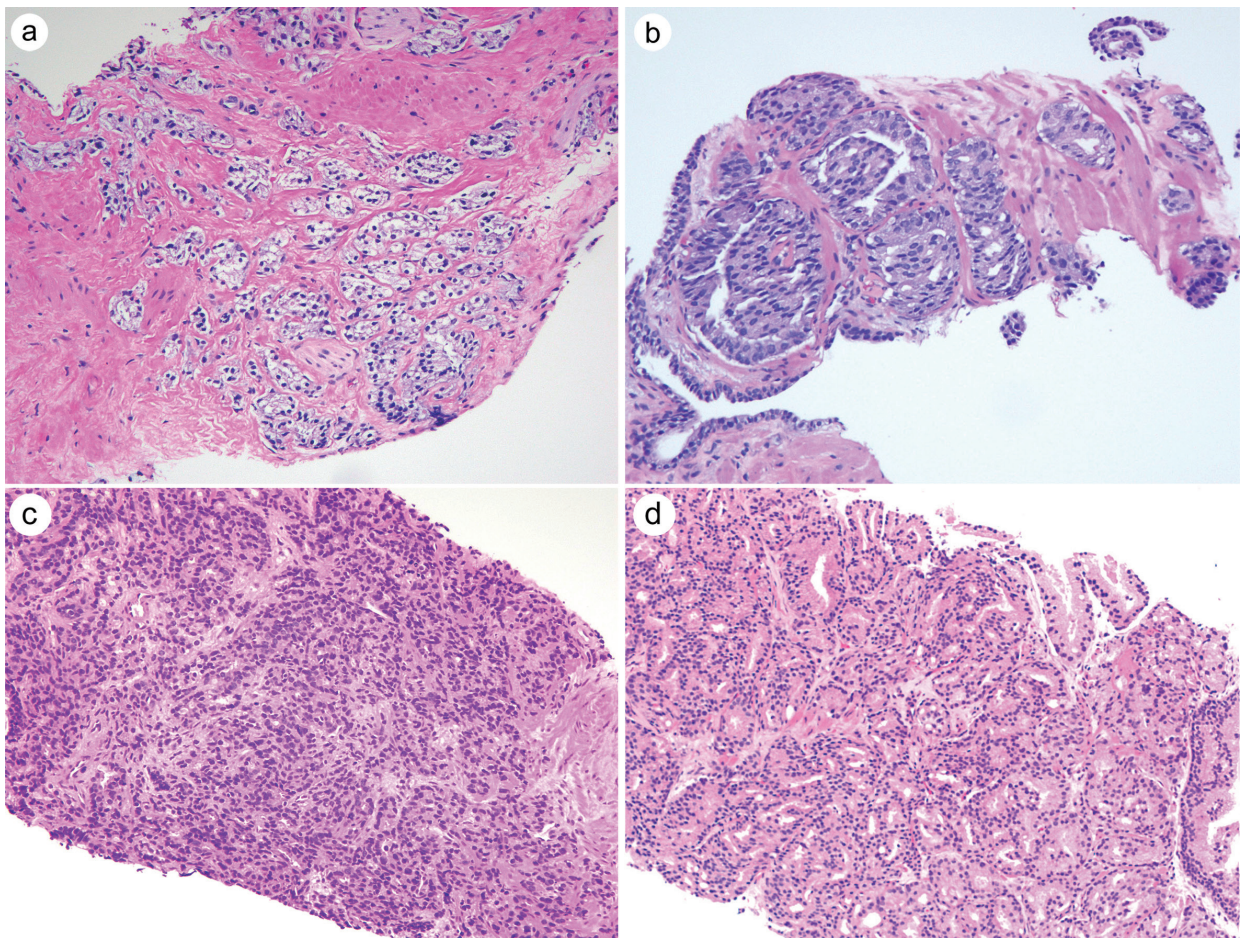

Fig. 2. Representative images of Gleason pattern 4 sub-patterns. These comprise the poorly formed glands (a), glomeruloid glands (b), fused glands (c) and cribriform glands (d). Each image was magnified to 100×.

