The modified ISUP system improves concordance between biopsy and prostatectomy tumour grade, independent of pre-biopsy MRI and biopsy method.

David M.Z.B. Hennes1, James Sewell1, Michael Kerger2, Christopher M. Hovens2, Justin S. Peters1, Anthony J. Costello1,2, Andrew Ryan3, Niall M. Corcoran1,2,4
1Department of Urology, Melbourne Health, Royal Melbourne Hospital, Parkville, VIC, Australia
2Department of Surgery, University of Melbourne, Parkville, VIC, Australia
3TissuPath, Mount Waverly, VIC 3149, Australia
4Victorian Comprehensive Cancer Centre, Melbourne, VIC 3000, Australia

Correspondence:
Dr David Hennes
Email: drdavidhennes@gmail.com
Phone: +614 3140 4710
The modified ISUP system improves concordance between biopsy and prostatectomy tumour grade.

Abstract

Objectives
- To assess the concordance between biopsy and radical prostatectomy (RP) specimens using the 2005 Gleason Score (GS) and the ISUP 2014/WHO 2016 modified GS system, accounting for the introduction of transperineal biopsy and pre-biopsy multiparametric MRI (mpMRI).

Patients and Methods
- Between 2002 and 2019, we identified 2431 patients with paired biopsy and RP histopathology, from a prospectively recorded and maintained prostate cancer database.
- Biopsy specimens were graded according to the 2005 GS or ISUP 2014 modified GS system, according to the year of diagnosis.
- Multivariable logistic regression analysis was conducted to retrospectively assess the impact of PSA, PSA density, age, pre-biopsy mpMRI, and biopsy method, on the rate of upgraded disease.
- The Kappa coefficient was used to establish the degree of change in concordance between groups.

Results
- Overall, 24% of patients had upgraded disease and 8% of patients had downgraded disease when using the updated ISUP 2014 criteria.
- Agreement in the updated ISUP 2014 cohort was 68% compared with 55% in the 2005 GS group, which was validated by a kappa co-efficient that was good (k=0.5 ± 0.4) and poor (k=0.3 ± 0.1),
Prostate cancer (PCa) grading, first implemented by Donald Gleason between 1966 and 1974, is consistently the most important predictor of overall outcome in localised disease, and therefore strongly guides clinicians’ treatment decisions [1]. Prostate cancer diagnoses crucially rely on needle biopsy findings, however, the discordance rate between needle biopsy and post-radical prostatectomy (RP) histopathology is reported to be as high as 50% using the ISUP 2005 Gleason Grading System (GS) [2]. In fact, low-grade disease on needle biopsy (Gleason ≤ 6) has shown an upgrading rate of 36% (range 14-51%) in post-prostatectomy specimens using this classification. Conversely, high-grade disease (Gleason ≥ 8) has shown a downgrading rate of up to 56% (range 29-56%) between needle biopsy and prostatectomy specimens using the older system [3,4].

At the International Society of Urological Pathology (ISUP) conference in 2014, Epstein and colleagues made several modifications to the original Gleason Grading System to improve interobserver reproducibility and decrease the rate of upgrading from biopsy to RP specimens [5, 6]. Under the new classification, all tumours with cribriform patterns of invasive adenocarcinoma were re-classified as Gleason Pattern (GP) 4, due to growing evidence of tumour aggression and adverse outcomes of all cribriform patterns [7]. Therefore, GP4 now includes cribriform glands, poorly formed glands, fused glands, and glomeruloid glands [7]. The 2014 ‘WHO Classification of Tumours

Respectively.

In multivariable models, a change in ISUP grading system independently improved overall disease concordance (p=0.02), and there were no other co-segregated patient or pathological factors such as PSA, total number of cores, maximum cancer length, biopsy route or the use of mpMRI that impacted this finding.

**Conclusion**

- The 2014 ISUP modification of the Gleason grading system improves overall concordance between biopsy and surgical specimens, and thus allows more accurate prognostication and management in high-grade disease, independent of more extensive prostate sampling and the use of mpMRI.

**Key words**: Adenocarcinoma, Prostate, Grading, Concordance, Gleason Score, ISUP Grade Group, Upgrade, Biopsy

**Introduction**

Prostate cancer (PCa) grading, first implemented by Donald Gleason between 1966 and 1974, is consistently the most important predictor of overall outcome in localised disease, and therefore strongly guides clinicians’ treatment decisions [1]. Prostate cancer diagnoses crucially rely on needle biopsy findings, however, the discordance rate between needle biopsy and post-radical prostatectomy (RP) histopathology is reported to be as high as 50% using the ISUP 2005 Gleason Grading System (GS) [2]. In fact, low-grade disease on needle biopsy (Gleason ≤ 6) has shown an upgrading rate of 36% (range 14-51%) in post-prostatectomy specimens using this classification. Conversely, high-grade disease (Gleason ≥ 8) has shown a downgrading rate of up to 56% (range 29-56%) between needle biopsy and prostatectomy specimens using the older system [3,4].

