

## Gleason Score and Lethal Prostate Cancer: Does 3 + 4 = 4 + 3?

Jennifer R. Stark, Sven Perner, Meir J. Stampfer, Jennifer A. Sinnott, Stephen Finn, Anna S. Eisenstein, Jing Ma, Michelangelo Fiorentino, Tobias Kurth, Massimo Loda, Edward L. Giovannucci, Mark A. Rubin, and Lorelei A. Mucci

### A B S T R A C T

#### Purpose

Gleason grading is an important predictor of prostate cancer (PCa) outcomes. Studies using surrogate PCa end points suggest outcomes for Gleason score (GS) 7 cancers vary according to the predominance of pattern 4. These studies have influenced clinical practice, but it is unclear if rates of PCa mortality differ for 3 + 4 and 4 + 3 tumors. Using PCa mortality as the primary end point, we compared outcomes in Gleason 3 + 4 and 4 + 3 cancers, and the predictive ability of GS from a standardized review versus original scoring.

#### Patients and Methods

Three study pathologists conducted a blinded standardized review of 693 prostatectomy and 119 biopsy specimens to assign primary and secondary Gleason patterns. Tumor specimens were from PCa patients diagnosed between 1984 and 2004 from the Physicians' Health Study and Health Professionals Follow-Up Study. Lethal PCa (n = 53) was defined as development of bony metastases or PCa death. Hazard ratios (HR) were estimated according to original GS and standardized GS. We compared the discrimination of standardized and original grading with C-statistics from models of 10-year survival.

#### Results

For prostatectomy specimens, 4 + 3 cancers were associated with a three-fold increase in lethal PCa compared with 3 + 4 cancers (95% CI, 1.1 to 8.6). The discrimination of models of standardized scores from prostatectomy (C-statistic, 0.86) and biopsy (C-statistic, 0.85) were improved compared to models of original scores (prostatectomy C-statistic, 0.82; biopsy C-statistic, 0.72).

#### Conclusion

Ignoring the predominance of Gleason pattern 4 in GS 7 cancers may conceal important prognostic information. A standardized review of GS can improve prediction of PCa survival.

*J Clin Oncol* 27:3459-3464. © 2009 by American Society of Clinical Oncology

### INTRODUCTION

Gleason grading<sup>1</sup> is a strong predictor of survival among men with prostate cancer (PCa).<sup>2</sup> The Gleason system, introduced in 1974,<sup>3</sup> is an architectural grading system that ranges from 1 (well differentiated) to 5 (poorly differentiated). The Gleason score (GS) is the sum of the primary and secondary patterns with a range of 2 to 10. It has long been appreciated that patients with GS  $\geq 7$  are at greater risk for extraprostatic extension and biochemical recurrence.<sup>4</sup> However, more recent evidence suggests that the application of GS has changed, leading to changes in the distribution of scores over time.<sup>5-7</sup> Further, studies using surrogate end points have shown that the prognosis of GS 7 cancers varies considerably.<sup>8-10</sup>

Analyses of GS assignment before and after the introduction of prostate-specific antigen (PSA) screening<sup>7,11,12</sup> and studies that have involved blind rereviews of original specimens,<sup>5-7,13</sup> have noted increases in GS with more contemporary readings. The change in GS over time appears to be largely due to a systematic shift in Gleason grading, which is attributed primarily to two factors.<sup>14</sup> It is widely accepted that Gleason grade from biopsy is frequently upgraded at prostatectomy, resulting in a reluctance to assign a low GS at diagnosis.<sup>15</sup> Previously undescribed benign lesions (ie, atypical adenomatous hyperplasia) were historically mistaken for Gleason 1 + 1 tumors, but contemporary scoring more often correctly classifies those lesions as benign.<sup>16</sup> Because systematic upgrading in Gleason results in improved survival for all Gleason

From the Departments of Epidemiology, Biostatistics, and Nutrition, Harvard School of Public Health; Channing Laboratory and Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital; Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA; and the Department of Pathology and Laboratory Medicine, Weill Medical College, Cornell University, New York, NY.