Table 2. Gleason pattern 4 quantification reproducibility in prostate cores stratified by sub-patterns

	N	Icc/kappa value	95% CIL	95% CIU	P value
P	18	0.43	0.22	0.55	n.a.
F	7	0.57	0.12	0.77	P vs F: <0.05
C/G	10	0.74	0.38	0.88	F vs (C/G): <0.05
No consensus	20	0.57	0.39	0.67	n.a.

Biopsy cores were stratified based on the most common sub-pattern (P: poorly formed glands; F: fused glands; C or G: cribriform or glomeruloid; no consensus). icc, intraclass correlation; 95% CIL, 95% confidence intervals lower limit; 95%CIU, confidence intervals upper limit.

Group 1), and 21 (35%) individual measurements as 10–50% (Grade Group 2); therefore, 38.3% (23/60) were considered misclassified. When the consensus of GP4 was ≥61% (Grade Group 3), there were 19/144 (13.2%) individual measurements as 10–50% (Grade Group 2); therefore, they were considered misclassified. The misclassification rate was significantly higher when the consensus of GP4 measurements was 41–60% (Table 3).

Discussion

The GP4 quantity in Grade Group 2 PCa is an important prognostic factor. In Grade Group 2 PCa, GP4 is associated with the adverse histopathological outcomes and an increased risk of biochemical failure in patients undergoing radical prostatectomy.[1,2,4–6,8,9] In recent studies, Dean *et al*. and Perera *et al*. demonstrated several GP4 quantification methods in Grade Group 2 PCa, including the maximum percentage of GP4 in any single core. The overall percentage of GP4 (GP4 mm/total cancer mm) and total length of GP4 in mm in all cores were significantly associated with an increased risk of adverse pathology in RP and biochemical recurrence risk.[3,7] Cole *et al*. also reported that the incremental percentage of GP4 in the biopsies was an important predictor of adverse pathology and PSA recurrence across the entire range of GS 7–8 PCa.[2]

The quantity of GP4 has been used in determining the candidacy for active surveillance, which has been increasingly used for patients with the National Comprehensive Cancer Network (NCCN) very low-risk/low-risk PCa. The guidelines now also consider active surveillance for select favorable intermediate risk patients, which includes low-volume Grade Group 2 disease, depending on life expectancy and other clinical/radiologic factors, and for the selection of different radiation therapy protocols.[15] Therefore, WHO, GUPS, ISUP and CAP have all recommended including the percentage of GP4 in the prostate biopsy reports.[10–13]

However, only one study so far investigated the interobserver reproducibility and histological features of PCa that may affect GP4 quantification in prostate biopsies.[16] Sadimin *et al*. compared the quantification of GP4 by the primary author and his four trainees. It found that in 32% and 75% of cases, the GP4 quantification between the primary author and his trainees were exact match and within +10%, respectively, with a weighted kappa value (κw) = 0.67. No significant difference was observed when the cases were stratified based on (1) GP4 component was scattered vs clustered in the background of GP3 cancer, (2) cancer in the biopsy was continuous vs discontinuous, and (3) the cribriform/glomeruloid pattern only versus poorly formed/fused pattern versus mixed cribriform and poorly formed/fused pattern. However, kw for cases with >10% cancer involvement of the biopsy core was significantly higher than those with ≤10% involvement (0.70 vs 0.50). While it showed that the GP4 quantification can be reproducibly performed, and the GP4 quantification in a small focus of Grade Group PCa is less reliable, the study has several limitations. There were only five participants in the study, and they were the primary author and his four trainees at the same institution. To understand the true degree of and the factors that affected the interobserver reproducibility of quantifying GP4, survey of more pathologists with more diverse practice settings is needed.

