The impact of the 2005 International Society of Urological Pathology Gleason grading consensus on active surveillance for prostate cancer

Venkat M. Ramakrishnan[1], Karolin Bossert[2], Gad Singer[3], Kurt Lehmann[2], Lukas J. Hefermehl[2]

[1]University of Louisville, School of Medicine, Louisville, Kentucky, United States
[2]Division of Urology, Department of Surgery, Kantonsspital Baden, Switzerland
[3]Institute of Pathology, Kantonsspital Baden, Switzerland

Introduction Current treatment plans for localized prostate carcinoma (PC) are based on core needle biopsies (CNB) classified using the Gleason score (GS). Recently, many institutions have started using the latest version of International Society of Urological Pathology (ISUP) guideline revision from 2014 for PC grading. Interestingly, this adoption is occurring without first understanding whether the 2005 ISUP revisions had a positive clinical impact. CNB-based GS may underestimate tumor aggressiveness and, therefore, critically impact patient eligibility for active surveillance (AS). The 2005 ISUP recommendations bore a significant impact on the grading of Gleason 6 and 7 PCs – a range that is meaningful for AS. The objective of this study was to compare the concordance between GS in CNB and radical prostatectomy (RP) before and after the 2005 ISUP guideline revisions, with an emphasis on its clinical impact on AS.
Material and methods This was a single-center, prospective observational study. CNB were performed in a standardized manner. GS of CNB and RP specimens were compared across three time periods: 1999–2005 (pre-revision), 2006–2007 (transitional period), and 2008–2015 (post-revision). AS is usually employed in patients with GS 6 or GS 7 PC. Thus, we therefore focused on the analysis of patients with CNBs of GS ≤7.
Results Between 1999 and 2015, 380 men with GS ≤7 PC underwent RP at our institution (median age: 62y; median PSA: 5.8 ng/ml). Of these, 231 CNB specimens were classified as GS ≤6, while 149 were GS 7. 46% (pre-revision), 43% (transitional), and 54% (post-revision) of CNB with original scores ≤6 were later upgraded in corresponding RP specimens (p <0.001).
Conclusions The 2005 ISUP GS revisions did not lower the rates of GS upgrades in RP specimens when compared to corresponding initial CNBs. Thus, these revisions did not improve AS selection. Future advances in molecular diagnostics may provide additional valuable information that facilitates patient enrollment in AS programs.

Corresponding author
Lukas J. Hefermehl
Kantonsspital Baden
Switzerland
Division of Urology
Department of Surgery
Im Ergel 1
5404 Baden, Switzerland
phone: +41 56 486 30 83
lukas.hefermehl@ksb.ch

INTRODUCTION

Pathologists grade prostate carcinoma (PC) using the Gleason grading scheme, with grades ranging from 1 to 5 [1] corresponding with increasing atypia and loss of tumor differentiation. The Gleason score (GS) is the sum of two Gleason grades. A high GS is usually associated with more aggressive disease, while a low GS is associated with a more indolent disease course. Urologists have often used the GS to provide personalized treatment options for their patients [2].

While it is standard protocol to employ Gleason scoring on core needle biopsies (CNB), it is important

to recognize that CNBs are inherently subject to sampling issues and, consequently, do not always accurately define the heterogeneity of various tumor types and their distributions. Thus, CNBs cannot necessarily completely represent all Gleason patterns or fully characterize a tumor. The literature demonstrates that anywhere between 28 and 68% of radical prostatectomy (RP) specimens had higher GS than the GS in their corresponding CNBs obtained prior to surgery [3, 4]. This, of course, has significant ramifications. Not only does this suggest that more than 33% of therapeutic decisions in PC patients might be based on inaccurate Gleason scores, but it also impacts the clinical and patient decision to pursue less-invasive approaches such as active surveillance (AS) [5].

Recently, many institutions have started using the latest version of International Society of Urological Pathology (ISUP) guideline revision from 2014 for grading prostate carcinoma (PC) [6]. Interestingly, this transformation is taking place, without knowing if the former ISUP revision from 2005 had a positive clinical impact.

