Concordance between transrectal ultrasound guided biopsy results and radical prostatectomy final pathology: Are we getting better at predicting final pathology?

Richard Walker, BScH Cand.;^{*} Uri Lindner, MD;^{*} Alyssa Louis, MD;^{*} Robin Kalnin, MSc,^s Marguerite Ennis, PhD, PStat,[±] Michael Nesbitt, MD;^{*} Theodorus H. van der Kwast, MD, FRCPC,[†] Antonio Finelli, MD, FRCSC;^{*} Neil E. Fleshner, MD, FRCSC,^{*} Alexandre R. Zlotta, MD, FRCSC;^{*} Michael A.S. Jewett, MD, FRCSC;^{*} Robert Hamilton, MD, FRCSC;^{*} Girish Kulkarni, MD, FRCSC;^{*} John Trachtenberg, MD, FACS, FRCSC^{*}

^{*}Division of Urology, Department of Surgical Oncology, University Health Network, Toronto, ON; [†]Department of Pathology, University Health Network, Toronto, ON; [§]Meridian Software, Toronto, ON; ^{*}Applied Statistician, Markham, ON

Cite as: *Can Urol Assoc J* 2014;8(1-2):47-52. http://dx.doi.org/10.5489/cuaj.751 Published online February 10, 2014.

Abstract

Introduction: Inaccuracy in biopsy Gleason scoring poses a risk to men who may then receive inappropriate treatment. We assess whether there was a change in discordance rates between biopsy and radical prostatectomy at our institution in recent years, while considering the implementation of active surveillance and the shift in biopsy scores caused by the 2005 International Society of Urologic Pathology update to the Gleason scoring protocol.

Methods: We reviewed patients who underwent radical prostatectomy at our institution between May 2004 and April 2011. We analyzed clinical and pathological correlates of upgrading in 3 subgroups: Gleason sum (GS) 6/6, GS6/7 and GS7/7, where the sum preceding the dash was determined from biopsy and the subsequent sum was determined from the radical prostatectomy specimen. We applied the log-rank test and Cox model to a Kaplan Meier analysis of biochemical recurrence in the subgroups, and also mapped GS6/7 discordance over time.

Results: In total, 1717 patients met our inclusion criteria. The 3 subgroups had significantly different mean prostate-specific antigen, patient age, tumour volume, margin status, pathologic stage, prostate weight, transrectal ultrasound volume and rate of progression (p < 0.05). We noted a multiphasic trend with a fall in discordance after 2005. However, there was no sustained trend over the study period taken as a whole (p = 0.06).

Conclusions: Although no sustained trend was observed, the falling discordance after 2005 may reflect the accommodation to the Gleason scoring update, while the gradual adoption of active surveillance may have led to the otherwise increasing trends. However, our observations may also be spurious biopsy sampling errors.

Introduction

Transrectal ultrasound (TRUS)-guided sextant or systematic prostate biopsy is the clinical standard for definitive diagnosis of prostate cancer. The Gleason sum (GS) derived from such biopsy specimens is crucial for appropriate treatment selection and is an important predictor of outcome. However, biopsy may miss or under-grade prostate cancer: false-negative rates of the first sextant biopsy approach 17%, upgrading on subsequent biopsy approaches 20% and under-grading rates are 30% to 50%.¹⁻⁵

Of specific interest is the misclassification of GS7 tumours as GS6, as this is the threshold for the D'Amico risk classification systems between low and intermediate risk.⁶ Men who have low-risk, "insignificant" tumours, as coined by Epstein, must have no more than GS6 on biopsy.⁷ These men may be eligible for conservative management strategies, such as active surveillance, and new focal therapy strategies. Men with intermediate-risk tumours are more commonly treated with total-gland therapies with curative intent, such as radical prostatectomy (RP) and radiotherapy.^{6,8} Recently, the Prostate Cancer Intervention versus Observation Trial (PIVOT) reported no significant difference in all-cause or prostate-cancer-specific mortality between men randomized to active surveillance versus RP, suggesting that active surveillance is appropriate for men with low-risk tumours.⁹ This finding emphasizes the need to accurately stratify men by risk, as active surveillance continues to gain support.