Submitted October 5, 2008; accepted February 9, 2009; published online ahead of print at www.jco.org on May 11, 2009.

Supported by Grants No. 5R01CA058684-13 and 5R01CA042182-20 from the National Cancer Institute; Grant No. W81XWH-05-1-0562 from the Department of Defense; and Grant No. T32 CA009001-32 from the National Research Service Award Training Program in Cancer Epidemiology and Dana-Farber/Harvard Cancer Center SPORE Career Development Award (J.R.S.). The Physicians Health Study is supported by Grants No. CA34944, CA40360, and CA097193 from the National Cancer Institute and Grants No. HL-26490 and HL-34595 from the National Heart, Lung, and Blood Institute.

Presented in part in abstract format at the Annual Meeting of the American Association for Cancer Research, April 12-16, 2008, San Diego, CA.

M.A.R. and L.A.M. contributed equally to this article.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Jennifer R. Stark, ScD, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115; e-mail: stark@hsph.harvard.edu.

The Acknowledgment is included in the full-text version of this article, available online at www.jco.org. It is not included in the PDF version (via Adobe® Reader®).

© 2009 by American Society of Clinical Oncology

0732-183X/09/2721-3459/\$20.00

DOI: 10.1200/JCO.2008.20.4669

categories, the observed Gleason shift has been described as the Will Rogers phenomenon.<sup>5,7</sup>

Historically, PCa risk prediction models<sup>4,17</sup> and observational studies that have adjusted for GS utilized only the overall GS. Because numerous studies suggest that Gleason 3 + 4 tumors (ie, those where pattern 3 is most prevalent but some amount of pattern 4 is also observed) have a better prognosis than Gleason 4 + 3 tumors (ie, those where pattern 4 is more prevalent than pattern 3), contemporary clinical risk prediction now incorporates primary and secondary Gleason pattern.<sup>18</sup> However, studies that have specifically assessed differences in Gleason 3 + 4 and 4 + 3 cancers have relied on associations of Gleason pattern with other prognostic factors and biochemical progression<sup>8,10,12,19-21</sup> or, less frequently, development of metastases.<sup>8,21</sup> While biochemical recurrence is a widely used outcome for studies relating to PCa risk prediction and treatment efficacy, it is important to note that its definition varies,<sup>22,23</sup> and it is an imperfect surrogate for PCa mortality.<sup>24,25</sup> We undertook a standardized review of radical prostatectomy and needle biopsy tumor tissue samples from the Physicians' Health Study (PHS) and the Health Professionals Follow-Up Study (HPFS) PCa cohorts to assess the predictive ability of a contemporary review compared to original GS, as well as explore potential differences in GS 7 subtypes, with PCa mortality as the primary end point.

## PATIENTS AND METHODS

### Patient Cohort

The PHS<sup>26-28</sup> was initiated as a randomized, double-blind, placebo-controlled trial for the primary prevention of cardiovascular disease and cancer. The study included 29,021 healthy US male physicians age 40 to 84 years at baseline. The majority of participants were white (94%). Participants were observed through annual questionnaires to collect data on diet, health, and lifestyle behaviors, and medical history, and biannually to ascertain compliance and health end points, including PCa.

The HPFS was initiated in 1986, when 51,529 male health professionals, ages 40 to 75 years, completed a mailed questionnaire on demographic characteristics, risk factors and preventive behaviors, and diet and use of supplements. The cohort was predominantly white (> 91%). Through biennial follow-up mailed questionnaires, we update exposure information and medical events, including PCa.

In both cohorts, participants are asked to report new diagnoses of cancer on follow-up questionnaires. Subsequently, hospital records and pathology reports are requested and reviewed by study investigators. Through systematic medical record review of PCa patients, we obtain clinical and pathological data. When available, the original GS, major and minor patterns are recorded. Stage is recorded according to the TNM staging system or a modified Whitmore-Jewett classification scheme. We also observed PCa patients through questionnaires to collect information on their PCa clinical course, including PSA levels and development of metastases. Deaths were ascertained through mailings, telephone calls, and searches of the National Death Index, and cause of death is assigned after review of death certificates, information from the family, and medical records. Follow-up for mortality was more than 99% complete in the PHS and more than 98% complete in the HPFS.