In 2005, revisions to the Gleason classification system were proposed by the International Society of Urological Pathology (ISUP). Amongst other things, the rather high rate of reclassification was one of the reasons that led to revisions of the GS classification criteria over ten years ago [7]. Notable changes included the following: (1) Gleason scores of 2 through 4 per CNB were only to be reported rarely, if ever, in CNB specimens. (2) Most cribriform patterns were to be classified as GS 4. (3) Different Gleason scores were to be used for CNB and RP specimens. (4) High-grade tumors were to be included in the report and assigned a possible tertiary Gleason pattern.

Next to improving clinical reliability and reducing interobserver variability the 2005 ISUP recommendations could have potentially impacted AS, as this treatment option typically necessitates a GS of 6 (3+3) as well as the absence of any Gleason pattern 4. We expected that these changes had profound implications on our own AS program, which we initiated in 1999 [8]. Given that ours was one of the first centers in the world to perform AS, we developed a very cautious inclusion criteria (less than 3 positive cores, each with less than 5 mm extension; GS ≤6; and only unilateral cancer). We performed follow-ups biannually, with PSA-testing, digital rectal examinations, and re-biopsies every two years. Today, this program is fully adherent to the international recommendations for AS [2, 9, 10].

Prior to 2005, we and others recognized that an underestimation of tumor aggressiveness using the older Gleason criteria might itself be an independent risk factor for patients who planned to pursue AS [3]. To obviate this risk, we only offered AS to patients with CNBs of GS 5 with the rationale that any change to the GS classification would most likely lead to a relative upgrade of the Gleason grades. In 2006, according to the revised 2005 ISUP guidelines, pathologists at our institution only reported GS 6 in CNBs. We then started enrolling patients with GS 6 PC into AS.

Using 16 years of clinicopathological data from CNB and RP specimens collected at our institution, our study aims to determine whether our implementation of the 2005 ISUP guidelines lowered the rate of GS upgrades in RP specimens. These results are not only of particular interest and significance for AS, but might also be important for investigations regarding the impact of the most recent Gleason grading scheme proposed by the 2014 ISUP Consensus Conference, which uses a scale from 1–5 according to Gleason score 6–10 and is used internationally since 2016 [6, 11, 12].

MATERIAL AND METHODS

Scope of analysis

The objectives of this study were to (a) compare the GS of CNB with the GS of corresponding total RP specimen in three time periods (defined below) and (b) evaluate the impact of the 2005 ISUP revisions on the clinical decision to pursue AS. RP is the only option to determine a complete GS of an individual's PC. We limited our analysis to GS ≤6 and 7 (7a or 7b) specimens because they are crucial to AS programs limiting the patient inclusion criteria to those with GS 6 (3+3). Our rationale was that patients with PC classified as GS ≤6 may not receive adequate treatment if they undergo AS, whereas those patients with GS ≥7 would be excellent candidates for other treatment options (such as radiotherapy, prostatectomy, etc.). This study was approved by the Swiss Ethics Committee. Due to the long timeframe and study design, the Ethics Committee permitted us to analyze data without written informed consent; most of the patients after 2004 signed informed consents.

Biopsy & surgery techniques and histopathology

Only treatment-naive patients were included in this study. All patients in this study were exposed to the same CNB and RP techniques. We retrospectively analyzed data from patients with GS ≤7 biopsies who underwent open retropubic RP between July 1999 and July 2015. In these patients, CNB

were performed using a standardized trans-rectal ultrasound guided ten-core biopsy technique under local anesthesia in a left lateral position. CNB and RP specimens were fixed in 4% formalin, routinely processed, and analyzed by certified pathologists. The former GS classification criteria was used until December 2005. In January 2006, our pathologists started using the revised Gleason classification criteria [7]. Furthermore, we recorded the age at biopsy, results of the PSA testing, prostate volume, and results of the digital rectal examination. GS of all CNB and corresponding RP specimens were compared.

Time periods

Patients who underwent biopsies between 1999 and 2005 were placed into the 'pre-revision' cohort, whereas those who underwent biopsies between 2008 and 2015 were placed into the 'post-revision' cohort. The cohort with men who underwent biopsy between 2006 and 2007 fell into a 'transitional period'.

Statistics

Statistical analysis was performed using GraphPad Prism (v6.0). The Mann-Whitney U and Kruskal-Wallis tests and Chi-squared analysis were applied. All statistical tests were two-sided, with p ≤0.05 considered statistically significant.