At the International Society of Urological Pathology (ISUP) conference in 2014, Epstein and colleagues made several modifications to the original Gleason Grading System to improve interobserver reproducibility and decrease the rate of upgrading from biopsy to RP specimens [5, 6]. Under the new classification, all tumours with cribriform patterns of invasive adenocarcinoma were re-classified as Gleason Pattern (GP) 4, due to growing evidence of tumour aggression and adverse outcomes of all cribriform patterns [7]. Therefore, GP4 now includes cribriform glands, poorly formed glands, fused glands, and glomeruloid glands [7]. The 2014 ‘WHO Classification of Tumours

This article is protected by copyright. All rights reserved
of the Prostate’ also recommended and accepted these changes, and the reassignment of the calculated Gleason Scores to a new five-tier Gleason grade designation consisting of a series of Grade Groups’ (Table 1); GG1 (≤ GS 6), GG2 (GS 7, 3 + 4), GG3 (GS 7, 4 + 3), GG4 (any GS 8), and GG5 (GS 9-10) [8,9]. The differentiation between Gleason 7 (3+4, 4+3) cancers in this classification improves prognostication, and therefore more accurately guides treatment decisions [9]. Some cancers previously graded as Gleason score 6 under the 2005 system are now graded as Gleason 7 (GG2/3). Therefore, Gleason 6 or GG1 cancers classified in the ISUP 2014 modified GS system carry a better prognosis than Gleason 6 under the 2005 grading system[7].

Contemporaneous with these modifications, there have been significant changes in how prostate biopsies are performed, including the use of pre-biopsy multiparametric MRI (mpMRI) by up to 67% [8], and a change from the previously standard 12-16 core transrectal biopsy, to transperineal saturation and targeted prostate biopsies [9]. These changes alone have improved the diagnostic accuracy of biopsy specimens to a variable extent, making the incremental value of the new ISUP grading system in improving disease concordance somewhat unclear. The current study aimed to assess the concordance between needle biopsy and prostatectomy specimens in 2431 patients who underwent RP at a single centre, using the 2005 GS system and the updated ISUP 2014 modified GS system. In addition, we aimed to assess the impact of more extensive and more accurate prostate sampling on overall disease concordance. In doing so, we hoped to further validate the use of the modified ISUP 2014 in PCa grading, in an era where tissue sampling is changing.

**Patients and Methods:**

All patients undergoing RP between January 1st, 2002 and May 31st, 2019, with biopsy and corresponding RP histopathology were identified from a prospectively recorded and maintained prostate cancer database. Clinical and pathological data were recorded in the Urological Cancer Biorepository study, approved at the Epworth Medical Centre (HREC protocol #34506). For the current study, Quality Assurance (QA) ethics approval was obtained from the Royal Melbourne Hospital Research Ethics Committee (protocol #QA2019077). Patients receiving neoadjuvant androgen deprivation therapy, prior radiotherapy, hormonal therapy, or other experimental therapies that may interfere with RP histopathology were excluded from the analysis. 44 patients were excluded from analysis due to a lack of biopsy results.
From a final cohort of 2431 patients, we collected data regarding patient age, pre-operative PSA, PSA density, biopsy method, pre-biopsy MRI, Prostate Imaging Reporting and Data System (PIRADS) score, number of biopsy cores, number of cores positive, and maximum linear cancer length. MRI data was collected by searching medical imaging and patient records for mpMRI reports or the presence of a target biopsy. The concordance of pathology grade between biopsy and RP specimens was then retrospectively analysed to assess overall agreement, the rate of upgrading, and the rate of downgrading between specimens. Experienced uropathologists assessed the macroscopic, microscopic and histopathological findings of core and needle biopsy samples according to the ISUP grading system used at the time of the sample collection. Before January 1st, 2015, the 2005 GS grading system was used. After this date, the ISUP 2014 modified GS system was the grading system used (Table 2).

Patients were deemed concordant if there was no difference in tumour grade between biopsy and RP specimens using the 2005 GS or 2014 ISUP modification of the Gleason grading system, respectively. Patients in whom there was not agreement were classified as discordant. Examples of discordance included a GS increasing from ≤7 to ≥8, or a GS decreasing from ≥3 to ≤2.

