A nomogram to predict Gleason sum upgrading of clinically diagnosed localized prostate cancer among Chinese patients

Jin-You Wang\(^1,2,\)*, Yao Zhu\(^1,2,\)*, Chao-Fu Wang\(^1,3\), Shi-Lin Zhang\(^1,2\), Bo Dai\(^1,2\) and Ding-Wei Ye\(^1,2\)

Abstract

Although several models have been developed to predict the probability of Gleason sum upgrading between biopsy and radical prostatectomy specimens, most of these models are restricted to prostate-specific antigen screening-detected prostate cancer. This study aimed to build a nomogram for the prediction of Gleason sum upgrading in clinically diagnosed prostate cancer. The study cohort comprised 269 Chinese prostate cancer patients who underwent prostate biopsy with a minimum of 10 cores and were subsequently treated with radical prostatectomy. Of all included patients, 220 (81.8%) were referred with clinical symptoms. The prostate-specific antigen level, primary and secondary biopsy Gleason scores, and clinical T category were used in a multivariate logistic regression model to predict the probability of Gleason sum upgrading. The developed nomogram was validated internally. Gleason sum upgrading was observed in 90 (33.5%) patients. Our nomogram showed a bootstrap-corrected concordance index of 0.789 and good calibration using 4 readily available variables. The nomogram also demonstrated satisfactory statistical performance for predicting significant upgrading. External validation of the nomogram published by Chun et al. in our cohort showed a marked discordance between the observed and predicted probabilities of Gleason sum upgrading. In summary, a new nomogram to predict Gleason sum upgrading in clinically diagnosed prostate cancer was developed, and it demonstrated good statistical performance upon internal validation.

Key words Prostatic neoplasms, neoplasm staging, nomograms
incidence = 1.6 per 100,000[14]), PSA screening is not widely conducted in mainland China. Thus, greater proportions of Chinese patients are diagnosed with clinical symptoms and have advanced disease compared with their western counterparts. We developed a nomogram to predict the probability of Gleason sum upgrading in a cohort with clinically diagnosed prostate cancer. An external validation of the Chun et al. nomogram[7] was also performed.

**Patients and Methods**

**Patients**

The study population consisted of 269 assessable patients treated with radical prostatectomy and no neoadjuvant hormonal therapy at Fudan University Shanghai Cancer Center between April 2006 and May 2011. Of all the patients, 220 (81.8%) were referred to the hospital due to clinical symptoms such as urinary frequency, urgency, dysuria, and hematuria.

**Clinical and pathologic evaluation**

The indications for biopsy included elevated PSA level (>4 ng/mL), abnormal digital rectal examination, and/or suspicious findings in a radiologic examination. Clinical stage was assigned according to the 2002 TNM system. Pretreatment PSA level was measured before digital rectal examination and transrectal ultrasound (TRUS) examination. Prostate biopsies were performed under TRUS guidance, with the number of cores ranging from 10 to 16. Of the 269 men included in the study, 200 (74.3%) were biopsied with a 12-core schema. All outside biopsy slides were reviewed by a trained pathologist before the results were recorded in the database. Radical prostatectomy specimens were processed according to the Stanford protocol[8] and were also graded according to the Gleason system. Primary and secondary Gleason scores were assessed by two experienced genitourinary pathologists. Discrepancies were resolved by a joint review of the slides.

**Statistical analysis**

The data analyses in this study consisted of 3 stages. First, a logistic regression model was used to determine the probability of any Gleason sum upgrading between the biopsy and radical prostatectomy specimens. To develop a robust prediction model, we restricted the number of covariates to 20 events per variable. The PSA level, clinical T stage, primary Gleason score, and secondary Gleason score were used as predictors because they are readily available in the clinic and are well supported by evidence in the literature. For categorical variables, small units (≤5%) were merged into adjacent subgroups. Nonlinear or interaction effects were included in the final model if required to optimize the Akaike Information Criterion. A nomogram was constructed based on the results of multivariate logistic regression analysis. Second, the prediction model was further evaluated for its ability to predict significant upgrading, which was defined as a change in biopsy Gleason sum from ≤6 to ≥7 or from 7 to ≥8, as described in a previous report[9]. The predicted probability of any Gleason sum upgrading was correlated with the outcome of significant upgrading in patients with biopsy Gleason sum < 8. Third, we externally validated the nomogram proposed by Chun et al. to assess its predictive value in our set of patients clinically diagnosed with prostate cancer.