Previously, Pinthus and colleagues investigated the discordance of GS between prostate biopsy and subsequent RP surgical specimens in 698 men treated at our institution between 1989 and 2004.⁵ This report was prior to the widespread implementation of active surveillance, and before the 2005 update to the GS scoring system by the International Society of Urologic Pathology (ISUP).^{10,11} This update dictated that all, except for the most uniform cribriform Gleason pattern 3, should be re-classified as 4, and tertiary higher-grade patterns at biopsy should replace the secondary pattern in the conventional grading system. A longitudinal comparison of the new system to the conventional system found that the proportion of GS6 decreased from 48% to 22%, while the proportion of GS7 increased from 26% to 68%. This stage migration was accompanied by increased concordance between biopsy and specimen GS from 58% to 72%.¹²

In this study, we document the discordance between prostate biopsy and final pathology in an older cohort and a contemporary cohort. By doing so, we will elucidate how recent events – specifically, the advent of active surveillance and the update in the Gleason grading system – have influenced the discordance rates over time. The outcome of such cases will also be considered.

Methods

Patient selection and follow-up

After obtaining Institutional Review Board approval (12-5106-CE), we abstracted records from our institutional Prostate Centre Database of patients who were treated with RP at The University Health Network (Princess Margaret Cancer Centre, Toronto General Hospital) from May 2004 to April 2011. Prior to RP, all patients underwent systematic TRUS-guided prostate biopsy. The inclusion criteria were prostate biopsy performed within 2 years of RP and existing pathology report for biopsy and RP specimen. We abstracted GS, age at surgery, PSA closest to biopsy, number of biopsy cores, cancer volume, margin status, pathologic stage, and adjuvant and salvage treatments. Due to the confounding effect of androgen deprivation on both GS and PSA recurrence, we excluded records if neoadjuvant or adjuvant treatment was documented without PSA failure or recurrence.

Statistical analysis

Following Pinthus and colleagues' study at our centre in 2005, we repeated the statistical tests identifying differences in characteristics of patients in 3 patient groups: GS6/6, GS6/7, and GS7, where the sum preceding the dash belongs to the first positive biopsy pathology and the subsequent sum to that of the RP. One-way ANOVA was used for continuous variables and the chi-square test was used for categorical variables.

To investigate changes over time in the proportion of GS6/7, we calculated the proportion for each year of study. These points were turned into a smooth curve using the rolling average method, by repeatedly moving the 1-year

window forward 30 days and recalculating the proportion. A permutation test was used to test whether the pattern obtained was consistent with random fluctuations around the mean or whether it represented a true difference.¹³ The test took into account the smoothing process and the fact that the windows overlapped.

The Kaplan-Meier method was used to estimate biochemical recurrence-free survival in the 3 patient groups of interest (GS6/6, GS6/7 and GS7/7). The log-rank test and a Cox proportional hazards model were applied to the curves. Biochemical recurrence was defined as detectable PSA on 2 consecutive visits after RP, with at least 3 months between each measurement. Patients without progression were censored at their most recent follow-up. Time to biochemical recurrence was measured from RP to the second consecutive recurred PSA. All analyses were performed using S-PLUS, version 6.2 (Insightful, Seattle, WA).

Results

.....

Preoperative clinical and postoperative pathological characteristics

Our database search yielded 1743 records. We excluded a total of 26 records: 17 due to adjuvant treatment and 9 due to the fact that there was more than 2 years between biopsy and RP. In the end, there were 1717 records for analysis. Of these, 1422 records were divided into 3 groups: (1) 356 men were Gleason 6 on both biopsy and specimen pathology (GS6/6), (2) 356 were upgraded from Gleason 6 to Gleason 7 (GS6/7) and (3) 710 were Gleason 7 on both (GS7/7). We tallied the preoperative and postoperative characteristics of these 3 groups (Table 1). Of the preoperative characteristics, age, PSA at diagnosis and TRUS volume were statistically different among groups, while there was no difference in number of biopsy cores and body mass index. Mean ages in the GS6/6, GS6/7 and GS7/7 groups were 58.99, 60.42 and 61.47, respectively (p < 0.0001). All postoperative characteristics were statistically different between groups (p < 0.05). In our study, 295 patients did not fit into a GS6/6, GS6/7, or GS7/7 group. The may have been GS7/8 or other. These 295 were excluded from comparisons between the 3 groups, but were included in overall discordance frequency statistics.

