Paper

Interobserver reproducibility of Gleason histological grading of prostatic carcinoma: a study done in a teaching hospital setting in Colombo, Sri Lanka

R D A S Francisco, M C Joseph

Abstract

Introduction: Gleason grading is one of the most important prognostic factors in prostatic carcinoma.

Objective: To assess the interobserver reproducibility of Gleason grading of prostatic cancer among three pathologists in the local setting.

Method: Fifty haematoxylin and eosin stained slides of previously diagnosed prostatic adenocarcinoma were retrieved and graded by using Gleason grading by three pathologists. These were re-graded by the same three pathologists after a time interval of 4 weeks after studying literature highlighting more descriptive criteria on the above grading system. The interobserver reproducibility was analysed using generalised kappa values.

Results: The generalised kappa value for interobserver variation among the three pathologists in the first instance was 0.306 indicating a fair level of agreement and 0.488 in the second instance, indicating a moderate level of agreement. The lowest level of interobserver agreement was observed in the Gleason score 2-4 and 7 categories.

Conclusion: The results emphasises the importance of continuous education of pathologists with regard to the Gleason grading system to reduce the inter observer variation in the grading of prostatic adenocarcnoma

Key words Interobserver variation, Gleason grading, Prostate carcinoma

Introduction

Prostate cancer is the third common cancer in men in the world with 543,000 new cases being diagnosed every year (1). According to latest cancer registry data prostate cancer is the fourth commonest type of cancer among Sri Lankan men over the age of 65 years (2).

Department of Histopathology, Sri Jayewardenepura General Hospital, Sri Lanka.

Correspondence: R.D.A.S. Francisco, email:sewwandifrancisco@yahoo.com

Gleason grading of prostate cancer is an important histopathologic predictor of prognosis which is routinely incorporated into the histopathology report (3). Donald F Gleason in 1966 created this unique grading system for prostate cancer based solely on the architectural pattern of the tumour. At that time there were no other screening methods for prostate cancer except for digital rectal examination.

With the new advances in medicine related to prostate pathology which includes serum prostate specific antigen (PSA) assay, new methods of sampling the prostate and the incorporation of immunohistochemical methods in histological assessment of prostate, there was a need to revise the original criteria introduced by Gleason. Gleason grading system was modified in 1974 and 1977 and the 2005 International Society of Urological Pathology (ISUP) consensus conference was an attempt by a group of urological pathologists to achieve a consensus regarding controversial areas of Gleason grading (3).

Gleason grading is done by low power assessment of the architectural pattern of the tumour with subsequent more detailed assessment under the high-power of the microscope. The most predominant architectural pattern and the next predominant (secondary) architectural pattern are identified and grades or scores are assigned for each pattern. Both scores are added up to make the final score. If there is a minor component which is of a higher grade than the most predominant (dominant) and next most predominant (codominant) components, the former component is also mentioned in the report as a significant tertiary pattern (3).

In Sri Lanka, where there are no specialised uropathologists dealing exclusively in the area of uropathology, Gleason grading of prostatic adenocarcinoma is the responsibility of the general histopathologists. For any grading system to be effective it is very important to ensure an accepted level of inter observer variation. Therefore, a study of the interobserver variation of the Gleason grading in a Sri Lankan setting provides an important insight into the effectiveness of this important prognostic marker which determines patient management.

Literature review did not reveal any published data regarding similar studies done in the Sri Lankan setting. The authors also attempted to assess whether allowing individual pathologists to study more descriptive criteria based on published literature with incorporation of reference images for each grade would improve the existing level of inter observer reproducibility of the grading system as this could then be utilized as one of the continuing medical development tools in our country.

Method

This was carried out in the department of histopathology at the Sri Jayewardenepura General Hospital. 50 cases of previously diagnosed adenocarcinoma of the prostate were retrieved from the files of the department of histopathology at the Sri Jayewardenepura General Hospital. The three pathologists who assessed the slides were involved in independent practice as consultant pathologists and all of them were serving as consultant pathologists in a teaching hospital setting. Olympus microscopes with a high power field diameter of 0.65 mm were used by all three pathologists.

The 50 representative slides from each case were reviewed by these three pathologists independently, as they would do in their routine practice. A diagram illustrating Gleason grading was available for reference. Final Gleason score for each case was recorded for each slide after obtaining numerical values for the most predominant and the next predominant components and adding these values together to make up the final score.

The individual pathologists were allowed to study more descriptive criteria for the grading system based on published literature with incorporation of reference images for each grade. Subsequently, the same set of 50 slides was again circulated amongst each pathologist and a final Gleason score for each slide was recorded again as previously described.

The slides were anonymised so that the initial identity of the slide was not known to the pathologist who performing the grading but was only known to the researcher. A period of 4 weeks was allowed between the two occasions to eliminate the recall bias.

Gleason scores by each pathologist were grouped into four categories for the convenience of statistical analysis.

Category 1. Gleason Score 2-4

Category 2. Gleason Score 5-6

Category 3. Gleason Score 7

Category 4. Gleason Score 8-10

Reproducibility of the Gleason score on both occasions was obtained by using generalized kappa values. Kappa values were calculated based on the equations presented by Fleiss (4).

Interobserver agreement was quantified by the kappa (k) statistics (Table 1).

The following scale was used in the interpretation of kappa values.

