Acta path. microbiol. immunol. scand. Sect. A, 93: 67–71, 1985.

# PROSTATIC CARCINOMA
# REPRODUCIBILITY OF HISTOLOGIC GRADING

HANS SVANHOLM[1] and HENRIK MYGIND[2]

Departments of Pathology, Aalborg and Slagelse Hospital, Denmark

[1] Department of Pathology, Sönderborg Hospital, 6400 Sönderborg, Denmark.
[2] Department of Pathology, Odense Hospital, 5000 Odense, Denmark.

Svanholm, H. & Mygind, H.: Prostatic carcinoma; reproducibility of histologic grading. Acta path. microbiol. immunol. scand. Sect. A, 93: 67–71, 1985.

The inter- and intraobserver variations of the histologic grading systems of prostatic carcinoma proposed by Gleason and by Böcking et al. were tested by the authors. Both grading systems have previously been shown to have good correlation with prognosis. After studying 91 cases of prostatic carcinomas, the interobserver agreements of the Gleason pattern score and the Böcking combined grade were found to be 36% and 69% respectively. Both observers regraded 31 randomly selected prostatic carcinomas. With the Gleason pattern score intraobserver agreements of 65% and 42% were found, whereas the intraobserver agreements of the Böcking combined grade were 90% and 71% respectively. Since the percentages of inter- and intraobserver agreements were higher with the system of Böcking et al., this system is recommended prior to the system of Gleason as a means of evaluating the prognosis of patients with prostatic carcinoma.

Key words: Prostatic carcinoma; histologic grading.

Hans Svanholm, Department of Pathology, Sönderborg, Hospital, 6400 Sönderborg, Denmark.

Accepted as submitted 19.ix.84

The importance of histopathologic grading in malignant disease was recognized by *Broders* (2). Several grading systems of prostatic carcinoma subsequently have been described, but the prognostic accuracy of many of these systems have been inconsistent (3, 11). The two most widely used systems of histologic grading of prostatic carcinomas presently employed are those proposed by *Gleason* (7) and by *WHO* (12). The Gleason classification has achieved wide recognition for its prognostic value and relative simplicity. The Gleason grades have been shown to correlate well with survival (7), incidence of pelvic lymph node metastases (8), and to be reproducible by other observers (1, 8). The system proposed by WHO does not have the same established clinical significance as the grading system of Gleason. However, a new histologic grading system, using diagnostic categories in accordance with the histologic classification published by WHO, has been published by *Böcking et al.* (3), and this system has shown good correlation with survival and presence of metastases. Furthermore, this study found the mean interobserver reproducibility to be 91%, and the intraobserver reproducibility to be 87,5%.

Before the beginning of a study of prostatic carcinoma, it was desirable to find a histopathological grading system which is reproducible. The tests chosen were the inter- and intraobserver variations of the systems proposed by Gleason and by Böcking et al., because these two systems have shown equally good correlation to the clinical course of prostatic carcinoma.

TABLE 1. *Grading Systems of Gleason and Bocking et Al. for Prostatic Carcinoma*

| | Gleason | Bocking et al. |
|---|---|---|
| Number of histologic growth patterns. | 1-5 | 1-4 |
| Nuclear anaplasia. | | 1-3 |
| Pattern (total) score. | Sum of the two quantitatively predominating patterns = 2-10. | Sum of growth pattern and nuclear anaplasia from the qualitatively poorest grade of differentiation = 2-7. |
| Combined grade | | I = Total score of 2-3.<br>II = Total score of 4-5.<br>III = Total score of 6-7. |
| Combined grading-staging. | Sum of pattern score and clinical stage = 3-15. | |

## MATERIALS AND METHODS

The material of the study consists of paraffin embedded tissue blocks from 100 consecutive cases of prostatic carcinoma from the files of the Department of Pathology, Aalborg Hospital. The tissue available for 96 cases were from transurethral resections, while four cases were from punch biopsies. All paraffin blocks were recut with an average of five blocks per case (range 1-10). The sections were stained with Hematoxylin and Eosin and were evaluated separately by the authors.

