Integrative Cancer Science and Therapeutics

Research Article ISSN: 2056-4546

An audit of inter-observer variability in Gleason grading of prostate cancer biopsies: The experience of central pathology review in the North West of England

Emile N Salmo*

Department of Histopathology, Bolton NHS Foundation Trust, UK

Abstract

Gleason score, which is an important histological parameter in determining therapeutic decisions for prostate cancer, has a high level of interobserver variability amongst general and specialist urological pathologists. A total of 96 prostate biopsies were reviewed and complete agreement was seen in 72% of cases following central pathology review. Amongst cases which demonstrated Gleason score change, 75% of cases these were downgraded and 25% were upgraded. Most of the discrepancy involved pattern 3 and 4, however, in our series, there was evidence of over interpretation of grade 3 and 4 and this might indicate the influence of the International Society of Urological Pathology (ISUP) modification of Gleason scoring which was adopted in 2005.

Introduction

Prostate cancer is the second most frequently diagnosed cancer and the sixth leading cause of cancer death in males, accounting for 14% of the total new cancer cases and 6% of the total cancer deaths in males in 2008 globally [1]. In the histological reporting of prostate cancer, Gleason scoring system is an important prognostic parameter for therapeutic decision and in the overall management of prostate cancer patients [2] and it has emerged as a strong predictor of recurrence and prediction of organ-confined disease [3]. In 2005, the ISUP introduced modifications of Gleason scoring system [4] one of which include assigning any cribriform pattern to grade 4 and this has been shown to decrease the under grading observed in biopsies when compared to the prostatectomy specimens. However, as with any other grading system, Gleason score has a level of subjectivity which depends on the pathologist experience. The aim of the audit is to determine the level of inter-observer variability in the reporting of Gleason scoring of prostate adenocarcinoma before and after a central review process.

Materials and methods

In a unique practice established in the North West of England cancer sector which involve Wigan Royal Infirmary, Bolton NHS Foundation Trust and Salford Royal Foundation Trust, all prostate cancer biopsies are subjected to central review by three pathologists with special interest in urological pathology. These biopsies are reviewed before discussion takes place at the weekly Sector Urology Multi-disciplinary Team Meeting (MDTM) where all urological cancer cases are discussed and treatment decisions are made at that time.

At the central review meeting, all prostate cancer biopsies are reviewed and a consensus opinion is reached and in case of discrepancy from the original report, a supplementary report is issued after discussion at the Sector Urology MDTM.

A total of 96 prostate biopsy cases were reviewed during a 6 months period (March 2014-September 2014), all of which are related to cases originated at Bolton NHS Foundation Trust. The original diagnosis was compared with the consensus diagnosis established at the urology peer review meeting. The usual clinical practice by the clinicians in our hospital is to send 6 biopsies from each side (left and right) in 2 separate pots labeled right and left.

Kappa value was used to measure the degree of agreement between the diagnoses and classified as follows [5]: A value of 0-0.2 indicates slight agreement; 0.21-0.4 fair agreement; 0.41-0.6 moderate agreement; 0.61-0.8 substantial agreement and >0.81 as almost perfect agreement.

Results

Amongst the 96 cases reviewed, total agreement was present in 69 cases (72%) with a Kappa value of 0.666 (substantial; good agreement) (Table 1) and 91(95%) cases were within +/- 1 score. In 3 cases, the overall Gleason score remained unchanged, however, there was a grading change between grade 3 and 4.

In the 24 cases which demonstrated score change, 18 (75%) cases were downgraded and 6 (25%) were upgraded. Amongst the 6 cases that were upgraded, 5 cases were upgraded from grade 3 to 4 and in one cases a small focus of grade 5 was missed (grade changed from 4+3 to 4+5).

Amongst the cases that were downgraded, 15 were downgraded from grade 4 to 3 and 3 cases from 5 to 4.

Correspondence to: Salmo EN, Department of Histopathology, Bolton NHS Foundation Trust, Minerva Road, Bolton, BL4 0JR, United Kingdom, Tel: 00 44 7403001111; **E-mail:** emilsalmo@hotmail.com

Key words: *prostate, cancer, gleason score, inter-observer variability*

Received: January 28, 2015; **Accepted:** February 26, 2015; **Published:** March 02, 2015

Table 1. Kappa value by Gleason score.