Frequency of discordance between prostate biopsy and postoperative pathology Gleason score

Of the 728 tumours graded GS6 at prostate biopsy, 356 (48.9%) matched the pathological findings while 361 (49.6%) were upgraded and 11 (1.5%) were downgraded. Tumours graded GS7 at biopsy (n = 838) had higher concordance; 710 (84.7%) matched at RP, 35 (4.1%) upgraded to GS8 or 9, and 93 (11.1%) downgraded to GS6 or 5 (Table 2).

Table 1. Preoperative and postoperative characteristics amongst GS6/6, GS6/7, and GS7/7						
Programatica	GS 6/6	GS 6/7	GS 7/7	n velve		
Preoperative	n=356	n=356	n=710	<i>p</i> value		
Age						
Mean	58.99	60.42	61.47	<0.0001		
Median (range)	59.5 (38–76)	61 (40–75)	62 (41–76)			
n	356	356	710			
PSA closest to biopsy						
(continuous)						
Mean	5.42	7.34	8.10	<0.0001		
Median (range)	4.83	5.91	6.27			
	(0.05–47.8)	(0.72–65.89)	(0.05–56.16)			
n	323	320	675			
No. biopsy cores						
Mean	11.59	11.53	11.63	0.6393		
Median (range)	11 (3–28)	11 (5–35)	11 (4–23)			
n	294	272	601			
BMI						
Mean	27.26	27.52	27.66	0.3402		
Median (range)	27.13	27.17	27.22			
moulan (rango)	(19.36–41.21)	(19.59–36.98)	(17.36–42.1)			
n	271	252	581			
TRUS volume						
Mean	43	39.07	39.29	0.0050		
Median (range)	40 (16–141)	33 (14–131)	34 (15.6–184)			
n	219	191	424			
Postonerative	GS 6/6	GS 6/7	GS 7/7	n value		
	n=356	n=356	n=710	pvalue		
% Cancer volume						
Mean	5.03	9.54	10.93	<0.0001		
Median (range)	3 (1–30)	6 (1–75)	7 (1–95)			
n	339	346	692			
Margins				<0.0001		
Negative	319 (90.1%)	283 (79.7%)	562 (79.6%)			
Positive	35 (9.9%)	72 (20.3%)	144 (20.4%)			
Total	354 (100%)	355 (100%)	706 (100%)			
Pathologic stage				<0.0001		
pT2	319 (90.4%)	249 (71.1%)	403 (57.7%)			
pT3/4	34 (9.6%)	101 (28.9%)	295 (42.3%)			
Total	353 (100%)	350 (100%)	698 (100%)			
Gleason pattern at RP				<0.0001		
3+4	_	329 (92.4%)	529 (74.5%)			
4+3	_	27 (7.6%)	181 (25.5%)			
Total	_	356 (100%)	710 (100%)			
Prostate weight						
Mean	49.86	47.56	46.59	0.0424		
Median (range)	46 (5–170)	42 (19–155)	41 (20–198)			
n	345	347	698			
GS: Gleason score; PSA: prostate-specific and	tigen; BMI: body mass index; TRUS: transre	ctal ultrasound; RP: radical prostatectom	у.			

Proportion of Gleason score discordance over time

In the first year of study, a progressive increase in the proportion of GS6/7 tumours was observed (Fig. 1). At its high-

est (May 2005 to end of April 2006) the GS6/7 proportion was 0.612. Over the next 12 months, the mean proportion had fallen to 0.387. From about May 2008, a progressive increase was again observed. In the study's final year, the

Biopsy GS 6		Biopsy GS 7		
Difference*	Frequency (%)	Difference* Frequency		
		-3	0 (0.0)	
-2	0 (0.0)	-2	2 (0.2)	
-1	11 (1.5)	-1	91 (10.9)	
0	356 (48.9)	0	710 (84.7)	
1	356 (48.9)	1	18 (2.1)	
2	3 (0.4)	2	17 (2.0)	
3	2 (0.3)			
Total	728 (100.0)	Total	838 (100.0)	

proportion of GS6/7 was 0.624. The permutation test found that the overall variation of the global proportion (0.489) was random (p = 0.06).