RESULTS

Definition of cohort & sub-cohorts

Between July 1999 and June 2015, our institution completed 482 radical prostatectomies. Of these,

380 had GS ≤7 PCa on CNB. The median age of this cohort was 62 (range: 58–66) years, while the median PSA was 5.8 (4.4–8.1) ng/ml. Within this cohort, 231 specimens were GS ≤6 and 149 were GS 7.

The 'pre-revision' (1999–2005), ' transitional period' (2006–2007), and 'post-revision' (2008–2015) sub-cohorts consisted of 180, 55 and 145 men, respectively. The proportion of GS ≤6 to GS 7 specimens within these sub-cohorts were approximately 4.29, 1.75, and 0.53, respectively (Table 1). The sub-cohorts were not significantly different with respect to their clinical parameters (Tables 2 and 3). For total prostatectomy specimens, the median score was 6. However, a group-wise comparison using the Kruskal-Wallis test calculated a significantly lower score (p = 0.004) for the pre-revision cohort.

Comparison of pre- and post-revision criteria

We first sought to compare the upgrade and downgrade rates in the GS ≤6 and GS 7 (Figure 1) cohorts as well as their sub-cohorts. Broadly, 89 of 231 (38.5%) patients with PC specimens graded as GS ≤6 on CNB demonstrated GS 7 in their corresponding RP specimens.

46% (67 of 146), 43% (15 of 35) and 54% (27 of 50) of those in the pre-revision, transitional and post- revision groups, respectively, demonstrated higher GS

Table 1. *Breakdown of specimens in each time period*

Sub-cohort	Sample size (n) of GS ≤6	Sample size (n) of GS = 7	Ratio of GS ≤6 to GS = 7
Pre-revision GS	146	34	4.29
Transitional GS	35	20	1.75
Post-revision GS	50	95	0.53

Table 2. *Clinical characteristics for GS ≤6 cohort*

Sub- Cohort	Sample Size (n)	Age median y	PSA median ng/ml	Prostate Volume ml	DRE -/+	Biopsy GS ≤6 mean 95% CI	RP GS mean 95% CI	Paired Wilcoxon t Test
Pre-revision GS	146	61 (57–65)	6.1 (4.3–8.5)	31 (24–400)	102/44	5.4 (5.3–5.9)	6.1 (5.9–6.2)	<0.001
Transitional GS	35	63 (60–66)	5.1 (4.3–8.4)	40 (22–47)	29/6	6.0	6.6 (6.3–6.9)	0.002
Post-revision GS	50	64 (60–67)	5.1 (4.3–6.4)	37 (27–51)	48/2	6.0	6.5 (6.4–6.7)	<0.001

Table 3. *Clinical characteristics for GS 7 cohort*

Sub- Cohort	Sample Size (n)	Age median y	PSA median ng/ml	Prostate Volume ml	DRE -/+	Biopsy GS mean 95% CI	RP GS mean 95% CI	Paired Wilcoxon t Test
Pre-revision GS	34	62 (59–67)	7.3 (7.7–11.1)	31 (26–42)	17/17	7	7.1 (7.0–7.2)	0.336
Transitional GS	20	62 (56–66)	5.4 (4.7–8.5)	27 (21–38)	10/10	7	7.1 (6.8–7.4)	0.75
Post-revision GS	95	64 (60–67)	6.0 (5.4–7.4)	39 (30–49)	71/24	7	7.1 (7.10–7.2)	0.064

with RP specimens when compared against their corresponding CNB GS scores. In all three sub-cohorts, these changes were significant (p <0.001, = 0.002, < 0.001, respectively).

11% (16 of 149) of the patients initially classified as GS 7 demonstrated higher GS scores in their RP specimens. Specifically, 18% (6 of 34), 10% (2 of 20) and 8% (8 of 95) showed upgraded RP GS in the pre- revision, transitional, and post-revision groups. In all three of these sub-cohorts, these changes were not significant (p = 0.336, 0.75, 0.064, respectively). Remarkably, significantly more patients had to be reclassified from GS ≤6 to GS >7 when comparing the post-revision and pre-revision cohorts (p = 0.006).

