Inter-Observer Reproducibility before and after Web-Based Education in the Gleason Grading of the Prostate Adenocarcinoma among the Iranian Pathologists

Alireza Abdollahi[1*], Sara Sheikhbahaei[1], Alipasha Meysamie[2], Mohammadreza Bakhshandeh[1], and Hasan Hosseinzadeh[1]

[1] *Department of Pathology, Faculty of Medicine, Tehran University of Medical Sciences, Tehran, Iran*

[2] *Department of Community Medicine, Faculty of Medicine, Tehran University of Medical Sciences, Tehran, Iran*

Received: 11 Feb. 2013; Accepted: 7 Jun. 2013

Abstract- This study was aimed at determining intra and inter-observer concordance rates in the Gleason scoring of prostatic adenocarcinoma, before and after a web-based educational course. In this self-controlled study, 150 tissue samples of prostatic adenocarcinoma are re-examined to be scored according to the Gleason scoring system. Then all pathologists attend a free web-based course. Afterwards, the same 150 samples [with different codes compared to the previous ones] are distributed differently among the pathologists to be assigned Gleason scores. After gathering the data, the concordance rate in the first and second reports of pathologists is determined. In the pre web-education, the mean kappa value of Interobserver agreement was 0.25 [fair agreement]. Post web-education significantly improved with the mean kappa value of 0.52 [moderate agreement]. Using weighted kappa values, significant improvement was observed in inter-observer agreement in higher scores of Gleason grade; Score 10 was achieved for the mean kappa value in post web-education was 0.68 [substantial agreement] compared to 0.25 (fair agreement) in pre web-education. Web-based training courses are attractive to pathologists as they do not need to spend much time and money. Therefore, such training courses are strongly recommended for significant pathological issues including the grading of the prostate adenocarcinoma. Through web-based education, pathologists can exchange views and contribute to the rise in the level of reproducibility. Such programs need to be included in post-graduation programs.

Acta Medica Iranica, 2014;52(5):370-374.

Keywords: Inter-observer reproducibility; Gleason grading; Prostate adenocarcinoma; Pathologist; Web-based education

Introduction

According to ACS (American Cancer Society), prostate cancer is the most prevalent type of cancer and the second leading cause of cancer death in men (1).

Gleason scoring, the most commonly used grading system for prostate adenocarcinoma with worldwide acceptance, is an important indicator of its' pathologic stage and outcome (1). Regarding the close correlation between the Gleason score of a tumor and the clinical course of the disease (2-4) and its important role in predicting patient prognosis, guiding patient counseling and making treatment decisions, we can realize the importance of reproducibility in Gleason scoring among pathologists (5-7).

In the previous studies, the concordance rate of reports varied from 0.16 to 0.84(8-10).

Factors such as the experience of pathologist have influenced this index of concordance.

However, there are reasons to believe that we can increase the concordance rate, regardless of the experience of pathologist or other factors. In this study, regarding the necessity of reproducible and concordant reports among different pathologists, we want to assess the effect of a web-based educational course on the concordance rate among the Gleason score reports of pathologists from selected hospitals of Tehran University of Medical Sciences. A simple search indicated t that no similar study has already been conducted in Iran or neighboring states and therefore the results of this study

Corresponding Author: A. Abdollahi
Department of Pathology, Faculty of Medicine, Tehran University of Medical Sciences, Tehran, Iran
Tel: +98 9121220588, Fax: +98 21 88269844, E-mail address: Abdollahi_a@tums.ac.ir

could serve as a yardstick for regional states.

Materials and Methods

In this self-controlled study, 150 tissue samples of prostatic adenocarcinoma obtained through Needle Biopsy from the archives of three teaching hospitals of Tehran University of Medical Sciences laboratories. The samples were collected to be re-examined and graded anew. This study was carried out between September 2009 and December 2010 and the biopsy slides belong to that period of time. This method is the first of the kind to identify and separate sample tissues for which a diagnosis of adenocarcinoma is made. After the preparation of slides and staining with hematoxylin and eosin, they are sent to 3 randomly selected pathologists in the three hospitals for Gleason scoring. (According to previous studies and based on the views of the statistician involved in the study, pathologists from different hospitals affiliated with Tehran's University of Medical Sciences were picked. These hospitals have large urology wards).