The statistical performance of the prediction model was assessed by discrimination and calibration. Discrimination was measured using the concordance index (C-index), which is similar to an area under the receiver operating characteristic curve. Bootstrap-corrected C-indexes were used to better gauge expected future predictive accuracy. Calibration was assessed by visually inspecting the plots of predicted probability of Gleason sum upgrading versus actual outcomes and by the Spiegelhalter Z-test.

Statistical tests were performed using software R 2.10.0, with a 2-sided significance level of 0.05.

**Results**

**Clinicopathologic data**

**Table 1** shows the descriptive characteristics of the entire cohort compared with those of Chun’s cohort. Overall, a higher PSA level, more advanced tumor stage, and higher Gleason score (biopsy and radical prostatectomy) were observed in our cohort. **Table 2** shows the concordance between the biopsy and prostatectomy Gleason sums. Of the 269 patients, the Gleason sum was consistent in 158 (58.7%), upgraded in 90 (33.5%), and downgraded in 21 (7.8%). We further divided the entire cohort into three groups according to the biopsy Gleason sum (≤6, 7, ≥8). The probability of Gleason sum upgrading was the highest in patients with a low biopsy grade (Gleason sum ≤6) (Figure 1). In those patients with a biopsy Gleason sum of 7, 45.5% of the Gleason sum upgrades were increased by 2 or 3 units. High-magnitude Gleason sum upgrading was less frequently observed among the patients with low-grade disease with Gleason sum upgrading (19.1%).

**Nomogram development**

The PSA level, clinical stage, and primary and secondary Gleason scores were included in multivariate analysis. Because there was a strong association between the primary and secondary Gleason scores (Table 3), the interaction term between the primary and secondary Gleason scores was added to the model for assessment. We found that the C-index was significantly improved from 0.758 to 0.789 after including the interaction term (P < 0.001). The bootstrap-corrected C-index was 0.789, indicating acceptable discrimination. The calibration plot demonstrates that the rate of predicted Gleason sum upgrading closely paralleled the observed rate (Figure 2). Based on these results, a nomogram was constructed to predict Gleason sum upgrading in a user-friendly manner (Figure 3).
To investigate the nomogram's ability to predict significant upgrading, we used it to evaluate 214 patients with a biopsy Gleason sum < 8. In this subgroup, 72 (33.6%) patients had significant upgrading. The prediction model showed a good discrimination ability for significant upgrading [C-index = 0.795, 95% confidence interval (CI) = 0.735–0.855]. Although the calibration plot shows a slight overestimation of the probability of significant upgrading, the Spiegelhalter Z-test indicated good calibration (P = 0.758) (Figure 4).

The statistical performance of the nomogram proposed by Chun et al. was assessed in our entire cohort. The C-index was 0.755 (95% CI = 0.691–0.819). The calibration plot, however, shows poor concordance between the predicted and observed probabilities of Gleason sum upgrading (P < 0.001) (Figure 5). Compared to the actual probability, the risk of underestimated or overestimated Gleason sum upgrading was more than 5%, with a wide range of predicted probability.
Figure 1. The distribution and magnitude of Gleason sum upgrading in three groups with different biopsy Gleason scores. For example, the Gleason sum was consistent in 80 (84.2%), upgraded by 1 unit in 6 (6.3%), upgraded by 2 or 3 units in 5 (5.3%), and downgraded in 4 (4.2%) patients with a biopsy Gleason sum of 7.

Table 2. Concordance between biopsy and radical prostatectomy Gleason sums

| Biopsy Gleason sum | Radical prostatectomy Gleason sum |
|--------------------|----------------------------------|
|                    | 5 | 6 | 7 | 8 | 9 | Total |
| 4–5                | 1 | 6 | 3 | 0 | 0 | 10   |
| 5                  | 0 | 51| 51| 3 | 4 | 109  |
| 6                  | 0 | 4 | 80| 6 | 5 | 95   |
| 7                  | 0 | 0 | 11| 15| 11| 37   |
| 8                  | 0 | 0 | 2 | 3 | 13| 18   |
| 9–10               | 0 | 0 | 2 | 3 | 13| 18   |
| Total              | 1 | 62| 147|27|33|269  |

