Discrepancies Between Biopsy Gleason Score and Radical Prostatectomy Specimen Gleason Score: An Iranian Experience

Shahaboddin Dolatkhahi1, Maryam Mirtalebi1, Parnaz Daneshpajouhnejad2,3, Ahmadreza Barahimi2, Hamid Mazdak1, Mohammad Hossein Izadpanahi1, Mehrdad Mohammadi1, Diana Taheri1,7*

Purpose: Considering the importance of treatment decisions for prostate cancer (PCa) and the utility of Gleason scoring system (GS) in this field, we aimed to assess the percent of agreement and disagreement between needle biopsy (NB) Gleason score and radical prostatectomy (RP) specimen Gleason score.

Materials and Methods: In this retrospective study, consecutive patients with PCa, who underwent NB and subsequently RP were enrolled. GS of both NB and RP specimens were recorded for each patient. Patients were classified according to the GS as low-grade (≤3+3), intermediate-grade (3+4 and 4+3), and high-grade (GS: 8-10). The levels of agreement and discrepancy of NB GS was compared to its corresponding RP GS using Kappa coefficient of agreement. Over-grading and under-grading of NB GS were also determined.

Result: A total of 100 embedded RP and corresponding NB were analyzed. The rate of discrepancy for group and individual scoring of GS was 41% and 56%, respectively. The rate of under and over-grading was 34% and 7%, respectively. Kappa value for group and individual scoring was .443 (95%CI: .313 -.573) and .411 (95%CI: .291 -.531), respectively.

Conclusion: The findings of our study indicate that though the agreement between NB GS and RP GS are fair to moderate, but the feature of discrepancy, i.e. under-grading in low and intermediate grades and over-grading in high grades of NB GS, could help us in making more appropriate clinical decision especially considering other biochemical and pathological factors such as the level of PSA or peri-neural invasion.

Keywords: gleason score; grading; needle biopsy; prostate cancer; radical prostatectomy

INTRODUCTION

Prostate cancer (PCa) is considered as the most frequent cancer in men according to the annually-updated cancer statistics from the National Cancer Institute. PCa is the second leading cause of cancer-related deaths in men. It is estimated that the rate of new cases of PCa and its related deaths would be 1.7 million and 499000 by 2030 in the world. One study in Iran showed the standardized incidence of PCa from 2003 to 2009 to be 5.4, 7.24, 9.22, 9.57, 10.91, and 12.80 cases per 100,000 people, respectively. Although the RP GS represents the “true” grade of PCa, the use of NB GS has recently increased as a diagnostic and therapeutic alternative to RP. Some characteristics of GS such as ease of learning and reproducibility make this system as an appropriate diagnostic tool for prognostic and therapeutic management of PCa. However, previous studies in this field have reported a significant discrepancy between the GS of NB and RP specimens. Factors such as multifocal nature of PCa and the inherent sampling error of diagnostic NB could explain the cause of this discrepancy. Present evidence indicates that depending on the series and the periods of examination, NB GS underestimates and overestimates the RP GS in 18%-60% and 0%-25% of cases, respectively.
Moreover, it seems highly beneficial to evaluate the correlation between NB and RP GS in an Iranian population with different clinical settings and PCa causes compared to developed countries(12). Thus, considering the increasing rate of PCa and its related morbidity and mortality among Iranian males(7), and the differences in clinical settings in Iran, we aimed to evaluate the discrepancy between NB GS and RP GS in a series of Iranian patients with PCa undergoing prostate NB and subsequent RP. To our knowledge, no previous studies have been performed in Iran on this topic. The results of this study would definitely be useful in treatment decisions especially between active surveillance and curative intent therapy, as well as the utility of GS in this field(13).

**MATERIALS AND METHODS**

**Study Population**

In this retrospective study, all consecutive patients diagnosed with PCa, who underwent radical retro-pubic prostatectomy in Al-Zahra hospital, which is affiliated to Isfahan University of Medical Sciences, Isfahan, Iran, from May 2009 to May 2012, were enrolled with simple sampling method. The protocol of this study was approved by regional ethics committee of Isfahan University of Medical Sciences.

Inclusion and Exclusion Criteria:

Inclusion criteria were the availability of both preoperative NB and corresponding RP pathologic specimens. All patients underwent NB prior to RP. Patients with a history of neoadjuvant or adjuvant hormone therapy were excluded in order to eliminate bias in the histopathologic evaluation of samples and defining the Gleason score.

