External validation of the novel International Society of Urological Pathology (ISUP) Gleason grading groups in a large contemporary Canadian cohort

Helen Davis Bondarenko, MD*1; Marc Zanaty, MD*1; Sabrina S. Harmouch, MD1; Cristina Negrean, MD1; Raisa S. Pompe, MD1,2; Daniel Liberman, MD1; Naeem Bhojani, MD1; Pierre I. Karakiewicz, MD1; Kevin C. Zorn, MD1; Assaad El-Hakim, MD1,3

1Division of Urology, Department of Surgery, Centre Hospitalier de l’Université de Montréal (CHUM), Montreal, QC, Canada; 2Martini-Klinik Prostate Cancer Centre, University Hospital Hamburg-Eppendorf, Hamburg, Germany; 3Division of Robotic Urology, Department of Surgery, Hôpital du Sacré-Coeur de Montréal, Montreal, QC, Canada

*Co-first authors

Abstract

Introduction: We sought to test the discriminatory ability of the 2014 International Society of Urological Pathology (ISUP) Gleason grading groups (GGG) for predicting biochemical recurrence (BCR) after robot-assisted radical prostatectomy (RARP) in a large, contemporary, Canadian cohort.

Methods: A total of 621 patients who underwent RARP in two major Canadian centres were identified in a prospectively maintained Canadian database between 2006 and 2016. Followup endpoint was BCR. Log-rank test, univariable, and multivariable Cox regression analyses were used.

Results: Mean followup was 27.9 months. All five ISUP GGG independently predicted BCR. Statistically significant differences in BCR rates were found between GGG 2 and GGG 3 strata (p<0.001). No statistically significant differences in BCR rates were found between GGG 4 and GGG 5 strata (p=0.3). Relative to GGG 1, the GGG 2, GGG 3, GGG 4, and GGG 5 yielded a 1.10-, 3.44-, 4.18-, and 4.74-fold hazard ratio (HR) increment in BCR, respectively.

Conclusions: This population-based Canadian cohort study confirms the added discriminatory property of the novel ISUP grading, specifically for GGG 2 and GGG 3 strata. No difference, however, was observed between GGG 4 and GGG 5, likely due to the lower number of patients in these groups. As such, after external validation, the 2014 ISUP GGG appears to retain clinical prognostic significance in a Canadian population.
Validation of new ISUP grading for predicting BCR

Methods

Data source

After institutional review board approval, data were extracted from a prospectively maintained Canadian database of patients who underwent RARP by one of two high-volume, fellowship-trained surgeons in two large academic Canadian centres, Hôpital du Sacré Coeur de Montréal and Hôpital Saint Luc du Centre Hospitalier de l’Université de Montréal, between 2006 and 2016.

Study population

A total of 621 patients diagnosed with adenocarcinoma of the prostate and treated with RARP between 2006 and 2016 were identified. BCR was defined as two consecutive prostate-specific antigen (PSA) values of ≥0.2 ng/dl or the use of salvage external beam radiation therapy and/or salvage androgen-deprivation therapy. Preoperative biopsy grades, as well as postoperative pathological grades were compared to BCR post-RARP. Clinical and pathological GS were categorized either according to the TGG of 6, 7, and 8–10, or according to the novel GGG: 1 (6), 2 (3+4), 3 (4+3), 4 (8), and 5 (9–10).

Covariates

For the clinical Gleason Grade (CGG), covariates included age at diagnosis, body mass index (BMI), clinical stage, PSA, prostate volume, and percentage of positive cores. For the pathological Gleason Grade (PGG), covariates included age at diagnosis, BMI, presence of extracapsular extension (ECE), surgical margin (SM) status, and lymph node status.

Statistical analyses

The Kaplan-Meier method was used for BCR analyses and the log-rank test to compare survival between groups. Univariable and multivariable Cox regression analyses were performed to test the impact of different GGG strata on BCR. In the clinical GGG, multivariable analyses were adjusted for preoperative PSA and clinical stage (T1, T2, or T3/4). In the pathological GGG, multivariable analyses were adjusted for preoperative PSA, surgical margin status, and pathological stage (pT2, pT3a, pT3b, or pT4). Harrell’s C-index was used to assess the discriminatory ability of four possible Gleason grading models: the TGG model (6, 7, and 8–10), the TGG and two separate groups for grade 7 model (6, 3+4, 4+3, and 8–10), the TGG and two separate groups for grade 8–10 model (6, 7, 8, and 9–10), and the novel GGG. All statistical tests were two-sided with a level of significance set at p<0.05. Analyses were performed using the R software environment for statistical computing and graphics (version 3.3.0).