		Original score					
	Gleason score	**6**	**7**	**8**	**9**	**10**	**Total**
Review opinion	6	23	4	3	0	0	30
	7	2	20	9	0	0	31
	8	1	1	16	2	0	20
	9	0	1	1	10	0	12
	10	0	0	0	0	3	3
	Total	26	26	29	12	3	96

Kappa value=0.666
95% confidence interval 0.549 to 0.782

Table 2: Kappa value by risk group.

		Original score			
	Gleason score	**6**	**7**	**8-10**	**Total**
Review opinion	6	23	4	3	30
	7	2	20	9	31
	8-10	1	2	32	35
	Total	26	26	44	96

Kappa value=0.669
95% confidence interval 0.545 to 0.793

When Gleason scores were grouped into risk categories (6 (low risk), 7 (3+4 and 4+3; intermediate risk), 8-10 (high risk)) [6] agreement was observed in 75% of cases with a mean Kappa value of 0.669 (good/substantial agreement) (Table 2).

The grading of prognostic groups were established according to the Gleason score and grouped as follows: Out of the 24 cases that demonstrated change in Gleason score, 21 (87%) cases showed major discrepancy which might have affected therapeutic decisions. A major Gleason score discrepancy was defined as a change to a different risk category (6, 7, 8-10) [6].

Discussion

Tissue biopsy is the gold standard in the diagnosis of prostate cancer, determining prognostic parameters which affect therapeutic decisions [7,8]. Gleason score has long been known as one of the most important prognostic factors for the outcome of treatment in prostate cancer and even determines the treatment of choice for the tumour [9,10], thus, a high degree of precision in its reporting is a crucial issue. Reporting of Gleason score has been shown to suffer a high degree of inter-observer variation amongst pathologists and it is seen to be higher amongst general pathologists than specialist urological pathologists [11].

Inter-observer agreement in Gleason score differs between studies as some literature reporting up to 71% exact agreement [12] with others reported a range between 9.9%-36% [11,13,14]. In our review, total agreement was demonstrated in 72% of cases which indicate a high degree of concordance between the original reports and the review opinion.

Previous literature consistently showed that training reduces the level of disagreement in Gleason scoring of prostate cancer and reduce inter-observer variability [14,15]. In recent literatures, it has been shown that the degree of inter-observer agreement depends on the experience of the pathologist and the training provided. This agreement, in general, has been seen to be high amongst urological pathologists than in general pathologists [16]. In a recent study, a kappa value of 0.7 was reported reflecting the experience of pathologists involved in the study [2]. In a study by Mulay et al, an agreement 0.36-0.64 was reported but the value increased after a simple web-based training, thus indicating the value of training in reducing the level of disagreement in the interpretation of prostate biopsies [14,15].

Mandatory second review also brings changes to the cancer grade on which major therapeutic decisions are based [6]. In the current review, 27 (28%) cases suffered a change in Gleason grading and in the majority of cases the changes involved migrating to a different risk group which might have affected treatment decisions. In the contemporary era any Gleason score change that places the patient in a different risk stratification category is considered a major change. The 3 categories used at most institutions are score 6, 7 and 8 –10 [6].

Part of the cause of reproducibility problems when diagnosing Gleason pattern 4 may be that not all pathologists are familiar with the changes recently brought to Gleason grading after the International Society of Urological Pathology consensus conference in 2005 [4]. In our series, the majority of the changes have been tumour downgrading which reflect the over-diagnosis of Gleason grade 4 by our pathologists.

A few studies have highlighted the importance of central pathology review of prostate biopsies before prostatectomy or further therapy and most of these have shown its value because it can result in a significantly different report that may affect therapy [6,17]. The majority of the current literature suggests that central pathology review should become routine practice [3,17] as this process has been shown to facilitate optimal prostate cancer management and improve quality of life to patients [3].

In the current review, it would have been beneficial to analyze how the changes in Gleason score would have affected treatment decision in these cases and this might be incorporated in future audits.

In Conclusion, this analysis of prostate cancer biopsies demonstrated that there was a high degree of concordance between the original Gleason scores and the consensus scores derived from the central review. This indicates that the reporting pathologists had a high degree of awareness of the Gleason score modification which was devised by the ISUP in 2005. The degree of interobserver variation between pathologists in the interpretation of Gleason score in prostate biopsies can be reduced by regular training and feedback following central review process.