**Procedures and Evaluations**

All selected pathological specimens were reviewed by the same expert genitourinary pathologist to avoid inter-observer variability. The uropathologist was blinded for the patient’s identity and for the original diagnosis and outcome (including any additional NP or RP biopsy results). Clinicopathologic characteristics of selected patients including the clinical stage of PCa, pre-biopsy PSA level and presence of peri-neural invasion (PNI) were recorded. The updated GS of both NB and RP specimens were determined and recorded for each patient. Patients were classified according to the International Society of Urological Pathology (ISUP) criteria on Gleason grading of PCa(14) as low-grade (GS ≤ 3 + 3), intermediate-grade (3 + 4 and 4 + 3) and high-grade (GS ≥ 8 - 10). The levels of agreement and discrepancy for each patient NB GS were assessed aligned with their corresponding RP GS. Over-grading and under-grading were defined as NB GS higher and lower than RP GS, respectively.

**Specimens’ preparation**

The NB specimens were performed using conventional trans-rectal, ultrasound-guided (TRUS) procedure under general anesthesia with antibiotic cover by the same surgeon. After placing the patient in the left lateral position, an ultrasound probe (BK Medical Pro-Focus 2202; BK Medical, Mileparken, Denmark) was placed in the rectum to visualize the prostate. Then, 12-24 TRUS guided core biopsies were taken from the right and left peripheral zones at the surgeon’s discretion. All biopsies were stained with hematoxylin and eosin. RP specimens were formalin fixed, paraffin-embedded sections which stained with hematoxylin and eosin. The specimens were sectioned at 4 mm intervals from apex to base. GS for both NB and RP specimens was assigned based on the sum of their primary and secondary tumor patterns.

---

**Table 1. Clinicopathologic characteristics of patients with prostate carcinoma**

<table>
<thead>
<tr>
<th>Clinicopathologic Characteristic</th>
<th>Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>100 (100%)</td>
</tr>
<tr>
<td>Mean age in years ± standard deviation (SD) (range)</td>
<td>63.0 ± 6.9 (42-78)</td>
</tr>
<tr>
<td>Number of patients at a clinical stage</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>2 (2%)</td>
</tr>
<tr>
<td>T2a</td>
<td>45 (45%)</td>
</tr>
<tr>
<td>T2b</td>
<td>18 (18%)</td>
</tr>
<tr>
<td>T3</td>
<td>19 (19%)</td>
</tr>
<tr>
<td>T4</td>
<td>16 (16%)</td>
</tr>
<tr>
<td>Mean ± SD pre-biopsy PSA* in each stage group (ng/mL)</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>9.0 ± 6.3</td>
</tr>
<tr>
<td>T2a</td>
<td>12.1 ± 9.4</td>
</tr>
<tr>
<td>T2b</td>
<td>10.2 ± 6.2</td>
</tr>
<tr>
<td>T3</td>
<td>19.1 ± 16.7</td>
</tr>
<tr>
<td>T4</td>
<td>20.1 ± 11.3</td>
</tr>
<tr>
<td>Mean ± SD pre-biopsy PSA in the total population (ng/mL)</td>
<td>14.3 ± 11.5</td>
</tr>
<tr>
<td>Biopsy Gleason score (%)</td>
<td></td>
</tr>
<tr>
<td>≤ 3 + 3</td>
<td>78 (78%)</td>
</tr>
<tr>
<td>3 + 4</td>
<td>13 (13%)</td>
</tr>
<tr>
<td>4 + 3</td>
<td>5 (5%)</td>
</tr>
<tr>
<td>8 – 10</td>
<td>4 (4%)</td>
</tr>
<tr>
<td>Radical prostatectomy Gleason score (%)</td>
<td></td>
</tr>
<tr>
<td>≤ 3 + 3</td>
<td>63 (63%)</td>
</tr>
<tr>
<td>3 + 4</td>
<td>24 (24%)</td>
</tr>
<tr>
<td>4 + 3</td>
<td>9 (9%)</td>
</tr>
<tr>
<td>8 – 10</td>
<td>4 (4%)</td>
</tr>
<tr>
<td>Positive perineural Invasion (%)</td>
<td></td>
</tr>
<tr>
<td>75 (75%)</td>
<td></td>
</tr>
</tbody>
</table>

*PSA, prostatic-specific antigen. All values expressed in numbers (percentages) unless expressed otherwise.
Statistical analysis
Obtained data were analyzed using SPSS software ver.21 (SPSS Inc., Chicago, IL, U.S.A.) and Student's t-test and the chi-square test was used for comparing quantitative and qualitative variables, respectively. The concordance between NB and RP GSs was evaluated through the coefficient of the agreement, the kappa and weighed kappa statistic. 95% Confidence Intervals (CIs) are also reported. The kappa statistic is a measure of agreement between two observations and considers the chance agreement\((15)\). Kappa was calculated for each individual\((2-10)\) GSs also. Kappa agreement was calculated using GraphPad software (2015 GraphPad Prism Software, California, USA). A P-value less than 0.05 was considered statistically significant.