The slides that cannot be scored and the slides from patients previously treated with radiotherapy or anti-androgenic drugs and those samples that contain less than 5 malignant acini are eliminated from the study. After the samples are chosen, a code is given to each sample, and the Gleason scores given by each pathologist will be recorded in a data sheet. Then all pathologists attend a free web-based course. This educational course is accessible at www.pathology.jhu.edu/prostate, and the course materials and photos will be the same for all pathologists. Then, the same 150 samples (with different code compared previous ones) are distributed differently among the same pathologists to be assigned Gleason scores. After gathering the data, the concordance rate of the reports of pathologists in the first and second reports with the consensus score, and also the concordance of the first and the second scores reported by each pathologist are calculated. Statistical analysis will include calculation of K for each pathologist in their first and second reports, and also between the pathologists and the consensus score. Kappa value of 0-0.20 indicates slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.81 and higher will be regarded as almost perfect agreement. At the end, the data will be analyzed, both descriptively and analytically, using the SPSS V.15 and STATA 8 software.

Results

150 tissue samples of prostatic adenocarcinoma were scored by 3 pathologists before and after attending a free web-based educational course.

Inter observer Agreement

Percentages of agreement & Kappa values of all possible pair combination of 3 pathologists' scores in pre web-education and post web-education were shown in Table 1.

Table 1. Kappa values & Percentages of agreement with 95% confidence intervals of all possible pair combination of 3 pathologist's Grading Scores after and before educational course

	P2 Pre	P3 Pre	P1 Post	P2 Posts	P3 Post	
P1 Pre	33% [25.5%-40.5%]	52% [44.0%-60.0%]	85% [79.3%-90.7%]	52% [44.0%-60.0%]	45% [37.0%-53.0%]	Agreement %
	0.14 [0.08-0.20]	0.39 [0.31-0.47]	0.8 [0.74-0.86]	0.39 [0.31-0.47]	0.3 [0.23-0.37]	Kappa
P2 Pre		41% [33.1%-48.9%]	35% [27.4%-42.6%]	57% [49.1%-64.9%]	54% [46.0%-62.0%]	Agreement %
		0.24 [0.17-0.31]	0.16 [0.10-0.22]	0.44 [0.36-0.52]	0.41 [0.33-0.49]	Kappa
P3 Pre			51% [43.0%-59.0%]	43% [35.1%-50.9%]	41% [33.1%-48.9%]	Agreement %
			0.37 [0.29-0.45]	0.27 [0.20-0.34]	0.25 [0.18-0.32]	Kappa
P1 Post				63% [55.3%-70.7%]	52% [44.0%-60.0%]	Agreement %
				0.52 [0.44-0.60]	0.39 [0.31-0.47]	Kappa
P2 Post					73% [65.9%-80.1%]	Agreement %
					0.65 [0.57-0.73]	Kappa

[P Value < 0.001], P1[Pathologist 1], P2[Pathologist 2], P3[Pathologist 3], Pre[Pre web-education], Post[Post web-education]

In the pre web-education, the mean kappa value The level of agreement in different Gleason scores compared for the 3 observer assigned to the 150 consensus cases was calculated both in pre web-education and post web-education (Table 2) was 0.25 [fair agreement]. Post web-education agreement significantly improved with the mean kappa value of 0.52 [moderate agreement].

Using weighted kappa values, there was a significant improvement in inter-observer agreement in higher scores of Gleason grade, as in score 10 the mean kappa value in Post web-education was 0.68 [substantial agreement].

