International Urology and Nephrology 28 (1), pp. 73–77 (1996)

Intraobserver and Interobserver Reproducibility of WHO and Gleason Histologic Grading Systems in Prostatic Adenocarcinomas

Ş. O. ÖZDAMAR,* Ş. SARIKAYA,** L. YILDIZ,* M. K. ATILLA,** B. KANDEMIR,* S. YILDIZ**

Departments of *Pathology and **Urology, Ondokuz Mayis University, Faculty of Medicine, Samsun, Turkey

(Accepted October 16, 1995)

In this study a total of 96 patients with prostatic carcinoma were evaluated retrospectively. Sections prepared from paraffin blocks were examined and all cases were scored according to the World Health Organization (WHO) and Gleason grading systems. We investigated intraobserver and interobserver reproducibility of two grading systems in prostatic adenocarcinomas. In our study the intraobserver reproducibilities of the WHO and Gleason systems were 75.0% and 78.1%, respectively. The interobserver reproducibilities of the WHO and Gleason grading systems were 60.4% and 70.8%, respectively. While there was no difference between intraobserver and interobserver variations in the Gleason system ($p>0.05$), there was significant difference between intraobserver and interobserver variations in the WHO system ($p<0.05$).

Introduction

Carcinoma of the prostate is the most common form of cancer in men and the second leading cause of cancer-related deaths. Determining the biologic behaviour of the tumour is very important for deciding the best method of therapy. Histopathological grading and staging are the most frequently used methods for predicting the biological behaviour of prostatic carcinomas. Recognizing the relationship between histologic appearance of prostatic tumours and clinical evaluation, many investigators have been searching for histologic classifications that fulfil the requirements of prognostic accuracy, reliability and objectivity [1–5]. A limitation of most grading systems is that histologic grading is a subjective procedure. The inconsistency in histologic grading may invalidate its use in treatment decision. In this way, the reproducibility has the same significance as the predictive character of prognosis. It is necessary to find the most reliable method before histologic grades are incorporated routinely into the diagnosis of prostatic carcinoma. Although at least 30 grading systems have been proposed, few of them have been shown to be reproducible. Today, the most widely used grading systems are those of the World Health Organization (WHO) and Gleason (see Table 1) [3, 5].

WHO accepted that the grading system might be based on the degree of nuclear anaplasia (nuclear grades), and the degree of glandular differentiation

VSP, Utrecht
Akadèmiai Kiadó, Budapest

Table 1
Comparison of adenocarcinoma of prostate grading systems

WHO system	Gleason system
Grade	Pattern (Gleason grade: sum of primary and secondary patterns)
1 Well-differentiated Tumours consisting of simple small or simple large glands	1 Very well-differentiated, small, closely packed, uniform glands in essentially circumscribed masses
2 Moderately differentiated Those with complex glands, fused glands or glands in glands	2 Similar to Pattern 1, but with moderate variation in size and shape of glands and more atypia in the individual cells, still essentially circumscribed, but more loosely arranged
3 Poorly differentiated Those with few glands	3 Similar to Pattern 2, but marked irregularity in size and shape of glands, with tiny glands or individual cells invading stroma away from circumscribed masses, or solid cords and masses with easily identifiable glandular differentiation within most of them; may be papillary or cribriform
4 Undifferentiated Those with columns and cords or solid sheets	4 Large clear cells growing in a diffuse pattern resembling "hypernephroma"; may also show gland formation
	5 Very poorly differentiated tumours; usually solid masses or diffuse growth with little differentiation into glands. In cases showing more than one pattern, the second pattern was classified and recorded

(histologic grades). Nuclear anaplasia was defined as variations of nuclear size, shape, and chromatin distribution and the character of the nucleoli (grades 1–3). The two parameters, nuclear grades and histologic grades, individually or combined with the clinical stage, have been found to provide a good prognostic index [4, 5].