In the current study, the interobserver reproducibility for 12 pathologists to quantify GP4 in 55 biopsy cores was moderate (κ = 0.57). This k value is lower than Sadimin's study (0.67). The difference could be explained by the aforementioned reason that participants in that study comprised the primary author and four of his trainees at one institution, while 12 pathologists in this study had diverse training and practice background. Our study may still have overestimated the interobserver reproducibility as most of the participants in our study had GU fellowship training and are practicing in a GU subspecialty signout. Sadimin's and our present studies do, however, suggest that training can significantly improve the reproducibility of the GP4 quantification.

We also found that the interobserver reproducibility in the biopsy cores with a tumor length >5 mm is significantly better than that for the cores with tumor lengths ≤2 mm

Table 3. Quantification of Gleason pattern 4 in prostate biopsies by individual pathologists stratified by consensus Gleason pattern 4 measurements

Consensus GP 4 measurement (%)	# biopsy cores	Total # classification events*	Classified as GP4% = 0, # (%)	Classified as GP4% 1–10, # (%)	Classified as GP4% = 11–50, # (%)	Classified as GP4% ≥ 51, # (%)	% mis-classification, % (#/total)	P value for mis-classification
1–10	2	2 × 12 = 24	4 (17)	18 (75)	2 (8.3)	0 (0)	8.3 (2/24)	<0.0001
11–40	31	31 × 12 = 372	12 (3.2)	132 (35.5)	200 (53.8)	28 (7.5)	10.8 (40/372)	
41–50	5	5 × 12 = 60	0 (0)	4 (6.7)	35 (58.3)	21 (35.0)	35 (21/60)	
51–60	5	5 × 12 = 60	2 (3.3)	0 (0)	23 (38.3)	35 (62.7)	41.7 (25/60)	
≥61	12	12 × 12 = 144	0 (0)	0 (0)	19 (13.2)	125 (86.8)	13.2 (19/144)	

*Quantification of a biopsy core by one pathologist is considered one event. The number of classification events for any core was 12. Total number of classification events = # of biopsy cores ×12.

and 2.1–5 mm (κ value 0.66 vs 0.50 and 0.51), indicating that GP4 quantification in a larger cancer focus is more reproducible. This finding is similar, although not identical, to Sadimin's study that showed the kw for cases with >10% involvement of the core was significantly higher than those with ≤10% involvement. We therefore advocate caution when quantifying the percentage of GP4 in a small focus of cancer. A recent survey by GUPS reported no consensus in quantifying the percentage of GP4 in needle biopsies with low-volume cancer with 58% of pathologists assigning and 42% not assigning the percentage of GP4.[11] A practical approach is to assign either a Grade Group 2 or 3 with a comment that the cancer focus is too small to accurately quantify the percentage of GP4.

We also studied the impact of the sub-patterns on the reproducibility of GP4 quantification. We showed that poorly formed glands are the most common GP4 sub-pattern in 18 (33%) cores and cribriform glands are the most common sub-pattern in nine (16%) cores. The reproducibility of GP4 quantification varies significantly among the four groups with predominantly poorly formed glands, fused glands, cribriform/glomeruloid gland, or no consensus. The biopsy cores with the poorly formed glands demonstrate lowest κ value (0.43), while the cribriform/glomeruloid glands exhibit the highest κ value (0.74). The low k value associated with the quantifying poorly formed glands may be explained by the significant grading variation of the poorly formed glands among the pathologists. Our previous study found that the poorly formed glands sub-pattern suffers definitional ambiguity and has the lowest diagnostic reproducibility (κ = 0.34).[17] It is obvious that grading and quantifying poorly formed glands need more standardization.