DISCUSSION

Our objective was to ultimately establish the effects of the 2005 ISUP modifications to the GS classification criteria on our decision to enroll patients in AS. One key finding is that a significant number of CNB specimens with GS ≤6 demonstrated higher Gleason scores in their corresponding RP specimens irrespective of the grading criteria used. This is very important, since patients with GS ≤6 PCs (presumed to be low-grade, low-volume cancers) are typically selected for AS. The definition of a PC as 'low grade' and 'low volume' critically depends on GS [2, 9, 10]. Thus, a significant number of patients with GS ≤6 may actually have higher Gleason scores in RP. In the spectrum of PC treatment options, AS plays a special role: the active treatment – if necessary – is postponed. Falsely down-graded scores in men who choose AS may turn into a missed opportunity to have intervened early-on. Thus, for AS patients, establishing concordant scores between CNB and RP is crucial. The number and percentage of men included in AS at our institution were 65 (36%), 22 (12%) and 94 (52%) in the pre-revision, the transitional, and post-revision phases, respectively. Reclassification from AS in the time groups showed no significant difference (2 test p = 0.7; data not shown). Thus, the probability of a bias is marginal.

The literature demonstrates conflicting results. Ozok et al. and Helpap et al. demonstrated improved concordance using the revised GS classification criteria in studies involving 97 and 368 patients, respectively [13]. Although these studies were published soon after the 2005 revision was implemented, they may not adequately or accurately represent the last ten years' worth of findings. Other studies have reported no improvement [14].

An inverse relation between prostate volume and reclassification may exist [15]. Sampling problems
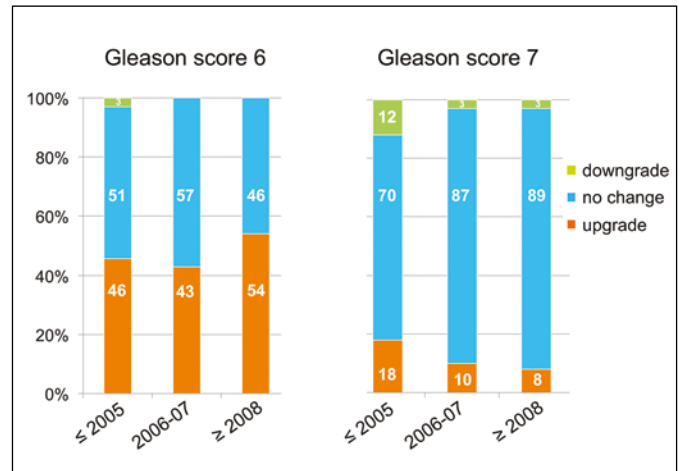

Figure 1. *Comparison of upgrade and downgrade rates in GS ≤6 and GS 7 cohorts and sub-cohorts. Those patients classified as GS ≤6 were generally upgraded pre- and post-revision, and such changes were significant. Those classified as GS >7 demonstrated relatively low upgrade rates pre- and post-revision, and such changes were not significant. Significantly more patients were reclassified from GS ≤6 to GS >7 when comparing post- and pre-revision cohorts (p = 0.006). GS – Gleason score*

by needle biopsy may be attributed to the high variability of cylinder lengths and/or poor sampling technique. Interestingly, larger tumors that were evaluated by CNB were associated with a significantly increased rate of reclassification to a higher GS in corresponding RP specimens.

Sampling problems in CNB specimens with limited tissue and/or tumor heterogeneity can have a significant impact on the discrepancy between Gleason scores in CNB and corresponding RP specimens. Although our biopsy technique remained stable over the whole study period, we believe that the reproducibility of any single biopsy session is low. In our opinion, two factors are error-prone: (a) the placement and direction of the biopsy needle and (b) the effectuated cylinder lengths. Sampling error (by missing the relevant cancer load during a NB) might be an important cause for secondary reclassification. Even though it is hard to prove, reclassification from AS to active treatment could be due to sampling error during the initial prostate biopsy and less commonly due to real tumor progress.

Pfirrmann et al. demonstrated that the number of positive cores does not correlate with upgrade in 'very low risk' tumors; rather, an increase in the number of cores taken correlates with a decrease in the probability of downstream GS upgrading [16]. This is a strong argument for performing early re-biopsies in patients who plan to pursue AS. At the

time of enrollment into an AS program, these patients must be informed that they carry a significant risk of being under-graded. Considering the recent upswing of technologies such as MRI and fusion biopsy, investigations looking at different biopsy and imaging techniques that address GS discordance should be emphasized [17].