Results

Table 1 summarizes the perioperative baseline patient and tumour characteristics. Overall, mean age was 60.34 years (95% confidence interval [CI] 59.83–60.85) with a mean followup of 27.9 months. For all 621 patients treated with RARP, the observed three-year and five-year BCR-free survival was 86.7% and 80.7%, respectively. In Kaplan-Meier analyses, two-year BCR-free survival rates for clinical GGG 1–5 were 98.3% (95% CI 96.4–100), 89.3% (95% CI 84.9–93.9), 86.9% (95% CI 78.7–96), 80.6% (95% CI 67.2–96.6), and 65.3% (95% CI 44.3–96.1), respectively (Fig. 1). Two-

<table>
<thead>
<tr>
<th>Table 1. Baseline characteristics of prostate cancer patients undergoing robotic-assisted radical prostatectomy between 2006 and 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>Mean age, years (95% CI)</td>
</tr>
<tr>
<td>Body mass index, kg/m² (95% CI)</td>
</tr>
<tr>
<td>Prostate volume, cc (95% CI)</td>
</tr>
<tr>
<td>Mean preoperative PSA, ng/ml (95% CI)</td>
</tr>
<tr>
<td>Mean followup, months</td>
</tr>
<tr>
<td>BCR rate, % (n)</td>
</tr>
</tbody>
</table>

BCR: biochemical recurrence; CI: confidence interval; ECE: extracapsular extension; LFU: lost at followup; PSA: prostate-specific antigen; PSM: positive surgical margins; SVI: seminal vesicle invasion.
year BCR-free survival rates for pathological GGG 1–5 were 95% (95% CI 88.3–99.3), 93.4% (95% CI 92.5–97.6), 82.4% (95% CI 70–94.6), 70.5% (95% CI 54.6–93.4), and 61.5% (95% CI 43.6–87.8), respectively (Fig. 2). In Kaplan-Meier analyses for the clinical GGG, statistically significant differences in BCR rates were observed between GGG 1 and GGG 2 (p<0.001), as well as between GGG 4 and GGG 5 (p=0.02). In the pathological GGG, statistical significant differences in BCR rates were identified between GGG 2 and GGG 3 (p<0.001). Furthermore, multivariable Cox regression analyses (Table 2), adjusted for preoperative PSA, surgical margin status, and pathological stage (pT2, pT3a, pT3b or pT4), revealed a three-fold higher hazard ratio (HR) for GGG 3 compared to GGG 2 (HR 2.0; CI 1.42–6.29; p<0.001). However, no difference was found between HR for GGG 4 and GGG 5 (HR 1.28; CI 0.52–3.11; p=0.92).

The univariate analyses Harrell’s C-index showed a higher discriminatory ability of the novel GGG when compared to the TGG in both clinical and pathological GGG (Table 3). Similarly, the novel GGG appears to have a stronger discriminatory ability in the multivariable analyses for the pathological GGG. However, in the clinical GGG, the multivariable analyses demonstrated a better discriminatory ability in the TGG and two separate groups for grade 8–10 model (6 vs. 7 vs. 8 vs. ≥9) rather than the novel GGG model. Moreover, a C-index increase of 0.002 from the TGG model to the TGG and two separate groups for grade 8–10 was noted in univariate analysis. Similarly, a small C-index increase of 0.005 was observed in multivariate analysis of the PGG classification for these groups. The latter may be due to the low prevalence of GS 9–10, constituting only 2.7% of the study cohort.