The Gleason system recognizes five histologic patterns of prostatic adenocarcinoma. Any given tumour may exhibit any of these patterns or a combination; therefore, in this grading system the two most predominant patterns and secondary patterns are added to arrive at a grade ranging from 2 to 10 [4, 6, 7].

In the present study, the intraobserver and interobserver reproducibilities of the WHO grading system based on the degree of glandular differentiation, and the Gleason scoring system were examined.

Material and methods

A set of histological sections of 96 prostatic carcinomas stained with haematoxylin and eosin was graded twice by one and the same pathologist on two different occasions. Another pathologist who did not know the results of

the first also examined the same sections. They used the grading systems of tumour differentiation as described by WHO and Gleason. The intraobserver and interobserver variations were computed. The data were analyzed by the test for significant difference between two proportions for paired data. A p value less than 0.05 was considered significant.

Results

Distribution of the patients by grade for each grading system is shown in Table 2. The overall intraobserver reproducibility of the WHO system was 75.0% and of the Gleason grading system 78.1%. The interobserver reproducibilities of the WHO and Gleason grading systems were 60.4% and 70.8%, respectively. There was no difference between intraobserver and interobserver variations in the Gleason system ($p>0.05$), but there was a significant difference between intraobserver and interobserver variation in the WHO system ($p<0.05$). While there was no difference in intraobserver variations ($p>0.05$), there was a significant difference in interobserver variations between the WHO and the Gleason grading systems ($p<0.05$).

Table 2
Results of Gleason and WHO grading systems intraobserver and interobserver reproducibility

	First observation (n)	Reproducibility of second observation (%)	Reproducibility of second observer (%)
Gleason			
Grade 2	0		
Grade 3	10	80.0	60.0
Grade 4	7	57.1	42.8
Grade 5	25	72.0	68.0
Grade 6	29	86.2	75.8
Grade 7	16	68.7	75.0
Grade 8	3	100.0	100.0
Grade 9	5	100.0	80.0
Grade 10	1	100.0	100.0
Total	96	78.1	70.8
WHO			
Grade 1	11	81.8	63.6
Grade 2	43	81.3	72.0
Grade 3	34	61.7	44.1
Grade 4	8	87.5	62.5
Total	96	75.0	60.4

Discussion

Although staging is important in determining the treatment of patients with prostatic adenocarcinoma, it does not always predict the biologic behaviour of the tumour. Grading has been advocated as a method of improving our ability to predict the tumour's biologic behaviour. Unfortunately, the reproducibility and clinical significance of grading have been suspect [7–9]. These shortcomings are well recognized and are emphasized by different pathologists. In addition, there are obvious limitations in the accuracy of grading based on the small amount of tissue available from needle biopsies of the prostate. Nonetheless, developing morphologic parameters to assist in determining prognosis and treatment for adenocarcinoma of the prostate are important goals for both pathologists and clinicians [3–5].

We preferred the WHO grading system based on the degree of glandular differentiation, because of its similarity to the Broders grading system which is commonly used in other carcinomas. When the WHO grading system was used, the vast majority of cases (80.2%) in our series were either grade 2 or grade 3, similar to the results of Brawn et al. [3]. We found an intraobserver reproducibility rate of 75.0%, and an interobserver reproducibility rate of 60.4%. The intraobserver reproducibility rate was better than the interobserver reproducibility rate ($p<0.05$). The major conflicts were in grades 3 and 4. The experience of the pathologist may influence observer variations and bias may be minimized as much as possible by choosing pathologist who were involved in uropathology for the same period. On the other hand, definition of pathological features may change from one pathologist to another [5–7, 8]. These problems raise questions as to the value of a grading system in close cooperation between different pathologists.