RESULTS
In this study, a total of 100 embedded RP and corresponding NBs were analyzed according to the updated Gleason system. Clinicopathologic characteristics of studied patients are presented in Table 1. Mean age of subjects was 63.0 ± 6.9 ranging from 42 to 78 years. Mean of PSA level was 14.3 ± 11.5 before taking biopsy. The median GS of all NBs was 6, whereas for RP it was 7.

Discrepancies between the Gleason scores of the biopsies and prostatectomy specimens are illustrated in Table 2. It is reported that among the 55 patients with Gleason score of ≤ 3 + 3 on NB, an accuracy of 61.8% for Gleason scores of ≤ 3 + 3 is seen. Of the 27 patients with Gleason scores of 3 + 4 and 4 + 3 on NB, 48.1% were graded correctly, while 48.1% were under-graded and 3.7% were over-graded. From 18 cases with high-grade tumor in NB, 66.7% were graded correctly and reminder were under-graded. Overall rate of under and over-grading was 34% and 7%, respectively. The rate of concordance, over-grading and under-grading of Gleason score from NBs compared with RPs are presented in Figure 1.

For group scoring, the number of observed agreements was 59 (59%). The reliability of biopsy for group scoring using Kappa statistics yielded a value of .374 (95% CI: .240 - .509) reflecting fair agreement beyond chance. Weighted Kappa value was .443 (95% CI: .313 - .573), which represent moderate agreement. For individual scoring, the number of observed agreements was 44 (44%). The reliability of biopsy for group scoring using Kappa statistics yielded a value of .290 (95% CI: .173 - .406), reflecting fair agreement beyond chance. Weighted Kappa value was .411 (95% CI: .291 - .531), which represent moderate agreement. We considered the weighted kappa, because most of the discrepancies are related to closer scores.

The PNI was presented in 23 (62.2%), 23 (74.2%) and 30 (93.7%) of low, moderate and high-grade PCa according to the RP GS, respectively. PNI was present in 41 (74.5%), 21 (77.8%) and 14 (77.8%) of low, moderate and high-grade PCa according to the NB GS, respectively.

Mean of PSA in low, moderate and high-grade PCa according to the RP GS was 13.1 ± 10.7, 17.3 ± 11.6 and 20.2 ± 11.1, respectively. Mean of PSA in low, moderate and high-grade PCa according to the NB GS was 9.0 ± 5.1, 11.0 ± 5.3 and 22.1 ± 10.6, respectively.

DISCUSSION
In this study, we have evaluated the discrepancies between NB and RP GS scoring in our center. We have found a 41% and 56% discrepancies between group and individual scoring of the two methods of GS scoring, respectively. Most cases of discrepancies were related to low and intermediate grade of NB GS and were mainly represented with under-grading for low and intermediate-grades. Whereas, for high-grade scoring, all of the discrepancies were represented as over-graded NB GS. Recently, in accordance with the introduction of different therapeutic alternatives to RP, the use of biopitic GS has become as an important issue in the diagnosis and management of PCa\((6)\). On the other hand, several stud-
ies have investigated the correlation between biopitic GS and RP GS and reported discrepancies between the two mentioned GS. Thus, it is suggested that in order to optimize the utility of biopitic GS in the management of PCs, evaluating the discrepancies of the two methods in each center and its related factors, could provide us with baseline information to minimize the discrepancies and improve the diagnostic utility of NB GS.

Several studies with different designs have investigated the discrepancy and over and under-grading rates of NB GS compared with RP GS. In a study in Norway among 1116 patients with PCa, reported correlation between the two grading methods was 53%, and under-grading and over-grading were 38% and 9%, respectively. Arrabal-Polo et al. in Spain also have reported similar results. In our study the rate of concordance, under-grading and over-grading was 59%, 34%, and 7%, respectively. Our results were similar to most of the reported studies in this field. Noguchi et al. have reported lower rate of concordance (36%) and higher rate of over (18%) and under (46%) grading.