The reproducibility of the quantifying cores with the fused glands as the most common sub-pattern is fair (κ = 0.57; better than that of the poorly formed glands but worse than the cribriform/glomeruloid glands. The biopsy cores with the cribriform/glomeruloid patterns as the most common pattern have the substantial and highest reproducibility in the GP4 quantification (κ = 0.74). This finding is clinically relevant and assuring as recent studies found that the cribriform sub-pattern is associated with a worse prognosis than other GP4 sub-patterns and independently predicts biochemical recurrence and risk of distant metastasis.[1,18,19] Our study suggests that the most common GP4 sub-pattern may be reported along with the percentage of GP4, as the quantification reproducibility for different GP4 sub-patterns differs significantly with the cribriform pattern having the highest quantification reproducibility.

We also investigated how often individual participant's quantification would fall in the same clinically meaningful GP4 range as the consensus measurements (0%, 1–10%, 11–50%, and ≥51%), which would imply that individual participant's Grade Group is concordant with the Grade Group based on the consensus. When the consensus of GP4 was 1–10% (Grade Group 2), there were 4/24 (17%) individual measurements of 0% (Grade Group 1). Several studies found that Grade Group 2 PCa with a minor GP4 component has pathological features and clinical outcomes similar to Grade Group 1 PCa,[6,20] and patients with Grade Group 2 PCa with a minor GP4 component may still be eligible for active surveillance. Therefore, a measurement of GP4 = 0% in these cores was considered concordant with the consensus measurement of 1–10%. However, when the consensus of GP4 was 41–50% and 51–60%, the discordance is seen in 35% and 41.7% of the individual measurements, implying that these cores have a significant risk of being upgraded from Grade Group 2 to 3 or downgraded from Grade Group 3 to 2.

The findings of this study have several important implications for clinical practice. First, the reproducibility for GP4 quantification is moderate (κ value = 0.57), and is affected by the size of the PCa focus in the biopsy cores. Therefore, for a small focus of Grade Group 2 PCa, pathologists may consider not providing the percentage of GP4; instead, they should comment on the unreliability of the percentage of the GP4 quantification in such cases. Second, the quantification of GP4 as 40–60% has a significant error rate that may not only affect the GP4 quantification, but also the Grade Group, a caveat that both pathologists and clinicians need to be aware of. Assigning a smaller or larger percentage of GP4 imparts a greater confidence in the quantity of GP4. In contrast, GP4 between 40–60% may indicate that a tumor is borderline between Grade Groups 2 and 3. Third, the quantification reproducibility is affected by the GP4 sub-patterns with the poorly formed glands and cribriform glands having the lowest and highest reproducibility, respectively. The GP4 sub-patterns, especially the poorly formed glands, therefore need to be morphologically better defined. Methodology, i.e., the linear length vs the area of GP4, should be standardized. Education and training of pathologists may help improve the reproducibility of the GP4 quantification.

Conclusions

The reproducibility of quantifying GP4 PCa is only moderate, and significantly lower in a small focus of cancer and cancer with predominantly poorly formed glands sub-patterns. There is significant variability in quantifying GP4 when it is 40–60%. Both pathologists and clinicians should understand the limitations of GP4 quantification. More training and education for pathologists, and standardization of quantification methodology to improve the GP4 quantification is warranted.

Acknowledgments

Funding

Conflict of interest

Dr. Deng FM has been an editorial board member, and Dr. Zhou M has been an advisory board member of the *Journal of Clinical and Translational Pathology* since 2021. The other authors have no conflicts of interest related to this study.

Author contributions

Study concept and design (JL and MZ), acquisition of data (JL, ME, AA, RB, KD, FD, PL, AM, JM, SM, YT, OY, RS, and MZ), analysis and interpretation of data (JL and MZ), drafting of the manuscript (JL and MZ), critical revision of the manuscript for important intellectual content (JL, ME, AA, RB, KD, FD, PL, AM, JM, SM, YT, OY, RS, and MZ), administrative, technical, or material support (JL and MZ), and study supervision (MZ). All authors made a significant contribution to this study and approved the final manuscript.