For PHS patients, original specimens (blocks and hematoxylin and eosin slides) from needle biopsy or prostatectomy are requested from the diagnosing institution of every PCa patient. For HPFS patients, only prostatectomy specimens are requested. Of the 1,195 tissue blocks obtained, we excluded two patients determined not to have cancer on standardized review, as well as two patients found to have transitional cell bladder cancer. An additional 379 patients were excluded because an original GS from the same specimen type as the obtained tissue was not available. A total of 812 men with

PCa were included in the study. Blinded to the original pathology reports and clinical data, we undertook a standardized review (M.A.R., S.P., S.F.) of original hematoxylin and eosin slides from the referring hospitals and assigned a primary and secondary Gleason grade. The pathologists reviewed the slides independently, and then GSs were compared. For any patients with discrepant Gleason data between pathologists, slides were rereviewed until a consensus was reached.

### Statistical Analysis

All analyses were conducted separately for prostatectomy and needle biopsy specimens. To assess differences in original and standardized GS, we compared distributions among the men using a Wilcoxon signed-rank test. Trends in Gleason during three 4-year time periods (1985 to 1988, 1994 to 1997, and 2001 to 2004) were explored by plotting the original versus standardized scores and the best-fitting regression line through the points. To compare the ability of GS from original report and standardized review to predict PCa survival, the outcome event was defined as lethal PCa. Event dates were the date of diagnosis of bony metastases when such data were available, or the date of PCa death, otherwise. Patients who did not die of PCa and who did not develop metastases to bone were censored at time of death from other causes or the end of study follow-up (March 1, 2008). Follow-up time was calculated from the date of PCa diagnosis to the event date, or time of censoring.

Crude mortality rates were calculated within six strata defined by the original reports and standardized GS (2 to 5, 6, 3 + 4, 4 + 3, 8, and 9 to 10). Time-to-event analyses to predict lethal PCa were also conducted separately for the original and the standardized Gleason data. Cox proportional hazards models that controlled for age at diagnosis were used to estimate hazard ratios (HR) and 95% CIs by including in the model indicator variables for each stratum of Gleason.

The discrimination of models that included the GS based on the standardized review and historical grading was compared using 10-year survival as an end point. In this analysis, we contrasted two groups: men who lived at least 10 years after diagnosis without known metastases, and men who developed metastases or died of PCa within 10 years after diagnosis. To obtain C-statistics for both original and standardized scores, a 10-level ordinal variable for GS 2 to 10 (with separate codes for Gleason 3 + 4 and 4 + 3) was included in a logistic regression model. We fit crude models and models that adjusted for age, stage, and PSA at diagnosis. We obtained 95% CIs by repeating the analysis on 1,000 bootstrap samples.<sup>29</sup> All analyses were performed using SAS version 9.1.3 (SAS Institute, Cary, NC). The research protocol was approved by the institutional review board at the Harvard School of Public Health and Partners Healthcare.

## RESULTS

Selected characteristics of the study population are included in Table 1. The mean age at diagnosis was 65.5 years among the 693 patients with prostatectomy specimens and 71.2 years among the 119 patients with biopsy specimens. Most patients were diagnosed in the PSA screening era (85% of prostatectomy and 93% of biopsy specimens).

We observed a dramatic and statistically significant shift in Gleason grading on standardized review for both prostatectomy ( $P < .0001$ ) and biopsy specimens ( $P < .0001$ ). The shift in GS assignment of prostatectomy specimens is illustrated in Table 2. Among the prostatectomy patients, 171 were originally assigned a GS of 2 to 5, but on standardized review, only six patients had that assignment (all of whom were GS 5). At the upper end of the scale, 28 prostatectomy patients were originally assigned GS 9 or 10, compared to 45 on standardized review. For biopsy specimens, GS assignment of 2 to 5 decreased from 20 to 1 and GS of 9 to 10 decreased from 7 to 3.