Multiparametric MRI:

A 3-Tesla scanner was used to perform T1 and T2-weighted, high spatial resolution anatomical imaging. Standard MRI prostate protocol included axial diffusion-weighted imaging (b values 50, 800 and 1200), apparent diffusion coefficient (ADC) mapping, and dynamic contract enhancement with an IV gadolinium DTPA bolus of 10ml.

Statistical Analysis:

Data on baseline continuous variables were presented as medians with interquartile range, whilst categorical variables were described using numbers and proportions. Differences between continuous variables were analysed using the Mann Whitney U-test as appropriate. Differences between categorical variables were analysed using Pearson’s Chi-squared test as appropriate. Concordance between biopsy and RP histological findings was evaluated with the kappa coefficient of agreement: <0.4 poor agreement, 0.4- 0.75 good agreement, and >0.75 excellent agreement. The association of the ISUP grading system used and disease concordance was assessed using a multivariable binary logistic regression model that separately accounted for patient age, PSA, biopsy method, pre-biopsy MRI, and maximum cancer length. A multivariable logistic regression was then repeated, where MRI results were imputed in randomly selected patients in the 2014 modified ISUP
group, for whom no MRI data or target biopsy report was available. This was generated to a sample that represented 67% of the total cohort, consistent with the rate of MRI cited in literature [8]. Joint posterior distribution modelling of the incomplete data was used to generate a multivariate normal imputation model that was directly compared to the previous logistic regression analysis. All data were analysed with SPSS software statistics (IBM, Sydney, NSW, Australia).

Results:

Baseline clinical and demographic data for the 2431 patients meeting inclusion criteria for the study are shown in Table 1 (supplementary material). Differences in clinical and pathological characteristics between the 2005 Gleason Score and 2014 modified ISUP cohorts are summarized in Table 2. The median number of biopsy cores taken demonstrated a statistically significant increase between groups, consistent with the increased implementation of transperineal saturation biopsies (13 vs 20, \(p=0.01\)).

A comparison of the concordance rates between needle biopsy and final prostatectomy histopathology is shown in Figure 1. The GS system was used in 2062 patients; 18 of which had a suspicious lesion identified on MRI, and 87% of patients undergoing TRUS biopsies. There was agreement in 1162 (55%) GS needle biopsy specimens, with 684 patients (33.2%) having upgraded disease and 248 patients (23%) having downgraded disease at review of prostatectomy histology (Figure 1). The ISUP 2014 modified system was used in 369 needle biopsy specimens, with transperineal biopsy being performed in 25% of cases and 33% of patients undergoing per-biopsy MRI. In this group, 252 patients (68%) had agreement, 88 (23%) had disease upgrade and 29 (8%) had downgraded disease. Across both cohorts, a total of 66 patients had a suspicious lesion identified on mpMRI. Of the 137 patients who underwent mpMRI, PIRADS 2 was reported in 11 patients, PIRADS 3 was reported in 9 patients, PIRADS 4 was reported in 61, and PIRADS 5 was reported in 56 patients.

The highest rate of upgrading was found in the GS 3+3 group and the highest downgrading rate was seen in the GS 4+4 group, 48% and 43%, respectively (Figure 2). Conversely, a maximum upgrading rate of 41% and maximum downgrading rate of 9% was demonstrated in GG1 and GG4, respectively (Figure 3). The kappa-statistics measures of agreement between needle biopsy and radical prostatectomy specimens were poor and good, respectively, for the GS and 2014 modified GS system cohorts (2005 GS: \(k=0.384 \pm 0.14\) vs 2014 ISUP: \(k=0.522 \pm 0.38\)).
The 2014 ISUP modification of the Gleason grading system presented an overall improvement in clinically significant concordance between biopsy and prostatectomy histology when compared to the 2005 GS system \((p=0.001, \text{Table 4, Figure 2})\). Multivariable regression analysis demonstrated no significant change in disease concordance with the established patient and pathological characteristics, and most notably demonstrated no association between improved concordance and biopsy method used, or the use of pre-biopsy mpMRI.

**Discussion:**

Discordance of tumour grade between needle biopsy and RP specimens is an independent and clinically significant prognostic indicator in PCa. Patients with upgraded disease exhibit more aggressive pathological features than concordant tumours, and a higher risk of biochemical recurrence post RP[2]. The 2014 ISUP modification of the Gleason grading system was implemented to more accurately stratify cribriform architectural patterns as Gleason pattern (GP) 4, associated with greater tumour aggression. The effect of this change resulted in a lower incidence of upgraded disease in surgical specimens and thereby improved prognostic indicators such as biochemical recurrence, tumour progression, metastases and overall survival[2].