The prostatic carcinomas were histologically graded in accordance to the descriptions given by *Gleason* (7) and by *Bocking et al.* (3) (Table 1). The histologic grading system of Gleason considers only the glandular differentiation, and distinguishes five growth patterns 1-5. The Gleason pattern score appears either by addition of the numbers of the two most predominant growth patterns (primary and secondary), or in cases with only one growth pattern, by multiplying the number with two. The histologic grading system of Bocking et al. distinguishes four histologic growth patterns 1-4, and three grades of nuclear anaplasia 1-3. The combined grade of prostatic carcinoma results from the addition of the rating number of the histologic growth pattern to the rating number of nuclear anaplasia. The sum of rating numbers 2-3 corresponds to grade I, 4-5 to grade II, and 6-7 to grade III.

In five cases both observers found the tissue sections unsatisfactory for histologic grading, and in four cases one or both observers disagreed on the carcinoma diagnosis. These nine cases were excluded. Subsequently, the interobserver study was based on 91 cases of prostatic carcinoma.

To study the intraobserver variation, 31 randomly selected cases from the material were examined a second time separately by both observers after suitable time interval, to avoid possible recognition of previous graded samples.

The inter- and intraobserver variations were described using percentages of exact agreement and agreement within one score unit. Moreover, the chance-corrected weighted Kappa statistics ($K_w$) ± one standard error ($S_w$) of the weighted Kappa was employed (5, 6). It was decided to weigh the disagreements found, so that the line of agreement was given the weight zero, while the weights for disagreements were the difference in grades from the agreement diagonal treated as the exponent to the second power (difference in grade = 2, weight = $2^2$;.....; difference in grade = 9, weight $9^2$). The method to test the significance of the difference between two independent $K_w$'s has been previous described (4).

The weighted Kappa statistic ($K_w$) gives a chance-corrected proportion of weighted agreement. $K_w$ yields negative values when there is less observed agreement than is expected by chance, zero when observed agreement can solely be explained by chance, and $K_w = 1$ when there is complete agreement. The weights assigned are an integral part of how agreement is defined and must be decided early in the study (5). Because the number of categories in a grading system bear influence on the Kappa statistics, the method is not suitable to compare values of $K_w$ obtained from different grading systems with different numbers of categories.

## RESULTS

The interobserver variation of the grades assigned to 91 cases of prostatic carcinomas is illustrated in Tables 2 and 3. For the Gleason pattern scores (Table 2) exact agreement between the observers was seen in 36%, whereas agreement within one score unit was achieved in 69%. The chance-cor-

TABLE 2. *Distribution of the Gleason Pattern Scores Made Independently by Two Observers, of 91 Prostatic Carcinomas*

| OBSERVER A \ OBSERVER B | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| 10 | | | | | | | | | 4 |
| 9 | | | | 1 | 2 | 3 | 2 | 6 | 2 |
| 8 | | | | 1 | 8 | 7 | 6 | 13 | 2 |
| 7 | | | 1 | 2 | 2 | 6 | | | |
| 6 | | 1 | 2 | 1 | 2 | | | | |
| 5 | 2 | 3 | 3 | 8 | | | | | |
| 4 | | | 1 | | | | | | |
| 3 | | | | | | | | | |
| 2 | | | | | | | | | |

TABLE 3. *Distribution of the Bocking Combined Grades, Made Independently by Two Observers, of 91 Prostatic Carcinomas*

| OBSERVER A \ OBSERVER B | I | II | III |
|---|---|---|---|
| III | | 20 | 43 |
| II | 4 | 15 | 2 |
| I | 5 | 2 | |

TABLE 4A. *Distribution of the Gleason Pattern Scores, Made by Observer A, Grading and Re-Grading 31 Prostatic Carcinomas*

| 2. GRADING \ 1. GRADING | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| 10 | | | | | | | | | |
| 9 | | | | | | | 1 | 2 | |
| 8 | | | | | | | 10 | 3 | |
| 7 | | | | 1 | | 3 | 2 | 1 | |
| 6 | | | | | 1 | | 1 | | |
| 5 | | | | 4 | | 1 | | | |
| 4 | | | | 1 | | | | | |
| 3 | | | | | | | | | |
| 2 | | | | | | | | | |