![Fig. 1. Kaplan-Meier analysis for biochemical recurrence (BCR)-free survival following robotic-assisted radical prostatectomy (RARP) stratified by clinical Gleason grade groups. Black solid line: Gleason score 6, grade group 1. Dotted line: Gleason score 3+4, grade group 2. Dark grey solid line: Gleason score 4+3, grade group 3. Light grey solid line: Gleason score 8, grade group 4. Dashed line: Gleason score 9 and 10, grade group 5.](image1)

![Fig. 2. Kaplan-Meier analysis for biochemical recurrence (BCR)-free survival following robotic-assisted radical prostatectomy (RARP) stratified by pathological Gleason grade groups. Black solid line: Gleason score 6, grade group 1. Dotted line: Gleason score 3+4, grade group 2. Dark grey solid line: Gleason score 4+3, grade group 3. Light grey solid line: Gleason score 8, grade group 4. Dashed line: Gleason score 9 and 10, grade group 5.](image2)
Table 3. Discriminatory ability of the four Gleason grading classifications

<table>
<thead>
<tr>
<th>Clinical Gleason grade</th>
<th>Pathological Gleason grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Univariate</td>
</tr>
<tr>
<td>TGG (6 vs. 7 vs. ≥8)</td>
<td>0.671</td>
</tr>
<tr>
<td>TGG and 2 separate groups for grade 7 (6 vs. 3+4 vs. 4+3 vs. ≥8)</td>
<td>0.886</td>
</tr>
<tr>
<td>TGG and 2 separate groups for grade 8–10 (6 vs. 7 vs. 8 vs. ≥9)</td>
<td>0.673</td>
</tr>
<tr>
<td>Novel GGG (6 vs. 3+4 vs. 4+3 vs. 8 vs. ≥9)</td>
<td>0.692</td>
</tr>
</tbody>
</table>

GGG: Gleason grade groups; TGG: traditional three-tiered Gleason grading.

Discussion

Since its introduction in 1966, the GS has been the strongest predictor of cancer outcomes in PCa. After multiple revisions, including the 2005 ISUP consensus, the original GS evolved to a traditional three-tiered Gleason grading (TGG). Due to the heterogeneity of the TGG strata, specifically, the lack of distinction between GS 7 (3+4 and 4+3) and GS 8–10, in 2013, Pierorazio et al proposed a novel five-tiered GGG. After analyzing 7869 radical prostatectomies (RP) specimens between 1982 and 2011, they concluded that the ISUP groups had a better discriminatory ability compared to the TGG. Their five-year BCR-free survival rates for men with ISUP 1–5 at RP was 96.6, 88.1, 69.7, 63.7, and 34.5%, respectively (p=0.001). Separation of GS 7 into two prognostic groups was further warranted given the favourable prognosis of Gleason 3+4 (GGG 2, two-year BCR-free survival of 93.6%), whereas Gleason 4+3 carcinomas (GGG 3) behaved more similarly to GS 8 (ISUP 4).

Since 2014, the ISUP GGG has been externally validated in nine studies, including, one multi-institutional review, two European and two American population-based cohorts, as well as in two single-centre cohorts. All studies demonstrated significant differences between the five GGG. Furthermore, they confirmed that better prognostication was accomplished with distinction between GS 3+4 and 4+3 instead of GS 7. For example, He et al found that each Gleason grading strata approximately doubled the risk for PCa-specific mortality in 331 320 PCa patients that underwent RP, radiotherapy (RT), or other treatments between 2006 and 2012. Similarly, in a large North American population-based cohort of 91 565 patients, Pompe et al observed that the eight-year PCa-specific mortality-free survival rates differed significantly between the five GGG. The study showed that the eight-year PCa-specific mortality-free survival rates for pathological GGG 1–5 were 99.5% (95% CI 99.4–99.6), 99.1% (95% CI 98.9–99.2), 97.4% (95% CI 97.9–97.9), 95.2% (95% CI 94.5–96), and 85.8% (95% CI 84.4–87.2), respectively (p<0.001).

To the best of our knowledge, this is the first study that aims to externally validate the novel ISUP GGG in a contemporary Canadian cohort treated with RARP. Several of our findings were noteworthy.

First, two-year BCR-free survival rates for clinical GGG 1–5 were 98.3, 89.3, 86.9, 80.6, and 63.3% (Fig 1) and two-year BCR-free survival rates for pathological GGG 1–5 were 95, 93.4, 82.4, 70.5, and 61.5%, respectively (Fig 2). These findings are consistent with Pierorazio et al, who noted that two-year BCR-free survival rates for men with clinical GGG 1–5 were 97.1, 90.6, 79.9, 70.9, and 51.5% (p<0.001), and that two-year BCR-free survival rates for pathological GGG 1–5 were 98.8, 93.6, 85.6, 73.7, and 58.5%, respectively (p<0.001).