Natural history

Biochemical recurrence was assessed for the 1422 patients belonging to the 3 patient groups of interest (Fig. 2). Due to absent postoperative PSA records, 56 were excluded from analysis; therefore, 1366 records were analyzed. The median follow-up times were 34.2, 28.9, and 24.6 months for GS6/6, GS6/7, and GS7/7 groups, respectively, with an overall median of 28.0 months. The survival curves of the 3 groups differed significantly (p < 0.0001). The 5-year biochemical recurrence-free survival estimates were 88% (95% confidence interval [CI]: 83%-94%), 83% (95% CI: 78%-89%), and 64% (95% CI: 58%-70%) for the GS6/6, GS6/7 and GS7/7 groups, respectively. A comparison between GS6/6 and GS6/7 failed to yield a trend towards better survival in the GS6/6 group (p = 0.077).

A multiple regression Cox model was fitted on 1220 of the 1366 available patients due to missing pathologic stage or cancer volume information (Table 3). When corrected for age, PSA at diagnosis and pathologic stage, GS was an independent factor in upgrading (p < 0.0001), but the progression of GS6/7 cancers did not differ significantly from that of GS6/6 cancers (p = 0.8460).

Discussion

Under-grading of GS6 prostate cancers remains as prevalent today as it was 20 years ago. In Pinthus and colleagues' retrospective study of 698 patients with RP performed between August 1989 and April 2004, 50.3% of tumours diagnosed by TRUS-guided biopsy as GS6 were upgraded to GS7 after postoperative pathological analysis.⁵ Our analysis yielded a similar value of 48.9% in the May 2004 to April 2011 treatment cohort. However, our data were generated from a



Fig. 1. Proportion of patients graded Gleason score (GS) 6 at biopsy who were upgraded to GS7 at radical prostatectomy using a 365-day window moved every 30 days.

single Canadian academic centre and may not be applicable to other settings.

Despite the absence of a sustained decrease in discordance, we interpret the 20% decrease from late 2005 to 2006 to be a result of the ISUP update. The ISUP update well-documents the stage migration from GS 6 to GS 7 and a concurrent increase in grading accuracy.¹² The migration of higher risk GS6 patients to GS7 would artificially improve the prognosis of GS6, according to the Will Rogers phenomenon.¹⁴ We demonstrated that GS6/6 cancers have better clinical and pathological characteristics than GS6/7 upgrades. With less severe GS6 cancers being graded under the 2005 GS classification, an associated decrease in undergraded cancers and discordance were observed.

Active surveillance as a conservative management option emerged during the study period. Dall'Era and col-



Fig. 2. Biochemical-recurrence free survival in 1366 patients.

Table 3. Cox model of biochemical-recurrence	free	survival
in 1228 patients		

	HR	Lower Cl	Upper Cl	<i>p</i> value	
GS group				<0.0001	
GS 6/7 vs. GS 6/6	1.06	0.59	1.90	0.8460	
GS 7/7 vs. GS 6/6	2.27	1.36	3.78	0.0016	
Age ≥65 vs. Age <65	1.08	0.78	1.49	0.6570	
Biopsy PSA				0.1030	
PSA 5 to 10 vs. PSA <5	1.19	0.82	1.72	0.3530	
PSA 10 to 20 vs. PSA <5	1.77	1.11	2.83	0.0171	
PSA ≥20 vs. PSA <5	1.58	0.79	3.18	0.2000	
Pathologic stage T3/4 vs. T2	2.20	1.58	3.07	<0.0001	
Log (cancer volume)	1.19	1.00	1.42	0.0486	
GS: Gleason score; HR: hazard ratio; CI: confidence interval; PSA: prostate-specific antigen.					

leagues observed a sharp increase in yearly accrual into an active surveillance program in the early 1990s to today.¹⁰ We hypothesize that active surveillance could explain the increasing GS6/7 discordance from late 2006 to the study end-date. Active surveillance allows the possibility for disease progression. Re-biopsy increases the detection of GS 7 tumours due to increased sampling; however, a frequently used trigger for radical therapy is biochemical progression indicated by consecutive rises in PSA or increasing PSA velocity.¹⁵ The use of PSA as a surrogate for disease progression would contribute to discordance if, as is frequently the case, an additional biopsy is not performed between the PSA rise and RP. Also, active surveillance protocols tend to be adopted by patients with very low-risk disease, a subset of the GS6 patients with PSA <10 ng/mL.¹⁶ By removing these favourable risk GS6 patients, the proportion of patients presenting for RP will have worse disease features, which may increase the rate of upgrading.