TABLE 4B. *Distribution of the Gleason Pattern Scores, Made by Observer B, Grading and Re-Grading 31 Prostatic Carcinomas*

| 2. GRADING \ 1. GRADING | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| 10 | | | | | | | | | |
| 9 | | | | | | | | 2 | 1 |
| 8 | | | | | 1 | 1 | 1 | 3 | |
| 7 | | | | 1 | 3 | 5 | 3 | | |
| 6 | | | 1 | 2 | 1 | | | | |
| 5 | | | | 3 | 1 | | | | |
| 4 | | | | | | | | | |
| 3 | | 1 | | 1 | | | | | |
| 2 | | | | | | | | | |

rected agreement coefficient $K_a$ was 0,700 ± 0,044. The agreement within Gleason primary pattern was better than the agreement within the secondary pattern, the values of $K_a$ being 0,642 ± 0,051 and 0,234 ± 0,089 respectively, and the difference betweeen the two values of $K_a$ was significant ($P < 0,001$, data not given). Using the system of Bocking et al. exact agreement of combined grades was found in 69%, whereas the agreement within one score unit was 100% (Table 3). The chance-corrected agreement coefficient $K_a$ was 0.649 ± 0,066. For the Bocking growth pattern and nuclear anaplasia, exact agreements were seen in 62% and 65% respectively, the values of $K_a$ being 0,613 ± 0,070 (growth pattern) and 0,527 ± 0,078 (nuclear anaplasia, data not given). As seen from Tables 2 and 3, the grades assigned by observer A of both grading systems have a systematic tendency to higher values compared to the grades given by observer B.

The variation of the grades from the intraobserver study is illustrated in Tables 4 and 5. For the Gleason pattern score (Tables 4A and 4B) exact agreements of the first and second gradings were found in 65% (observer A) and 42% (observer B), and both observers had agreement within one score unit of 87%. The chance-corrected agreement coefficients $K_a$ were 0.801 ± 0,071 (observer A) and 0,809 ± 0,057 (observer B). As regards the Bocking combined grade (Tables 5A and 5B), exact agreements of the first and second gradings were seen in 90% (observer A) and 71% (observer B). Values of $K_a$ were 0,867 ± 0,084 (observer A) and 0,488 ± 0,135 (observer B), and this difference is significant ($0,01 < P < 0,05$).

TABLE 5A. *Distribution of the Bocking Combined Grades, Made by Observer A, Grading and Re-Grading 31 Prostatic Carcinomas*

| 2. GRADING | 1. GRADING I | 1. GRADING II | 1. GRADING III |
|---|---|---|---|
| III | | | 20 |
| II | | 6 | 3 |
| I | 2 | | |

TABLE 5B. *Distribution of the Bocking Combined Grades, Made by Observer B, Grading and Re-Grading 31 Prostatic Carcinomas*

| 2. GRADING | 1. GRADING I | 1. GRADING II | 1. GRADING III |
|---|---|---|---|
| III | | 2 | 11 |
| II | 1 | 11 | 6 |
| I | | | |

## DISCUSSION

Both *Gleason* (7) and *Böcking et al.* (3) have shown good correlation between their histologic grading systems and the prognosis. While the reproducibility of the system proposed by Gleason has been tested by others (1, 8), this is to the best of our knowledge not the case for the grading system proposed by Böcking et al.

The histologic grading system of Gleason demands an evaluation of the tumour volume of the different growth patterns, and does not take smaller areas with lower degree of differentiation in account. However, in a recent work utilizing the Gleason system, *McGowan et al.* (9) found in disagreement with Gleason, that the qualitatively highest malignancy identified (the highest histologic grade) is most important in respect of prognosis. The histologic grading system of Böcking et al. uses the same histologic and cytologic categories as the WHO classification, but differes from the WHO classification by assigning the qualitatively highest malignancy found to the tumour as a whole (3, 12). Furthermore, the system of Böcking et al. differs from that of *Mostofi* (10) by not having a fixed combination of the histologic growth patterns and the three degrees of nuclear anaplasia, and the system of Mostofi designates a tumor grade on the basis of predominant features (13).