Table 3. Statistical results of covariates in the prediction model for Gleason sum upgrading

| Variate                 | β (SE)         | Chi-square statistic | P    |
|-------------------------|----------------|----------------------|------|
| PSA (continuous)        | 0.022 (0.008)  | 6.62                 | 0.010|
| pGS (4–5 vs. 2–3)       | -1.540 (0.447) | 17.26                | <0.001|
| sGS (4–5 vs. 2–3)       | -3.666 (0.773) | 22.66                | <0.001|
| Clinical T stage (vs. T1c) | 3.45            |                      | 0.327|
| T2a         | 0.459 (0.457)  |                      |      |
| T2b         | -0.239 (0.417) |                      |      |
| T2c–3      | -0.276 (0.559) |                      |      |
| pGS * sGS (interaction) | 3.447 (0.932)  | 13.69                | <0.001|

PSA, prostate-specific antigen; pGS, primary biopsy Gleason score; sGS, secondary biopsy Gleason score.
Discussion

We successfully developed and validated a model to predict Gleason sum upgrading from biopsy to final pathology using four clinical variables. Our model is 78.9% accurate, and its prediction closely approximates the observed rate of Gleason sum upgrading.
between biopsy and final pathology. At the same time, the prediction model showed good discrimination ability for significant upgrading, with a C-index of 0.795. Conversely, testing of a previously published model predicting Gleason sum upgrading showed poor concordance between the predicted and observed probabilities of Gleason sum upgrading.

Biopsy upgrading has important clinical implications in terms of selecting candidates for watchful waiting, surgery, and radiotherapy approaches. The previous studies were focused on Gleason sum upgrading in low- or intermediate-grade prostate cancer. However,
this phenomenon is also meaningful in high-grade prostate cancer because Gleason score is a significant predictor of disease progression and survival even in the subgroup of patients with high-grade disease. Albertsen et al.\textsuperscript{18} have reported that men with a biopsy Gleason score of 7 to 10 had a high risk (42% to 87%) of death from prostate cancer when treated conservatively, even when cancer was diagnosed as late as 74 years of age. Brachytherapy also showed poor outcomes in patients with intermediate- or high-grade disease: Gleason score 7 tumors had an approximately 50% probability of biochemical relapse within 5 years\textsuperscript{19}. After radical prostatectomy, the incidence of extracapsular extension, seminal vesical invasion, and lymph node metastases was significantly higher in patients with high-grade disease than in patients with low-grade disease\textsuperscript{17}. The results of the RTOG 92-02 trial showed that long-term androgen deprivation and radiotherapy provided a survival advantage only in patients with Gleason score 8–10 tumors\textsuperscript{16}. Even in the subgroup with biopsy Gleason score > 7, the prognosis of prostate cancer worsens with increasing Gleason sum. The 8-year progression-free survival rates after radical prostatectomy were 40%, 32%, and 27% for patients with final Gleason sums of 8, 9, and 10, respectively (\(P = 0.043\))\textsuperscript{19}. Audenet et al.\textsuperscript{18} also reported distinctive prognostic in high-grade prostate cancer treated with surgery; the individual 5-year global PSA-free recurrence rates in patients with Gleason sums of 8 and 9 were 50% and 35%, respectively (\(P = 0.002\)). Therefore, inaccurate tumor grade estimation definitively results in improper risk attribution and treatment assignment.

The prevalence of Gleason sum upgrading in this study was 33.5%, which is within the range of previously reported values. It should be noted that the prevalence of Gleason sum upgrading was related to the number of prostate biopsy cores taken. King et al.\textsuperscript{19} found that the probability of Gleason sum upgrading was 25% when using the sextant biopsy scheme but only 13% when using the extended 10-core biopsy scheme (\(P = 0.045\)). The risk of Gleason sum upgrading was reported to decrease when more than 12 biopsy cores were taken\textsuperscript{16}, although this biopsy scheme is not widely accepted as the first choice for prostate cancer diagnosis. In our study, all patients had a minimum of 10 biopsy cores, which represents the current standard of care.

Consistent with previous studies, we found that PSA level and primary and secondary Gleason scores were independent prognostic factors for Gleason sum upgrading. The inclusion of an interaction term in the final model was motivated by the strong association between primary and secondary Gleason scores. The improved model performance confirmed the hypothesis that the effect of the secondary Gleason score depended on that of the primary Gleason score.

King et al.\textsuperscript{18} defined significant upgrading as a Gleason sum upgrading either from \(\leq 6\) to \(\geq 7\) or from \(7\) to \(\geq 8\) between biopsy and radical prostatectomy specimens. We observed that most (91.1%) of the Gleason sum upgrading in patients with low- or intermediate-grade disease was significant. When applied to this subgroup, our nomogram showed satisfactory discrimination and calibration.