When the Gleason grading system was used, the majority of cases (30.2%) in our series were grade 6, similar to Gleason's study [2]. We found intraobserver reproducibility rate to be 78.1% which is better than the reproducibility found by other workers [9–11]. Employing his own system, Gleason estimated his intraobserver reproducibility rate to be 80% which is similar to our result. Although our interobserver reproducibility rate was lower than the intraobserver reproducibility rate, the difference was statistically not significant ($p>0.05$). In interobserver evaluations there was a significant difference between the WHO and the Gleason systems ($p<0.05$). It seems that agreement among pathologists in the Gleason grading system is greater than expected by chance, and reproducibility rates in interobserver evaluation are higher in the Gleason system as compared to the WHO system.

In the absence of a gold standard, the value of a histological indicator can be measured by examining the inter- and intraobserver reproducibility rates. Those items with poor agreement cannot be regarded as reliable indicators and cannot be used for making a decision. A histological grading system showing higher intra- and interobserver reproducibility rates seems ideal. In addition, if there is no statistically significant difference between inter- and intraobserver

reproducibility rates it may be safely used by different pathologists. In this study, we found higher intra- and interobserver reproducibility rates both in the WHO and the Gleason grading systems, but only a comparison of intra- and interobserver reproducibility rates was not different in the Gleason grading system. For this reason we suggested that the Gleason grading system was better than the WHO grading system based on the degree of glandular differentiation. Further studies including patients' survival, clinical stage, neuroendocrine differentiation, proliferative index (PCNA/Cyclin reactivity), immunohistochemical analysis (PSA and PSAP) and comparison of these parameters with the histologic grading systems will be helpful in understanding the roles of these grading systems in prostatic adenocarcinomas.

Acknowledgements

We would like to thank Esin Özyilkan, M. D. and Naziye Gürkan, M. S. for their help.

References

1. Boring, C. C., Squires, T. S., Tong, T., Montgomery, S.: Cancer statistics. *CA Cancer J. Clin., 44,* 7 (1994).
2. Gleason, D. F.: Histologic grading of prostate cancer: A perspective. *Hum. Pathol., 23,* 273 (1992).
3. Brawn, P. N., Ayala, A. G., Von Eschenbach, A. C., Hussey, D. H., Johnson, D. E.: Histologic grading study of prostate adenocarcinoma: The development of a new system and comparison with other methods – A preliminary study. *Cancer, 49,* 525 (1982).
4. Cintra, M. L., Billis, A.: Histologic grading of prostatic adenocarcinoma: Intraobserver reproducibility of the Mostofi, Gleason and Böcking grading systems. *Int. Urol. Nephrol., 23,* 449 (1991).
5. Mostofi, F., Sesterhenn, I. A., Davis, C. J.: Pathologist's view of prostatic carcinoma. *Cancer Suppl., 71,* 906 (1993).
6. Gleason, D. F.: Histologic grading of prostatic carcinoma. In: Bostwick, D. G. (ed.): Pathology of the Prostate. Churchill Livingstone, New York 1990.
7. Sogani, P., Israel, A., Lieberman, P., Lesser, M. L., Whitmore, W. F. Jr.: Gleason grading of prostate cancer: A predictor of survival. *Urology, 25,* 223 (1985).
8. Schroeder, F. H., Hop, W. C., Bloom, J. H., Mostofi, F. K.: Grading of prostatic carcinoma: Multivariant analysis of prognostic parameters. *Prostate, 7,* 13 (1985).
9. Gallee, M. P. W., Ten Kate, F. J. W., Mulder, G.: Histologic grading of prostate carcinoma in prostatectomy specimens: A comparison of prognostic accuracy of five grading systems. *Br. J. Urol., 65,* 368 (1990).
10. Murphy, G. P., Whitmore, W. F. Jr.: A report of the workshops on the current status of the histologic grading of prostate cancer. *Cancer, 44,* 1490 (1979).
11. Svanholm, H., Mygind, H.: Prostatic carcinoma reproducibility of histologic grading. *Acta Pathol. Microbiol. Scand., 93,* 67 (1985).