Using his own system Gleason had intraobserver reproducibility of 80% (13), which is better than the reproducibility of 65% and 42% found in the present study. *Bain et al.* (1) using Gleasons grading system, found interobserver agreement within $\pm$ 1 of a consensus score from 74,1% to 93,1%, and the chance-corrected agreement with the consensus score $\pm$ 1 ranged from 0,605 to 0,836 (unweighted Kappa statistics). Using the system of Gleason, we found interobserver agreement within one score unit of 69%. Because the weighted Kappa statistics were used, $K_w$ cannot be compared to the value of K given by Bain et al.. However, from our study the calculated unweighted Kappa statistic using agreement $\pm$ 1 was K = 0,467 and S = 0,077 as regards the interobserver agreement of the Gleason pattern score. This value of K is lower than found by Bain et al., but the difference may be explained by the fact that Bain et al. have used a consensus score as the standard for judging the scores assigned by the seven participating pathologists, and furthermore that study has graded 58 slides from 58 cases of prostatic carcinoma. The gradings of this study are the results of the evaluation of an average of five sections from each tumor.

With the system of Bocking et al. the present study found intraobserver agreements of 90% and 71% which is close to the reproducibility of 87,5% found by Böcking (3). However, our interobserver agreement of 69% is lower than the interobserver agreement of 91% found by Böcking et al. (3).

The Gleason system does not utilize the cellular or nuclear features of tumours. In a report of the workshops on the current status of the histologic grading of prostatic cancer, it is recommmended that nuclear and cytologic characteristics be considered in prospective studies to further the discriminative capabilities of the Gleason system (13). The system of Böcking et al. is based on the evaluation of both the histologic growth pattern and the nuclear anaplasia. In this respect the system of Böcking et al. may hold more information concerning the tumour biology than the Gleason system which takes into account only the histologic growth pattern.

This study finds that both the systems of Gleason and of Böcking et al. are reproducible, but the percentages of the inter- and intraobserver agreements are higher for the Böcking combined grade than for the Gleason pattern score. Though statistic comparison between the two systems is difficult, the results speak in favour of the Böcking

system. Therefore, the authors recommend this system as a means of evaluating the prognosis of patients with prostatic carcinoma.

---

We greatly thank Cand.Scient. *T. Arnfred*, Department of Clinical Chemistry, Aalborg Hospital, for the computerprogram of the weighted Kappa statistics, and for his statistical assistance.

## REFERENCES

1. *Bain, G. O., Koch, M. & Hanson, J.:* Feasibility of grading prostatic carcinomas. Arch. Pathol. Lab. Med. *106*: 265-267, 1982.
2. *Broders, A. C.:* Grading and practical application. Arch. Pathol. *2*: 376-381, 1926.
3. *Bocking, A., Kiehn, J. & Heinzel-Wach, M.:* Combined histologic grading of prostatic carcinoma. Cancer *50*: 288-294, 1982.
4. *Cohen, J.:* A coefficient of agreement for nominal scales. Educational and Psychological Measurement *20*: 37-46, 1960.
5. *Cohen, J.:* Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin *70*: 213-220, 1968.
6. *Fleiss, J. L., Cohen, J. & Everitt, B. S.:* Large sample standard errors of kappa and weighted kappa. Psychological Bulletin *72*: 323-327, 1969.
7. *Gleason, D. F.:* Histologic grading and clinical staging of prostatic carcinoma. In: *Tannenbaum, M.* (Ed.): Urologic pathology: The prostate. Lea & Febiger, Philadelphia 1977, pp. 171-197.
8. *Kramer, S. A., Spahr, J., Brendler, C. B., Glenn, J. F. & Paulson, D. F.:* Experience with Gleason's histopathologic grading in prostatic cancer. Jour. Urol. *124*: 223-225, 1980.
9. *McCowan, D. G., Bain, G. O. & Hanson, J.:* Evaluation of histological grading (Gleason) in carcinoma of the prostate: Adverse influence of highest grade. The Prostate *4*: 111-118, 1983.
10. *Mostofi, F. K.:* Grading of prostatic carcinoma. Cancer Chemother. Rep. *59*: 111-117, 1975.
11. *Mostofi, F. K.:* Problems of grading carcinoma of prostate. Seminars in Oncology *3*: 161-169, 1976.
12. *Mostofi, F. K., Sesterhenn, I. & Sobin, L. H.:* Histological typing of prostate tumours. International histological classification of tumours, No. 22, World Health Organization, Geneva, 1980.
13. *Murphy, G. P. & Whitmore, W. F.:* A report of the workshops on the current status of the histologic grading of prostate cancer. Cancer *44*: 1490-1494, 1979.