The largest series of patients have been investigated by Epstein et al. by analyzing 7643 patients for the correlation between NB and RP GS. They reported a 36.3% undergrading for GS 5-6 and a 58% similar results for GS 9-10. Rajinikanth and colleagues in the USA showed that most of under-graded cases in NB GS were related to GS ≤ 6 and over grading were more in NB GS of 8-10. The results of our study were similar to this study. In our study 6 of 7 cases of over-grading, were for NB GS of 8-10 and 1 was for NB GS of 7.

Recently, Walker et al. in Canada have investigated the trend and change in discordance rates between NB and RP after implementation of active surveillance and updating of the Gleason scoring protocol by the International Society of Urologic Pathology in 2005. They indicated that the rate of discordance have decreased since 2005 in a way that the proportion of under-grading by NB has decreased for 50%. It seems that under-grading of NB GS is considered to be the most important part of reported discrepancies. Some factors including pathologic diagnosis error or experience of the pathologist, borderline cases, sampling error and reverse sampling error could explain the findings as well as the higher rate of its related under-grading.

There are evidences that increasing the number of biopsies would decrease the rate of discrepancy. In a regional experience in Australia, Ooi et al. have reported a concordance rate of 45% and under-grading rate of 46%. They concluded that the number of biopsies could improve scoring accuracy.

In a population-based study, Rapiti and colleagues have investigated the degree of concordance between NB and RP GS in 371 cases of Pca, in Geneva, Switzerland. They used Kappa statistic for evaluating the concordance. Their findings indicated that in 67% of studied population the grading was similar and in 26% was under graded by NB GS. The Kappa agreement was 0.42. They also indicated that the concordance rate would be improved by increasing the number of biopsy cores. Kappa agreement in our study was similar to the mentioned study.

Another explanation for obtained discrepancy is the time interval between biopsy and RP. Evidences suggest that increasing the time period between biopsy and RP, could increase the rate of under-grading especially for cases with lower grade tumors. It is worth to mention that in our study we used TRUS biopsy method in evaluation of PCa. One study has compared transperineal template prostate biopsy to TRUS and concluded that transperineal template prostate biopsy results in an almost 4-fold higher rate for PCa detection compared to TRUS biopsy. Another study has also suggested transperineal sector biopsy as a first-line diagnostic strategy which can be used as a safe and effective approach with high cancer detection rates compared to TRUS biopsy.

In this study, PNI was reported in 75% of all cases and the rate had increasing trend with increasing the grade of PCa. The trend was more significant by using RP GS. Mean of PSA was also higher in higher grade of PCa both in NB GS and RP GS. It seems that in cases with lower grade of PCa, clinical condition of the patients in accordance with factors such as PNI and level of PSA could help us for making more appropriate treatment approach.

The limitations of this study were small sample size of studied population, single center evaluation and retrospective design of the study. Furthermore, we have not recorded the number of biopsies in each NB and RP, core length of biopsy and prostate weight due to missing data in the medical files of the patients. Previous studies showed that the concordance between NB and RP GS scoring is higher in a larger number of biopsy specimens. Reis et al. in Brazil have reported the association between core length of biopsy as well as prostate weight (inverse relation) with RP GS up-grading. Moreover, due to the small sample size we could not investigate the role of different factors such as age, level of PSA, size of the gland, etc. in predicting the discrepancies between NB and RP GSs. The small sample size further resulted in a low number of cases with GC > 7 which may under power the findings of our study in generalizability to high-grade tumor patients.

CONCLUSIONS

The findings of our study indicated that though the agreement between NB GS and RP GS are fair to moderate, the feature of discrepancy, i.e. under-grading in low and intermediate grades and over-grading in high grades of NB GS, could help us for making more appropriate clinical decision specially if other biochemical and pathological factors such as the level of PSA or PNI are considered. This study has utmost advantage for interpretation of results in our center, and urges us to improve the biopsy techniques and pathology reports in our center to be able to rely more on the pathology readings for patient on active surveillance. It is recommend to plan future studies to determine factors which could predict discrepancies between the two methods as well as strategies to reduce it in order to provide more appropriate treatment strategies using NB GS.

ACKNOWLEDGEMENT

This study was approved and under financial support by Isfahan University of Medical Sciences as a research project.

CONFLICT OF INTEREST

The authors report no conflict of interest.
REFERENCES


25. Eroğlu M, Doluoglu OG, Sarici H, Telli O, Ozbuz GC, Bozkurt S. Does the time from biopsy to radical prostatectomy affect Gleason score upgrading in patients with clinical t1c