Intra Observer Agreement

By comparing pre web-education and post web-education scores of each pathologist, the intra-observer agreement was calculated, Low Kappa implies greater change in pathologists' scoring, whereas high Kappa indicate little change in scoring after the educational course. Kappa values of Intra-observer reproducibility is shown in Figure 1.

Table 2. Interobserver reproducibility of Gleason's Grading system for prostatic carcinoma after and before educational course

Gleason Score	Pre Test		Post Test	
	Kappa	P value	Kappa	P value
6	0.36	<0.001	0.51	<0.001
7	0.06	0.12 *	0.41	<0.001
8	0.28	<0.001	0.48	<0.001
9	0.34	<0.001	0.59	<0.001
10	0.25	<0.001	0.68	<0.001
Combined	0.25	<0.001	0.52	<0.001

*: Not Significant

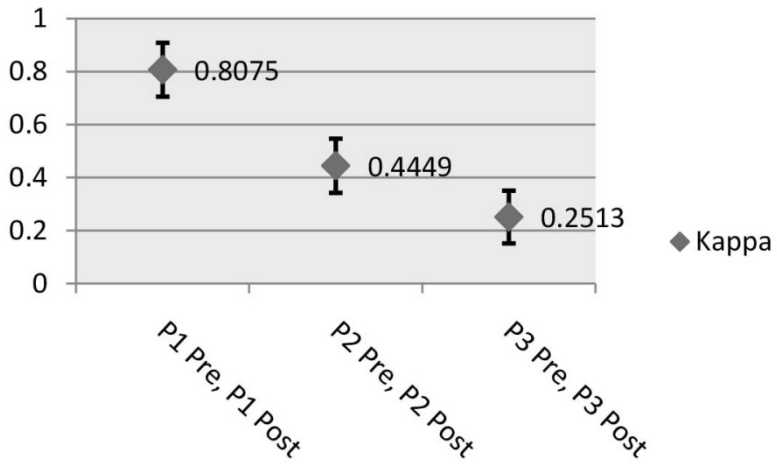

Figure 1. Intra Observer agreement

P1[Pathologist1],P2[Pathologist2], P3[Pathologist3], Pre[Pre web-education],Post[Post web-education]

Discussion

Carcinoma of the prostate is the most common internal malignancy among men in the United States and is responsible for 10% of cancer deaths in this population (2-4). 207 Each year in New York State more than 11,000 men are diagnosed with prostatic cancer and more than 2,300 die from it (4). Prostate cancer is the leading cause of new cancer in men and is second only to lung cancer as a leading cause of cancer-related deaths in men (4).

The prostate cancer is very lethal and pathologists make a great contribution to treating this disease by its grading. Any treatment [surgery, chemotherapy, radiotherapy and hormone therapy] is chosen based on the tumor's type, grading and staging. That is why an accurate grading is crucial. An option to boost the reliability of grading for a tumor is the existence of

acceptable reproducibility among various pathologists. In other studies, various levels of reproducibility ranging from insufficient to good have been reported.

In different studies conducted on this issue, the reproducibility among pathologists has been examined. A study conducted by Oz Damar et al on the prostatic needle biopsy from 96 patients with prostatic cancer reached an acceptable inter-observer variation for the Gleason-style grading (11).

Moreover, the study carried out by Allsbrook et al examined 46 cases of prostatic cancer for grading; ten pathologists were involved. The reproducibility stood at an acceptable level (12).

The study by Mulay *et al.,* on 40 prostatic cancer patients reached an inter-observer reproducibility between 0.36 and 0.64. After a web-based training course for the pathologists contributing to this project, the indicator soared significantly. This study is well indicative of the significance of regular training of pathologists (13).