Figure 1 illustrates the trends in Gleason grading of prostatectomy specimens during the study period. Our pathologists tended to give scores that were higher than those originally assigned from 1985

**Table 1.** Selected Characteristics of Study Population by Source of Gleason Data

| Characteristic                | Prostatectomy (n = 693) |      | Needle Biopsy (n = 119) |      |
|-------------------------------|-------------------------|------|-------------------------|------|
|                               | No.                     | %    | No.                     | %    |
| Mean age at diagnosis, years  | 65.5                    |      | 71.2                    |      |
| SD                            | 6.0                     |      | 6.4                     |      |
| Mean follow-up time, years    | 10.7                    |      | 8.4                     |      |
| SD                            | 4.0                     |      | 3.3                     |      |
| Median PSA at diagnosis       | 6.8                     |      | 7.2                     |      |
| Q1-Q3                         | 4.9-10.0                |      | 5.2-12.4                |      |
| Pathological tumor stage      |                         |      |                         |      |
| T2                            | 505                     | 73.9 | —                       | —    |
| T3/T4                         | 162                     | 23.7 | —                       | —    |
| N1/M1                         | 16                      | 2.4  | —                       | —    |
| Clinical tumor stage          |                         |      |                         |      |
| T1/T2                         | —                       | —    | 93                      | 93.9 |
| T3/T4/N1/M1                   | —                       | —    | 6                       | 6.1  |
| Prostate-specific antigen era |                         |      |                         |      |
| Diagnosed before 1992         | 103                     | 14.9 | 8                       | 6.7  |
| Diagnosed during/after 1992   | 590                     | 85.1 | 111                     | 93.3 |

Abbreviations: SD, standard deviation; Q1-Q3, interquartile range (quartile 1-quartile 3).

to 1988. Compared to the scores assigned in 1994 to 1997, the standardized GS at the upper end of the scale were generally upgraded from their original score. This trend appeared to persist through 2001 to 2004. When we divided the data into three time periods (pre-1994, 1994 to 2000, and 2001 to 2004), the weighted  $\kappa$  statistics for concordance between original and standardized scores increased in each subsequent time period (from 23% to 33% to 44%), while the percent of patients with original GS that were lower than the standardized scores simultaneously decreased (from 73% to 55% to 39%).

A total of 53 lethal PCa events, which included the development of bony metastases (n = 11) and PCa-specific deaths (n = 42), occurred during the study period. Of the 241 total patients (biopsy and prostatectomy) assigned GS 2 to 5 or 6 on standardized review, none developed lethal PCa. Among the 693 prostatectomy patients, 37 developed lethal PCa, with postdiagnostic survival time ranging from 0.1 to 21.1 years (median, 10.4). As presented in Table 3, crude cancer

mortality rates increased in each consecutive stratum of standardized GS for the prostatectomy specimens.

Because no lethal events occurred among prostatectomy cases scored as GS 2 to 6 in the standardized review, we designated 3 + 4 as the reference group when estimating HRs for lethal PCa. After adjusting for age at diagnosis, rates of lethal PCa for prostatectomy specimens assigned as GS 4 + 3, 8, and 9 to 10 on standardized review were significantly higher than GS 3 + 4 (Table 3), and rates increased with each level of Gleason. Among those with an prostatectomy specimen, patients with a standardized GS of 4 + 3 were 3.1 times more likely to develop lethal PCa than patients with 3 + 4 (95% CI, 1.1 to 8.6), while patients with a standardized GS of 9 to 10 were more than 19 times more likely to develop lethal PCa compared to patients with GS of 3 + 4 (HR, 19.1; 95% CI, 7.4 to 49.2). Using the original GS from prostatectomy, an elevated rate of lethal PCa was also found when we compared GS 4 + 3 versus 3 + 4 (HR, 2.4; 95% CI, 1.0 to 5.6). Patients with prostatectomy specimens assigned GS 9 to 10 in the original report were only 4.9 times as likely to develop lethal PCa compared to cases with GS 3 + 4 tumors assigned by the original pathologist (95% CI, 2.0 to 11.9; Table 3). After additional adjustment for pathological stage and log (PSA) at diagnosis among the subset of 502 patients for whom data was available, the point estimates for the standardized and original GS among prostatectomy cases were similar, albeit with wider CIs. The HRs for standardized GS of 4 + 3, 8, and 9 to 10 cancers compared to 3 + 4 were 2.6 (95% CI, 0.4 to 16.0), 6.2 (95% CI, 0.9 to 44.7), and 25.9 (95% CI, 4.7 to 145.2), respectively. For the original GS, there were no deaths in the GS 2 to 5 category among the subset of men with stage and PSA data. HRs comparing each remaining stratum of original GS to Gleason 3 + 4 were 0.2 for Gleason 6 (95% CI, < 0.1 to 1.8), 3.2 for Gleason 4 + 3 (95% CI, 0.9 to 11.1), 0.5 for Gleason 8 (95% CI, 0.1 to 4.3), and 3.3 for Gleason 9 to 10 (95% CI, 0.8 to 13.7).