There were different clinical features between the 3 subgroups. Compared to the GS6/6, the GS6/7 patients were older, had higher preoperative PSA, lower TRUS volume, higher tumour volume, higher positive margin rate and increased incidence of pT3 and pT4 disease. These results were similar to those reported by Pinthus and colleagues, except for age at time of treatment. Active surveillance enrolls from favourable GS6 men. Since GS6/6 is the most favourable subgroup, increased active surveillance enrolment will have its largest effect on this subgroup.⁸ The negative results concerning predictive power of biopsy core number match those reported by Pinthus and colleagues; however, it is common to find correlations between core number and accuracy.^{17,18} This may be explained by considering the biopsy regimen at our centre. Only 59 men in our cohort (3.4%) underwent the traditional sextant biopsy and the mean core number in the 3 patient subgroups was about 11.6, a comparatively extended protocol. Our results were consistent with the observation that larger tumours may harbour higher-grade disease.¹⁹

We found that, in contrast to Pinthus and colleagues' older cohort, GS6/7 tumours resembled GS6/6 in terms of biochemical progression. This can be expected considering that the 2005 ISUP update reclassified the most severe GS6 cancers as GS7, thus mostly improving the prognosis of the worst GS6 cancers and, by extension, the GS6/7 group. Active surveillance would have removed the most favourable GS6 cancers, thus worsening the prognosis of GS6/6. The net change makes the progression of GS6/6 and GS6/7 cancers more similar.

Conclusion

Over the last 20 years, the proportion of GS6 tumours upgraded to GS7 on final pathology was about half. There was some transient improvement in discordance corresponding to the ISUP update, but its effects may have been erased by the impact of active surveillance. Treatment decisions rely heavily on biopsy GS; therefore, factors that may predict discordance and strategies to minimize discordance remain important for the appropriate treatment of prostate cancer.

Competing interests: Dr. Walker, Dr. Lindner, Dr. Louis, Ms. Kalnin, Ms. Ennis, Dr. Nesbitt, Dr. Zlotta, Dr. van der Kwast, Dr. Kulkarni, Dr. Trachtenberg and Dr. Hamilton all declare no competing financial or personal interests. Dr. Jewett is a member of the Advisory Board for Pfizer. He has also received grants from Novartis, GSK and Pfizer. Lastly, he has participated in clinical trials with Novartis, GSK and Pfizer. Dr. Finelli has also participated in clinical trials in the past 2 years for Amgen, Astellas, Janssen and Ferring. Dr. Fleshner is a member of the Advisory Board for Amgen, Janssen, Astellas and Eli Lily. He has received honoraria from Amgen, Janssen, Astellas and Eli Lily. He is and has participated in clinical trials for Amgen, Janssen, Medivation, OICR, and Prostate Cancer Canada.

This paper has been peer-reviewed.

References

- Djavan B, Zlotta AR, Ekane S, et al. Is one set of sextant biopsies enough to rule out prostate cancer? Influence of transition and total prostate volumes on prostate cancer yield. *Eur Urol* 2000;38:218-24. http://dx.doi.org/10.1159/000020282
- Fleshner NE, Cookson MS, Soloway SM, et al. Repeat transrectal ultrasound-guided prostate biopsy: A strategy to improve the reliability of needle biopsy grading in patients with well-differentiated prostate cancer. Urology 1998;52:659-62. http://dx.doi.org/10.1016/S0090-4295(98)00226-X
- Carlson GD, Calvanese CB, Kahane H, et al. Accuracy of biopsy Gleason scores from a large uropathology laboratory: Use of a diagnostic protocol to minimize observer variability. Urology 1998;51:525-9. http:// dx.doi.org/10.1016/S0090-4295(98)00002-8
- Ooi K, Samali R. Discrepancies in Gleason scoring of prostate biopsies and radical prostatectomy specimens and the effects of multiple needle biopsies on scoring accuracy. A regional experience in Tamworth, Australia. ANZ J Surg 2007;77:336-8. http://dx.doi.org/10.1111/j.1445-2197.2007.04054.x
- Pinthus JH, Witkos M, Fleshner NE, et al. Prostate cancers scored as Gleason 6 on prostate biopsy are frequently Gleason 7 tumors at radical prostatectomy: implication on outcome. J Urol 2006;176:979-84. http://dx.doi.org/10.1016/j.juro.2006.04.102
- D'Amico AV, Whittington R, Malkovwicz SB, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. JAMA 1998;280:969-74. http://dx.doi.org/10.1001/jama.280.11.969