Acknowledgement

I would like to thank Dr S Verma and Dr K Chilman for their help in the central review process.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, et al. (2011) Global cancer statistics. *CA Cancer J Clin* 61: 69-90. [Crossref]

2. Lucia MS, Bostwick DG, Somerville MC, Fowler IL, Rittmaster RS. (2013) Comparison of classic and international society of urological pathology 2005. Modified gleason grading using needle biopsies from the reduction by dutasteride of prostate cancer eevents (REDUCE) trial. *Arch Pathol Lab Med* 137: 1740-1746. [Crossref]

3. D'Souza N, Loblaw DA, Mamedov A, Sugar L, Holden L (2012) Prostate cancer pathology audits: is central pathology review still warranted? *Can J Urol* 19: 6256-6260. [Crossref]

4. Epstein JI, Allsbrook WC Jr, Amin MB, Egevad LL; ISUP Grading Committee (2005) The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *Am J Surg Pathol* 29: 1228-1242. [Crossref]

5. Brennan P, Silman A (1992) Statistical methods for assessing observer variability in clinical measures. *BMJ* 304: 1491-1494. [Crossref]

6. Brimo F, Schultz L, Epstein JI (2010) The value of mandatory second opinion pathology review of prostate needle biopsy interpretation before radical prostatectomy. *J Urol* 184: 126-130. [Crossref]

7. Bostwick DG (1994) Gleason grading of prostatic needle biopsies. Correlation with grade in 316 matched prostatectomies. *Am J Surg Pathol* 18: 796-803. [Crossref]

8. Vira MA, Tomaszewski JE, Hwang WT, D'Amico AV, Whittington R, et al. (2005) Impact of the percentage of positive biopsy cores on the further stratification of primary grade 3 and grade 4 Gleason score 7 tumors in radical prostatectomy patients. *Urology* 66: 1015-1019. [Crossref]

9. Gleason DF (1992) Histologic grading of prostate cancer: a perspective. *Hum Pathol* 23: 273-279. [Crossref]

10. Allsbrook WC Jr, Mangold KA, Johnson MH, Lane RB, Lane CG, et al. (2001) Interobserver reproducibility of Gleason grading of prostatic carcinoma: general pathologist. *Hum Pathol* 32: 81-88. [Crossref]

11. Abdollahi A, Meysamie A, Sheikhbahaei S, Ahmadi A, Moradi-Tabriz H, et al. (2012) Inter/intra-observer reproducibility of Gleason scoring in prostate adenocarcinoma in Iranian pathologists. *Urol J* 9: 486-490. [Crossref]

12. Ozdamar SO, Sarikaya S, Yildiz L, Atilla MK, Kandemir B, et al. (1996) Intraobserver and interobserver reproducibility of WHO and Gleason histologic grading systems in prostatic adenocarcinomas. *Int Urol Nephrol* 28: 73-77. [Crossref]

13. McLean M, Srigley J, Banerjee D, Warde P, Hao Y (1997) Interobserver variation in prostate cancer Gleason scoring: are there implications for the design of clinical trials and treatment strategies? *Clin Oncol (R Coll Radiol)* 9: 222-225. [Crossref]

14. Abdollahi A, Sheikhbahaei S, Meysamie A, Bakhshandeh M, Hosseinzadeh H (2014) Inter-observer reproducibility before and after web-based education in the Gleason grading of the prostate adenocarcinoma among the Iranian pathologists. *Acta Med Iran* 52: 370-374. [Crossref]

15. Mulay K, Swain M, Jaiman S, Gowrishankar S (2008) Gleason scoring of prostatic carcinoma: impact of a web-based tutorial on inter- and intra-observer variability. *Indian J Pathol Microbiol* 51: 22-25. [Crossref]

16. Iczkowski KA, Lucia MS (2011) Current perspectives on Gleason grading of prostate cancer. *Curr Urol Rep* 12: 216-222. [Crossref]

17. Montironi R, Lopez-Beltran A, Cheng L, Montorsi F, Scarpelli M (2013) Central prostate pathology review: should it be mandatory? *Eur Urol* 64: 199-201. [Crossref]