Ethical statement

The study was performed in accordance with the ethical standards of the contributing authors' institutions and with the Declaration of Helsinki (as revised in 2013), and was ap-

proved by the corresponding author's Institutional Review Boards. The patients' consents were waived as the slides were anonymized. The study described in this publication was conducted and reported in accordance with the guidance from the Committee on Publication Ethics (COPE) and practices according to the Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals from the International Committee of Medical Journal Editors (ICMJE).

Data sharing statement

The data set used in support of the findings of this study are available from the corresponding author at mzhou3@tufts-medicalcenter.org upon request.

References

[1] Choy B, Pearce SM, Anderson BB, Shalhav AL, Zagaja G, Eggener SE, *et al*. Prognostic Significance of Percentage and Architectural Types of Contemporary Gleason Pattern 4 Prostate Cancer in Radical Prostatectomy. Am J Surg Pathol 2016;40(10):1400–1406. doi:10.1097/PAS.0000000000000691, PMID:27379821.

[2] Cole AI, Morgan TM, Spratt DE, Palapattu GS, He C, Tomlins SA, *et al*. Prognostic Value of Percent Gleason Grade 4 at Prostate Biopsy in Predicting Prostatectomy Pathology and Recurrence. J Urol 2016;196(2):405–411. doi:10.1016/j.juro.2016.01.120, PMID:26920466.

[3] Dean LW, Assel M, Sjoberg DD, Vickers AJ, Al-Ahmadie HA, Chen YB, *et al*. Clinical Usefulness of Total Length of Gleason Pattern 4 on Biopsy in Men with Grade Group 2 Prostate Cancer. J Urol 2019;201(1):77–82. doi:10.1016/j.juro.2018.07.062, PMID:30076908.

[4] Delahunt B, Steigler A, Atkinson C, Christie D, Duchesne G, Egevad L, *et al*. Percentage grade 4 tumour predicts outcome for prostate adenocarcinoma in needle biopsies from patients with advanced disease: 10-year data from the TROG 03.04 RADAR trial. Pathology 2022;54(1):49–54. doi:10.1016/j.pathol.2021.11.004, PMID:34955242.

[5] Hollemans E, Verhoef EI, Bangma CH, Rietbergen J, Helleman J, Roobol MJ, *et al*. Prostate Carcinoma Grade and Length But Not Cribriform Architecture at Positive Surgical Margins Are Predictive for Biochemical Recurrence After Radical Prostatectomy. Am J Surg Pathol 2020;44(2):191–197. doi:10.1097/PAS.0000000000001384, PMID:31592799.

[6] Kır G, Seneldir H, Gumus E. Outcomes of Gleason score 3 + 4 = 7 prostate cancer with minimal amounts (<6%) vs ≥6% of Gleason pattern 4 tissue in needle biopsy specimens. Ann Diagn Pathol 2016;20:48–51. doi:10.1016/j.anndiagpath.2015.10.013, PMID:26750655.

[7] Perera M, Assel MJ, Benfante NE, Vickers AJ, Reuter VE, Carlsson S, *et al*. Oncologic Outcomes of Total Length Gleason Pattern 4 on Biopsy in Men with Grade Group 2 Prostate Cancer. J Urol 2022;208(2):309–316. doi:10.1097/JU.0000000000002685, PMID:35363038.

[8] Perlis N, Sayyid R, Evans A, Van Der Kwast T, Toi A, Finelli A, *et al*. Limi-

[9] Sauter G, Steurer S, Clauditz TS, Krech T, Wittmer C, Lutz F, *et al*. Clinical Utility of Quantitative Gleason Grading in Prostate Biopsies and Prostatectomy Specimens. Eur Urol 2016;69(4):592–598. doi:10.1016/j.eururo.2015.10.029, PMID:26542947.

[10] Kench J, Berney D, De Marzo A, Egevad L, Kristiansen G, Litjens G, *et al*. Prostatic acinar adenocarcinoma. In: WHO Classification of Tumours Editorial Board (ed). Urinary and male genital tumours, 5th ed. Lyon (France)International Agency for Research on Cancer. 2022:203–219.