Table 1. Landis and Koch scale for interpreting kappa values (5)

Kappa value	Agreement
<0.2	Slight
0.2-0.4	Fair
0.4-0.6	Moderate
0.6-0.8	Substantial
0.8-1	Almost perfect

Results

The generalized kappa value for interobserver agreement among the three pathologists on the first instance was 0.306 indicating a fair level of agreement (Table 2).

The generalized kappa value for interobserver agreement among the three pathologists on the second instance was 0.488 indicating moderate agreement (Table 3).

Table 2. Category wise kappa values in the first instance are as follows.

Category	Kappa value
1 (Gleason score 2-4)	0.379
2 (Gleason score 4-6)	0.336
3 (Gleason score 7)	0.280
4 (Gleason 8-10)	0.307

$p < 0.05$

Table 3. Category wise kappa values on the second instance (after the introduction of descriptive literature) were as follows.

Category	Kappa value
1 (Gleason score 2-4)	0.493
2 (Gleason score 4-6)	0.658
3 (Gleason score 7)	0.303
4 (Gleason 8-10)	0.545

$p < 0.05$

The p values on both instances were less than 0.05 indicating interobserver agreement exceeding chance on both occasions.

Discussion and conclusions

The generalized kappa values for inter observer agreement in the first round which indicated a fair level of agreement with a kappa value of 0.306, increased up to the level of moderate agreement with a kappa value of 0.488 after pathologists were allowed to study the literature containing descriptive criteria for each grade with the aid of reference images.

A study involving several centres in UK regarding the interobserver variation among 10 pathologists (6) also gave results similar to our study indicating a moderate level of agreement.

Kappa values for each category was also seen to have increased in the second round

indicating improvement of the level of agreement after studying the literature.

On both occasions, the lowest level of agreement was seen in the grading of Gleason score 2-4 and 7 tumours. This is comparable to a UK based study which also showed least agreement in these two categories of Gleason scores (6).

Another study conducted in Cardiff, UK comparing interobserver agreement of Gleason score interpretation among uropathologists and general pathologists showed that inter observer agreement among general pathologists is significantly lower than that among uropathologists (7).

The Cardiff study also showed an improvement of the level of agreement after each rater is allowed to go through a web based tutorial (7,8). This is comparable to the improvement in interobserver agreement seen in our study after the raters were allowed to study the descriptive literature and reference images.

Another study in USA has revealed that level of agreement among pathologists is significantly lower if they diagnose fewer cases per year or have not learned about the Gleason score in a course or consensus meeting (9,10).

Presence of lower interobserver agreement in Gleason score 2-4 and 7 calls for further refinement of criteria in the diagnosis of these two categories.

The results also suggest that the consensus on the criteria of grading prostatic carcinoma among the pathologists is important to increase the level of interobserver reproducibility.

Therefore, it is prudent to carry out continuous medical education sessions(CME), workshops and other external and internal quality assurance processes which will help to update the knowledge on the prostatic grading system.

Thus, it becomes apparent that CME sessions, workshops and other quality assurance programmes should include updates and practice sessions on Gleason grading of prostatic adenocarcinomas for consultant histopathologists in Sri Lanka.

The refinement of the grading criteria of those grades which showed lower levels of inter observer agreement, is also important to increase the effectiveness of the grading system.

Peer reviewing and seeking second opinions also has an important place in increasing the interobserver reproducibility of a grading system which will ultimately benefit the management of patient management.

Acknowledgement

The authours wish to thank Dr. Prabhath Ambawatta (Consultant Histopathologist), Dr.

Sonali Rodrigo (Consultant Histopathologist), Dr. Supun Wijesingha, Dr. Chamila Nagodavithana and staff members of the Department of Histopathology, Sri Jayewardenepura General hospital for the guidance and support given to conduct this study.

References

1. WHO. *World cancer report* , Lyon. IARC Press; 2003

2. Cancer Incidence Data Sri Lanka.2001-2005. National Cancer Control Programme publications 2009.

3. Recent trends in Gleason grading of prostate cancer, Prognosis reproducibility and reporting., Egevad L, *Urologe A*, 2007 Jan; 46 (1):59-62

4. Jonathan I. Epstein, William Allsbrook Jr, Mahul B. Amin, Lars L. Egevad The 2005 International conference of Urological Pathology (ISUP) Consensus Conference on Gleason grading of Prostate carcinoma. *American Journal of Surgical Pathology,* Vol 29 ; September 2005: 1228-1241.

5. Statistical Methods for Rates and Proportions. 1981: 229-232 http://www.ccitonline.org/jking/homepage/kappal.xlx

6. Landis JR, Koch GG, The measurement of observer agreement for categorical data. *Biometrics.* 1977 ,Vol. 33:159-174

7. Melia J, et al, A UK based investigation of inter and intra observer reproducibility of Gleason grading of prostatic biopsies. *Histopathology* 2006; 48 : 644-654.

8. Griffiths DFR et al. A study of Gleason score interpretation in different groups of UK pathologists; techniques for improving reproducibility. *Histopathology* Vol 48; 2006 : 655-662

9. Kronz JD, Silberman MA, Allsbrook WC, Epstain JI. A web based tutorial improves practicing pathologists' Gleason grading of images of prostate carcinoma specimens obtained by biopsy: A validation of new medical education paradigm. *Cancer* 89; 2000, 1818-1823.

10. Allsbrook WC Jr, Mangold KA, Jhonson MH et al. Interobserver reproducibility of Gleason grading of prostatic carcinoma; general pathologists; *Human Pathology,* 32; 2001: 81-88