- Epstein JI, Walsh PC, Carmichael M, et al. Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer. JAMA 1994;271:368-74. http://dx.doi.org/10.1001/ jama.1994.03510290050036
- Scherr D, Swindle PW, Scardino PT. National comprehensive cancer network guidelines for the management of prostate cancer. Urology 2003;61:14-24. http://dx.doi.org/10.1016/S0090-4295(02)02395-6
- Wilt TJ, Brawer MK, Jones KM, et al. Radical prostatectomy versus observation for localized prostate cancer. N Engl J Med 2012;367:203-13. http://dx.doi.org/10.1056/NEJMoa1113162
- Dall'Era MA, Konety BR, Cowan JE, et al. Active surveillance for the management of prostate cancer in a contemporary cohort. *Cancer* 2008;112:2664-70. http://dx.doi.org/10.1002/cncr.23502
- Epstein JI. An update of the Gleason grading system. J Urol 2010;183:433-40. http://dx.doi. org/10.1016/j.juro.2009.10.046
- Hepalp B, Egevad L. The significance of modified Gleason grading of prostatic carcinoma in biopsy and radical prostatectomy specimens. *Virchows Arch* 2006;449:622-7. http://dx.doi.org/10.1007/ s00428-006-0310-6
- Katz BP. A computer enhanced runs test for smoothed data. *Stat Med* 1988;7:795-804. http://dx.doi. org/10.1002/sim.4780070709
- Hernandez DJ, Nielsen ME, Han M, et al. Natural history of pathologically organ-confined (pT2), Gleason score 6 or less, prostate cancer after radical prostatectomy. *Urology* 2008;72:172-6. http://dx.doi. org/10.1016/j.urology.2007.10.055

- Klotz L. Active surveillance for prostate cancer: Patient selection and management. *Curr Oncol* 2010;17:S11-7.
- Dahabreh IJ, Chung M, Balk EM, et al. Active surveillance in men with localized prostate cancer: A systematic review. Ann Intern Med 2012;156:582-90. http://dx.doi.org/10.7326/0003-4819-156-8-201204170-00009
- Emiliozzi P, Maymone S, Paterno A, et al. Increased accuracy of biopsy Gleason score obtained by extended needle biopsy. J Urol 2004;172:2224-6. http://dx.doi.org/10.1097/01.ju.0000144456.67352.63
- San Francisco IF, DeWolf WC, Rosen S, et al. Extended prostate needle biopsy improves concordance of Gleason grading between prostate needle biopsy and radical prostatectomy. J Urol 2003;169:136-40. http://dx.doi.org/10.1016/S0022-5347(05)64053-0
- McNeal JE, Villers AA, Redwine EA, et al. Histologic differentiation, cancer volume, and pelvic lymph node metastasis in adenocarcinoma of the prostate. *Cancer* 1990;66:1225:33. http://dx.doi. org/10.1002/1097-0142(19900915)66:6<1225::AID-CNCR2820660624>3.0.CO;2-X

Correspondence: Dr. John Trachtenberg, Fleck Tanenbaum Chair in Prostatic Diseases, Director, The Prostate Centre, Princess Margaret Hospital, Professor, Department of Surgery, University of Toronto, University Health Network, Princess Margaret Hospital, Suite 3 208, 620 University Ave., Toronto, ON M5G 2M9; john.trachtenberg@uhn.ca