De Nunzio et al. (2018) assessed the impact of the change to the updated ISUP 2014 criteria on the rates of downgrading and upgrading from biopsy to RP specimens in 9703 patients. The updated ISUP 2014 criteria presented a lower upgrading rate of 19.5% compared with 24.0% in the 2005 GS group \((p = 0.001)\), and a similar downgrading rate \((7.7\% \text{ vs } 8\%; p = 0.267)\) [13]. Factors previously associated with discordant disease include age \(\geq 60\) years, PSA density \(\geq 0.2\), \(\geq 2\) positive cores, \(\geq 5\%\) core tissue involvement and perineural invasion [14]. Concomitant with the ISUP 2014 modifications, other measures have been introduced to improve accuracy in PCa diagnoses. For instance, attention has turned to TP prostate biopsy to reduce rates of infection related complications, and improve rates of detection of clinically significant prostate cancer (csPCa). It is suggested that TP biopsy results in improved disease detection through extensive sampling of the anterior prostate where clinically significant disease is commonly missed by the posterior approach of the TRUS biopsy [15]. Another change in urologists’ practice has been the increased use of mpMRI, both as a triage test, and to aid target biopsies. The extent to which these practice changes have improved disease concordance yet to be fully quantified, though a local study demonstrated a lower rate of disease upgrade (17%) when a trizonal biopsy schema was used to target cores identified from mpMRI [16]. However, a particular limitation to the use of mpMRI has been its low
positive predictive value of 24-68% and mixed detection rates particularly in low grade disease (ISUP 1-2) [17].

This study demonstrated that the 2014 ISUP modification of the Gleason grading system has improved overall disease concordance (54.8% vs 68.2%, \( p=0.001 \)) in the GS and modified GS groups, respectively. In particular, when using the 2014 modified ISUP system, 23.8% of patients had upgraded disease, comparatively fewer than the 33.2% of patients classified using the 2005 GS system. Fewer patients also had downgraded disease at RP when the modified GS system was compared to the 2005 GS system, 7.8% vs 23.3%, respectively. Figures 1 and 2 demonstrate a clinically significant reduction in the rate of downgrading and upgrading disease in the vast majority of individual Gleason grade groups when the modified ISUP 2014 GS system is compared to the 2005 GS system. When using the GS system, the GS 3+3 group had an upgrading rate of 47.7%. There was a non-significant decrease in the rate of upgrading at 41.3% from GG1 when the 2014 ISUP modification \((p=0.3)\) was used. This is largely attributed to the limited power allowed by a sample size of 35. In order for this difference to reach statistical significance, there would need to be a total of 945 patients in the ISUP 2014 group. GG4 and GG5 also demonstrated more concordant disease, with a clinically significant decrease in the rate of downgrading when GS 4+4 was compared to GG4 (43.2% vs 18.5%, respectively) and GS 5+5 was compared to GG5 (12% vs 1.6%, respectively).

In a multivariable logistic regression model, we examined the impact of change in ISUP grading system on concordance co-linear variables including PSA, PSA density, histological grade, number of total cores, number of positive cores, the maximum linear length of tumour, biopsy method and the provision of mpMRI (table 3). We found that no co-linear variables demonstrated statistically significant change in disease concordance rate, unlike the change in ISUP grading system \((p=0.02, [CI 1.01 – 2.06])\). Pre-biopsy MRI did not reach significance in our analysis \((p= 0.2, [CI 0.7-2.6])\), similar to biopsy method \((p= 2.5, [CI 0.84-2.54])\). Over the last decade, within the time that the ISUP 2014 modifications were made, most academic centres in Australia transitioned to the use of transperineal saturation biopsy, pre-biopsy MRI and MRI-targeted biopsies. At the Epworth medical centre, Urologists adopted tranperineal template sampling as early as 2013, however our data reflect that uptake of this sampling method was initially mixed. This is perhaps explained by variance in the individual practice of private urologists, avoidance of the out-of-pocket cost incurred prior to its Medicare rebate, and a proportion of patients undergoing needle biopsy in provincial settings prior to being referred on to tertiary uro-oncologists. To improve the accuracy of our analysis in the setting of these low figures, we assumed randomly imputed data for 67% (250/369) of patients for
whom no MRI data was available. Joint modelling was used to re-compute the multivariable logistic regression analysis, which demonstrated that change in ISUP remained the only significant variable associated with improved concordance (p= 0.03, [CI 0.98 – 1.99]), and mpMRI did not reach significance in this model either (p=0.3, [CI 0.85- 1.4]).