For the 119 patients with biopsy specimens, 16 developed lethal PCa and postdiagnostic survival ranged from 0.1 to 16.3 years (median, 8.3). Crude cancer mortality rates in each stratum of standardized GS of the biopsy specimens were as follows: GS 2 to 5, no deaths; GS 6, no deaths; GS 3 + 4, 11.0/1,000 person-years; GS 4 + 3, 18.7/1,000 person-years; GS 8, 40.6/1,000 person-years; GS 9 to 10, 98.8/1,000 person-years). According to the original GS of the biopsy specimens, crude cancer mortality rates per 1,000 person-years were

**Table 2.** Comparison of Gleason Scores From Original Source and Standardized Review Among Prostatectomy Specimens (N = 693)

| Original Source | Standardized Review |   |   |   |     |       |       |    |    |    | Total |
|-----------------|---------------------|---|---|---|-----|-------|-------|----|----|----|-------|
|                 | 2                   | 3 | 4 | 5 | 6   | 3 + 4 | 4 + 3 | 8  | 9  | 10 |       |
| 2               | 0                   | 0 | 0 | 0 | 2   | 1     | 0     | 0  | 0  | 0  | 3     |
| 3               | 0                   | 0 | 0 | 0 | 5   | 3     | 0     | 0  | 0  | 0  | 8     |
| 4               | 0                   | 0 | 0 | 0 | 21  | 6     | 1     | 1  | 0  | 0  | 29    |
| 5               | 0                   | 0 | 0 | 3 | 58  | 56    | 9     | 3  | 2  | 0  | 131   |
| 6               | 0                   | 0 | 0 | 2 | 80  | 92    | 37    | 8  | 2  | 0  | 221   |
| 3 + 4           | 0                   | 0 | 0 | 1 | 31  | 73    | 43    | 14 | 9  | 0  | 171   |
| 4 + 3           | 0                   | 0 | 0 | 0 | 1   | 15    | 22    | 12 | 5  | 0  | 55    |
| 8               | 0                   | 0 | 0 | 0 | 1   | 8     | 16    | 10 | 10 | 2  | 47    |
| 9               | 0                   | 0 | 0 | 0 | 1   | 3     | 6     | 2  | 11 | 3  | 26    |
| 10              | 0                   | 0 | 0 | 0 | 0   | 0     | 0     | 1  | 0  | 1  | 2     |
| Total           | 0                   | 0 | 0 | 6 | 200 | 257   | 134   | 51 | 39 | 6  | 693   |



**Fig 1.** Comparison of standardized and original Gleason scores assigned to prostatectomy specimens during three time periods. The Gleason scores and the best fitting line through the points are plotted in gold. The blue line represents perfect concordance between original and standardized scores.

14.0 for GS 2 to 5, 8.5 for GS 6, 10.8 for GS 3 + 4, 45.2 for GS 4 + 3, 26.6 for GS 8, and 15.9 for GS 9 to 10.