Second, we observed statistically significant BCR rates difference between GGG 1 and GGG 2 (p<0.001), and GGG 4 and GGG 5 (p=0.02) in the clinical GGG, and between GGG 2 and GGG 3 (p<0.001) in the pathological GGG. Other reports, such as Epstein et al, noted that five-year BCR-free survival rates in 20 845 North American RP and 5501 RT patients were 96%, 88%, 63%, 48%, and 26% for GGG 1–5, respectively. Similar results were also reported by Pierorazio et al and Pompe et al, among others.

Third, multivariable analyses revealed a three-fold higher HR for GGG 3 compared to GGG 2 (3.44 vs. 1.10; p<0.001). However, no difference was found between the HR for GGG 4 and GGG 5 (4.18 vs. 4.74; p<0.013). Similarly, Pompe et al found that the GGG yielded a 1.5-fold or greater HR differences between GGG 2 and 3, and Epstein et al found that the HR for PCa progression in GGG 2 and GGG 3 relative to GGG 1 was 2.2 and 7.3, respectively. Unlike our results, both groups found two-fold or greater HR differences between GGG 4 and 5. This difference in results may be due to a low sample size of GGG 4 (6.9%) and 5 (2.7%) in our cohort.

Finally, the Harrell’s C-index demonstrated a higher discriminatory ability of the novel GGG compared to the TGG model in the univariate analyses for both clinical (0.692 vs. 0.671; C-index delta 0.021) and pathological GGG (0.692 vs. 0.647; C-index delta 0.045). Similarly, the novel GGG compared to the TGG model in multivariable analyses for the pathological GGG revealed a higher discriminatory ability of the novel GGG (0.808 vs. 0.790; C-index delta 0.018). Interestingly, the Harrell’s C-index showed a higher discriminatory ability of the TGG with separate groups for GS 8 and GS 9–10 compared to the TGG model in the multivariable analyses for the clinical GGG (0.745 vs. 0.740; C-delta 0.005). This finding solidifies the notion that GGG 5...
has worse outcomes than GGG 4. Tsao et al., among others, found similar results, showing that patients treated with RP with a GGG 5 had 1.74 increased risk of death compared to GGG 4 (HR 1.74; 95% CI 1.15–2.65).

Our study has numerous strengths that distinguish it from other studies. First, to the best of our knowledge, we are the first to have validated the ISUP GGG classification in a Canadian cohort treated with RARP. Second, we extracted our information from a prospectively maintained Canadian database, and therefore, did not have the limitations associated with large retrospective databases, such as missing information and coding errors. Third, even though we did not have a central pathology review, our patients were only treated at two different hospitals and, as such, we were able to reclassify patients according to their GS by matching the modified new score to the corresponding GS category.

Our study is not devoid of limitations. First, we relied on BCR-free survival rather than a more specific endpoint, such as PCA-specific mortality. Since the Gleason system was first introduced in 2005 and our methodology relies on a prospective database, this limitation was inevitable. An additional 10 years of followup would be required to be able to use PCA-specific mortality as an endpoint and our study mean followup was of 27.9 months. Second, we only had 60 patients who classified as GGG 4 and GGG 5, which may be underpowered to find statistically significant results.

Conclusion

The ISUP Gleason grading classification is an independent predictor of BCR in Canadian PCA patients treated with RARP. The ISUP Gleason grading classification showed added discriminate property, especially for the GGG 2 and GGG 3 strata.

Competing interests: Dr. Zorn has received honoraria as a lecturer/proctor for Greenlight (Boston Scientific), and participated in the WATER II clinical trial, supported by Procept Biorobotics. The remaining authors report no competing personal or financial interests.

This paper has been peer-reviewed.

References


Correspondence: Dr. Helen Davis Bondarenko, Division of Urology, Department of Surgery, Centre Hospitalier de l’Université de Montréal (CHUM), Montreal, QC, Canada; Helenxdbondarenko88@gmail.com