[11] Epstein JI, Amin MB, Fine SW, Algaba F, Aron M, Baydar DE, *et al*. The 2019 Genitourinary Pathology Society (GUPS) White Paper on Contemporary Grading of Prostate Cancer. Arch Pathol Lab Med 2021;145(4):461–493. doi:10.5858/arpa.2020-0015-RA, PMID:32589068.

[12] van Leenders GJLH, van der Kwast TH, Grignon DJ, Evans AJ, Kristiansen G, Kweldam CF, *et al*. The 2019 International Society of Urological Pathology (ISUP) Consensus Conference on Grading of Prostatic Carcinoma. Am J Surg Pathol 2020;44(8):e87–e99. doi:10.1097/PAS.0000000000001497, PMID:32459716.

[13] Paner GP, Srigley JR, Pettus J, Giannico GA, Sirintrapun J, Harik LR. Protocol for the Examination of Prostate Needle Biopsies From Patients With Carcinoma of the Prostate Gland: Case Level Reporting 2021. Available from: https://documents.cap.org/protocols/Prostate.Needle.Case.Bx_1.0.0.1.REL_CAPCP.pdf. Accessed December 1, 2022.

[14] Huang CC, Deng FM, Kong MX, Ren Q, Melamed J, Zhou M. Re-evaluating the concept of "dominant/index tumor nodule" in multifocal prostate cancer. Virchows Arch 2014;464(5):589–594. doi:10.1007/s00428-014-1557-y, PMID:24619626.

[15] Mohler JL, Antonarakis ES, Armstrong AJ, D'Amico AV, Davis BJ, Dorff T, *et al*. Prostate Cancer, Version 2.2019, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 2019;17(5):479–505. doi:10.6004/jnccn.2019.0023, PMID:31085757.

[16] Sadimin ET, Khani F, Diolombi M, Meliti A, Epstein JI. Interobserver Reproducibility of Percent Gleason Pattern 4 in Prostatic Adenocarcinoma on Prostate Biopsies. Am J Surg Pathol 2016;40(12):1686–1692. doi:10.1097/PAS.0000000000000714, PMID:27487740.

[17] Zhou M, Li J, Cheng L, Egevad L, Deng FM, Kunju LP, *et al*. Diagnosis of "Poorly Formed Glands" Gleason Pattern 4 Prostatic Adenocarcinoma on Needle Biopsy: An Interobserver Reproducibility Study Among Urologic Pathologists With Recommendations. Am J Surg Pathol 2015;39(10):1331–1339. doi:10.1097/PAS.0000000000000457, PMID:26099009.

[18] Dong F, Wang C, Farris AB, Wu S, Lee H, Olumi AF, *et al*. Impact on the clinical outcome of prostate cancer by the 2005 international society of urological pathology modified Gleason grading system. Am J Surg Pathol 2012;36(6):838–843. doi:10.1097/PAS.0b013e3182486faf, PMID:22592143.

[19] Kweldam CF, Wildhagen MF, Steyerberg EW, Bangma CH, van der Kwast TH, van Leenders GJ. Cribriform growth is highly predictive for postoperative metastasis and disease-specific death in Gleason score 7 prostate cancer. Mod Pathol 2015;28(3):457–464. doi:10.1038/modpathol.2014.116, PMID:25189638.

[20] Huang CC, Kong MX, Zhou M, Rosenkrantz AB, Taneja SS, Melamed J, *et al*. Gleason score 3+4=7 prostate cancer with minimal quantity of gleason pattern 4 on needle biopsy is associated with low-risk tumor in radical prostatectomy specimen. Am J Surg Pathol 2014;38(8):1096–1101. doi:10.1097/PAS.0000000000000235, PMID:24832163.

tations in Predicting Organ Confined Prostate Cancer in Patients with Gleason Pattern 4 on Biopsy: Implications for Active Surveillance. J Urol 2017;197(1):75–83. doi:10.1016/j.juro.2016.07.076, PMID:27457260.