CONCLUSIONS

In summary, our findings suggest that the decision-making criteria for AS are still not clear-cut. Our study shows that the 2005 revisions to the Gleason System did not significantly improve the concordance between the Gleason scores of CNBs and their corresponding RP specimens. A GS ≤6 on CNB ultimately yielded a higher GS in the RP specimen in more than 40% of patients. This may be due to inconsistencies with tissue sampling and/or tumor heterogeneity. Data regarding the impact of the new 2014 ISUP revisions on AS enrollment are still being collected. In the future, molecular diagnostic methods will provide additional, valuable information that facilitates a patient's selection for AS. In the meantime, we suggest that patients be informed about the possibility that their tumor may be under-graded at the time of AS enrollment. Consequently, close follow-up is warranted.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

References

1. Gleason DF, Mellinger GT. Prediction of prognosis for prostatic Staging, adenocarcinoma by combined histological grading and slinical. J Urol. 1974; 111: 580-564.

2. Heidenreich A, Bastian PJ, Bellmunt J, et al. EAU guidelines on prostate cancer. part 1: screening, diagnosis, and local treatment with curative intent-update 2013. Eur Urol. 2014; 65: 124-137.

3. Egevad L, Mazzucchelli R, Montironi R. Implications of the International Society of Urological Pathology modified Gleason grading system. Arch Pathol Lab Med. 2012; 136: 426-434.

4. Uemura H, Hoshino K, Sasaki T, et al. Usefulness of the 2005 International Society of Urologic Pathology Gleason grading system in prostate biopsy and radical prostatectomy specimens. BJU Int. 2009; 103: 1190-1194.

5. Epstein JI, Feng Z, Trock BJ, Pierorazio PM. Upgrading and downgrading of prostate cancer from biopsy to radical prostatectomy: Incidence and predictive factors using the modified gleason grading system and factoring in tertiary grades. Eur Urol. 2012; 61: 1019-1124.

6. Epstein JI, Egevad L, Amin MB, et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. Am J Surg Pathol. 2016; 40: 244-252.

7. Epstein JI, Allsbrook WCJ, Amin MB, Egevad LL. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. Am J Surg Pathol. 2005; 29: 1228-1224.

8. Hefermehl LJ, Disteldorf D, Lehmann K. Acknowledging unreported problems with active surveillance for prostate cancer: a prospective single-centre observational study. BMJ Open. 2016; 6: e010191.

9. Thompson I, Thrasher JB, Aus G, et al. AUA Guideline for the Management of Clinically Localized Prostate Cancer: 2007 Update. J Urol. 2014; 177: 2106-2131.

10. Interdisziplinäre Leitlinie der Qualität S3 zur früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms, Langversion 3.1, AWMF Registernummer: 034/022. http://leitlinienprogrammonkologie.de/Leitlinien70.html.

11. Danneman D, Drevin L, Delahunt B, et al. Accuracy of prostate biopsies for predicting Gleason score in radical prostatectomy specimens: Nationwide trends 2000-2012. BJU Int. 2017; 119: 50-56.

12. Epstein JI, Zelefsky MJ, Sjoberg DD, et al. A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score. Eur Urol. 2016: 69: 428-435.

13. Helpap B, Egevad L. The significance of modified Gleason grading of prostatic carcinoma in biopsy and radical prostatectomy specimens. Virchows Archi. 2006; 449: 622-627.

14. Zareba P, Zhang J, Yilmaz A, Trpkov K. The impact of the 2005 International Society of Urological Pathology (ISUP) consensus on Gleason grading in contemporary practice. Histopathology. 2009; 55: 384-391.

15. Sfoungaristos S, Perimenis P. Clinical and pathological variables that predict changes in tumour grade after radical prostatectomy in patients with prostate cancer. Can Urol Assoc. 2013; 7: E93-97.

16. Thomas C, Pfirrmann K, Pieles F, et al. Predictors for clinically relevant Gleason score upgrade in patients undergoing radical prostatectomy. BJU Int. 2012; 109: 214-219.

17. Sonn GA, Natarajan S, Margolis DJA, et al. Targeted biopsy in the detection of prostate cancer using an office based. J Urol. 2013; 189: 86-91. ■