To the best of our knowledge, 3 models have already been developed to predict the probability of any Gleason sum upgrading\textsuperscript{7,19}. D’Amico’s model was externally validated in a European cohort and yielded a C-index of 0.5, which is equivalent to the flip of a coin\textsuperscript{17}. The second model, published by Kulkarni et al.\textsuperscript{9}, showed a C-index of 0.71 upon internal validation. However, this model relied on 9 predictor variables, some of which are not routinely recorded in patients’ charts. The third model, developed by Chun et al.\textsuperscript{13}, demonstrated a C-index of 0.8 in the original patient cohort. It was later applied to two independent datasets from Italy and Japan for external validation\textsuperscript{12,15}. Both the development and validation cohorts showed a low PSA level (median value < 10 ng/mL) and early T stage (mostly < cT2b). The predictive accuracy was 74.9% in the Italian cohort and 79.2% in the Japanese cohort. The Chun et al. nomogram, however, was not validated in patients with clinically diagnosed prostate cancer, which had distinct characteristics. According to a multicenter study in mainland China, only 6.2% of prostate cancer patients presented with increased PSA level without other symptoms\textsuperscript{22}. Urinary symptoms were the major reason for prostate cancer diagnosis referral (75.9%) for Chinese patients. In populations without widespread PSA screening, patients who undergo radical prostatectomy tend to have higher PSA levels and more advanced tumor stage. When applied to our dataset, the accuracy of the Chun et al. nomogram in predicting Gleason sum upgrading was 75.5%; however, the calibration plot shows poor concordance between the predicted and observed probabilities of Gleason sum upgrading (\(P < 0.001\)) (Figure 5). The poor performance of the Chun et al. nomogram in our series confirmed that the application of nomograms in different patient populations should be performed with great caution.

The limitations of the present study include its single-center nature and limited sample size, which is mainly due to the low incidence of prostate cancer in mainland China. Furthermore, our nomogram was based on four readily available predictors. The accuracy of our model could potentially be improved by integrating additional variables. However, some potentially predictive factors, such as positive biopsy cores and the proportion or length of tumor involvement per biopsy core, were not recorded for patients who underwent biopsy outside of our center. Finally, we have not tested the performance of our nomogram in an external dataset. Despite these limitations, our model represents the first predictive model concerning the rate of Gleason sum upgrading between biopsy and final pathology among patients with clinically diagnosed prostate cancer.

**Conclusions**

We developed a new nomogram to predict Gleason sum upgrading in patients with clinically diagnosed prostate cancer. The model demonstrated good discrimination and calibration in internal validation. The existing nomogram constructed by Chun et al. should not be used in this setting.

**Acknowledgments**

This study was supported by the Grants for International
References

[1] Han M, Partin AW, Pound CR, et al. Long-term biochemical disease-free and cancer-specific survival following anatomic radical retropubic prostatectomy—the 15-year Johns Hopkins experience. Urology, 2001, 57:315-321.

[2] Zieba T, Oktar C, Salomons G, et al. The novel nomogram of Gleason sum upgrade: possible application for the eligible criteria of low dose rate brachytherapy. Int J Urol, 2010, 17:662-668.

[3] Jemal A, Bray F, Center MM, et al. Global cancer statistics. CA Cancer J Clin, 2011, 61:69-90.

[4] Capitanio U, Karakiewicz PI, Jeldres C, et al. The probability of Gleason score upgrading between biopsy and radical prostatectomy can be accurately predicted. Int J Urol, 2009, 16:526-529.

[5] Freedland SJ, Kane CJ, Amling CL, et al. Radical prostatectomy for pathological Gleason 8 or greater prostate cancer: influence of concomitant pathological variables. J Urol, 2002, 167:117-122.

[6] Kulkarni GS, Lockwood G, Evans A, et al. Clinical predictors of Gleason ups at surgery: implications for patients considering watchful waiting. Cancer, 2007, 109:2432-2438.

[7] Capitalio U, Karakiewicz PI, Valiquette L, et al. Biopsy core number represents one of the most important predictors of clinically significant Gleason ups at surgery in patients with low-risk prostate cancer. Urology, 2009, 73:1087-1091.

[8] Stackhouse DA, Sun L, Schroedl FR, et al. Factors predicting prostate biopsy Gleason score under grading. J Urol, 2009, 182:118-122; discussion 123-114.