The difference in the reproducibility levels in these studies could result from the training programs for pathologists, their experiences, their participation in follow-up courses, dealing with difficult cases and so on. In this study, the inter-observer reproducibility was measured from 150 biopsy slides of prostate adenocarcinoma. The reproducibility was not acceptable for the reasons mentioned earlier. But it rose tangibly after pathologists underwent a web-based training course before they examined the same slides for Gleason grading. The more the grades, the more was the reproducibility. The others studies reached similar results and a web-based program led to a big jump in the inter-observer reproducibility. Different studies have highlighted the value of the web-based education (14-16).

In view of the present levels of inter-observer reproducibility, the developing countries had better brace for training uropathologists to the higher levels.

Web-based training courses are attractive to pathologists as they will not need to spend much time and money. Therefore, such training courses are strongly recommended for significant pathological issues including the grading of the prostate adenocarcinoma. Through web-based education, the pathologists can exchange views and contribute to the rise in the level of reproducibility. Such programs need to be included in post-graduation programs.

The restrictions we encountered in this study included the small number of participating pathologists and the small size of samples. A multi-centre and more extensive study with participation of more pathologists with different levels of experience is recommended for the future.

Acknowledgment

The authors sincerely offer their gratitude to the vice-chancellor for research affairs at Tehran University of medical sciences for his financial support in favour of this research project and all contributors to this study.

References

1. Billis A, Guimaraes MS, Freitas LL, et al. The impact of the 2005 international society of urological pathology consensus conference on standard Gleason grading of prostatic carcinoma in needle biopsies. J Urol 2008;180(2):548-52
2. Bostwick, DG. Gleason grading of prostatic needle biopsies. Correlation with grade in 316 matched prostatectomies. Am J Surg Pathol 1994;18(8):796-803.
3. Vira MA, Tomaszewski JE, Hwang WT, et al. 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 2005;66(5):1015-9.
4. Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. 1974. J Urol 2002;167(2 Pt 2):953-8.
5. Oyama T, Allsbrook WC Jr, Kurokawa K, et al. A comparison of interobserver reproducibility of Gleason grading of prostatic carcinoma in Japan and the United States. Arch Pathol Lab Med 2005;129(8):1004-10.
6. Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. J Urol 1974;111(1):58-64.
7. Damico AV, Renshaw AA, Arsenault L, et al. Clinical predictors of upgrading to Gleason grade 4 or 5 disease at radical prostatectomy: potential implications for patient selection for radiation and androgen suppression therapy. Int J Radiat Oncol Biol Phys 1999;45(4):841-6.
8. Melia J, Moseley R, Ball RY, et al. A UK-based investigation of inter- and intra-observer reproducibility of Gleason grading of prostatic biopsies. Histopathology 2006;48(6):644-54.
9. Lotan TL, Epstein JI. Clinical implications of changing definitions within the Gleason grading system. Nat Rev Urol 2010;7(3):136-42.
10. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ

1992;304(6840);1491-4.
11. Oz Damar so, Sarrikaya S, Yildiz L, et al. Intraobserver and Interobserver Reproducibility of WHO and Gleason Histologic Grading System in Prostatic Adenocarcinoma. Int Urol Nephrol1996;28(1):73-7.
12. Allsbrook WC Jr, Mangold KA, Johnson MH, et al. Interobserver reproducibility of Gleason grading of prostatic carcinoma: general pathologists. Hum Pathol 2001;32(1):81-8.
13. Mulay K, Swain S, Jaimn S, et al. Gleason Scoring of Prostatic Carcinoma: impact of a web-based tutorial on inter-and intra-observer variability. Indian J Pathol Microbiol 2008;51(1):22-5.
14. Erickson RA, Chang A, Johnson CE, et al. Lecture versus web tutorial for pharmacy students learning of MDI technique. Ann Pharmacother 2003;37(4):500-5.
15. Barnes K, Itzkowitz S, Brown K. Teaching clinical management skills for genetic testing of hereditary nonpolyposis. Genet Med 2003;5(1):43-8.
16. Li Y, Brodlie K. Phillips N. Web-based VR training simulator for percutaneous rhizotomy. Stud Heath Technol Inform 2000;70(1):175-81.