Thus, this study suggests that the grading changes made in the ISUP 2014 modifications improve pre and post-RP concordance, independent of the introduction of transperineal saturation biopsy, pre-biopsy MRI and targeted biopsies. The improved agreement between biopsy and RP specimens demonstrated by the modified GS system can be understood by the rationale behind its change. In the original Gleason system, cribriform glands were typically graded as GP3; however this histological pattern has been associated with significant tumour aggression, resulting in poorer prognosis in multiple studies [18, 19]. On the basis of that evidence, Epstein and colleagues re-assigned tumours with even small cribriform gland predominance from GP3 to GP4. The 2014 modified ISUP changes also resulted in the subclassification of previous GS 7 cancers to GG2 (previous 3+4) and GG3 previous 4+3 (see Table 1)[9]. This change was implemented to more accurately identify those men who had higher risk of disease progression with Gleason scores of 4+3 but were previously grouped as Gleason 7 tumours, alongside those who had less cribriform predominance in the 3+4 pattern. Furthermore, smaller tumour volumes have been associated with higher risk of upgraded disease and biochemical recurrence [2,9]. This has been attributed to sampling error that occurs with decreased tissue volume resulting in inadequate biopsy samples, and subsequent upgrading. The 5-tier modified GS system inherently mitigates this error by reflecting the absolute volume of higher grade elements (e.g. cribriform predominance), rather than relative proportions of two predominant patterns [2]. A secondary change in the modified GS system is that not all mucinous (colloid) carcinoma of the prostate are assigned GP4 as was previously practiced, but they are now graded on the basis of underlying growth patterns. Theoretically, the effect of this change would be an increased rate of downgrading tumour (from GP4 to GP3) and potentially influential to our data. However, the material effect of this is small, as overall material mucinous change is not often encountered or described by our centre’s uropathologists [20].

It has been validated that even when adjusting for variables including clinical stage, PSA, total number and proportion of positive cores, disease upgrade after local curative surgery is an independent predictor of biochemical recurrence when using the 2005 GS system [2]. Turan et. al (2018) similarly demonstrated statistically significant decrease between the five-year biochemical recurrence free survival rates in patients with upgraded and concordant disease at RP when using
the 2014 ISUP modification of the Gleason grading system (55% vs 86%, respectively) [21]. For this reason, the overall accuracy of the biopsy grade in representing true disease status is crucial.

This study demonstrated that the 2014 ISUP modification of the Gleason grading system significantly improves overall concordance between biopsy and surgical specimens, independent of the concomitant transition to transperineal biopsy and pre-biopsy MRI. A crucial question that is yet to be studied is whether patients experiencing an upgrade in their Gleason score after surgery have a clinical outcome that is similar to that of concordant tumours of the higher grade, concordant tumours of the lower grade, or somewhere in between [22]. It is suggested that future studies might explore this question, to ultimately investigate whether specific clinical outcomes in PCa have been impacted by the ISUP 2014/WHO 2016 modifications. Urologist-led research will ensure accuracy and completeness of concordance data assessments and it is therefore imperative that any validation or replication studies that follow engage urologists early [23].

The limitations of this study include the lack of complete data availability inherent to retrospective analyses. There is a possibility of inherent patient selection bias due to the implementation of mpMRI, resulting in an unknown true PCa status amongst patients with PIRADS 2 disease who would have historically had biopsies performed [24]. The measurement of effect of ISUP grading system change is limited by the relatively recent changes in what is a large historical data-set, meaning that that data could only be collated in the four years subsequent to the grading system change. Future studies should implement a larger sample size with more equivalent representation of both cohorts to support the notion that the ISUP changes improve disease concordance independent of transperineal biopsy and pre-biopsy MRI. Furthermore, the impact of upgraded intermediate or high risk disease on overall cancer outcomes using the ISUP 2014 modified GS system should be assessed and compared to similar findings in previous studies assessing this using the 2005 GS system.

Source of Funding: departmental.

Conflicts of Interest:
None declared.

References:


This article is protected by copyright. All rights reserved


Author/s:
Hennes, DMZB; Sewell, J; Kerger, M; Hovens, CM; Peters, JS; Costello, AJ; Ryan, A; Corcoran, NM

Title:
The modified International Society of Urological Pathology system improves concordance between biopsy and prostatectomy tumour grade

Date:
2021-12

Citation:

Persistent Link:
http://hdl.handle.net/11343/298937