In predicting 10-year survival, we observed a marked improvement in discrimination of models that utilized strata of the standardized scores compared to original scores when modeled alone or with age. For men with prostatectomy specimens, this analysis included the 380 men who lived at least 10 years after diagnosis without known metastases and the 30 who experienced development of distant metastases or death from PCa within 10 years of diagnosis. When we modeled GS alone, the original score from prostatectomy was associated with a C-statistic of 0.82 (95% CI, 0.77 to 0.88), whereas the C-statistic was 0.86 (95% CI, 0.82 to 0.91) in the model using the standardized scores (Fig 2). When we modeled GS, age, pathologic stage, and log (PSA) at diagnosis, the C-statistics were 0.89 (95% CI, 0.82 to 0.96) for the original GS and 0.90 (95% CI, 0.83 to 0.97) for the standardized GS. For men with biopsy specimens, the comparison of model discrimination included the 37 men who lived at least 10 years after diagnosis without known metastases and the 12 who developed lethal PCa within 10 years of diagnosis. C-statistics for models of biopsy GS alone were 0.72 for the original scoring (95% CI, 0.58 to 0.86) and 0.85 for the standardized review (95% CI, 0.76 to 0.94).

## DISCUSSION

As noted in previous studies,<sup>5,7,13</sup> we observed a striking upgrading of GS in a standardized review compared to the original reading. No patients were assigned a GS of 2 on standardized review and fewer than 1% received GS 3 to 5. We observed a 61% increase in the number prostatectomy specimens assigned a GS of 9 to 10 in the standardized review. More importantly, we found that systematic changes in Gleason grading have improved the ability of GS to predict PCa mortality. Among men with GS lower than 7 assigned in a standardized review, no lethal PCa developed in more than 2,600 person-years of follow-up. Moreover, our analysis revealed that Gleason 4 + 3 cancers assigned to prostatectomy specimens have three times of the rate of PCa mortality compared with Gleason 3 + 4 tumors. Considering that capturing the amount of Gleason pattern 4 in tumors revealed remarkable differences in outcome, it is possible that the utilizing the percentage of Gleason 4 may provide further prognostic information, as suggested by a study of biochemical progression.<sup>30</sup>

One previous study found that standardized GSs of needle biopsy and transurethral resection of the prostate specimens were significantly better at discriminating indolent from aggressive disease than the original GS assigned in 1991 to 1996.<sup>13</sup> Our study confirms these findings in a population where the majority of GS data came from prostatectomy. Interestingly, including data on stage and PSA at diagnosis to our models of PCa mortality produced similar C-statistics for both original and standardized GS (0.89 and 0.90, respectively). These results suggest that stage and PSA levels are more important predictors of PCa mortality when original Gleason data are used, as would often be the case in epidemiologic studies. As PSA screening continues to increase the number of patients diagnosed with localized cancer and reduces variability in stage at diagnosis,<sup>31</sup> obtaining the most precise assessment of tumor grade will become more critical for population-based studies. Thus, both the primary and secondary Gleason patterns should be considered essential components of PCa data collection for

**Table 3.** Prostate Cancer Mortality Rates According to Prostatectomy Gleason Score

| Total Gleason Score | Standardized Review |              |     |   |               |             | Original Source   |              |     |   |               |             |
|---------------------|---------------------|--------------|-----|---|---------------|-------------|-------------------|--------------|-----|---|---------------|-------------|
|                     | No. of PCa Deaths   | Person-Years | No. | Mortality Rate (per 1,000 person-years) | Hazard Ratio* | 95% CI      | No. of PCa Deaths | Person-Years | No. | Mortality Rate (per 1,000 person-years) | Hazard Ratio* | 95% CI      |
| 2-5                 | 0                   | 64.6         | 6   | 0                                       | No deaths     | —           | 1                 | 2,178.8      | 171 | 0.5                                     | 0.1           | 0.0 to 0.5  |
| 6                   | 0                   | 2,216.0      | 200 | 0                                       | No deaths     | —           | 3                 | 2,331.5      | 221 | 1.3                                     | 0.2           | 0.1 to 0.6  |
| 3 + 4               | 6                   | 2,864.9      | 257 | 2.1                                     | 1.0           | Referent    | 12                | 1,701.8      | 171 | 7.1                                     | 1.0           | Referent    |
| 4 + 3               | 9                   | 1,419.1      | 134 | 6.3                                     | 3.1           | 1.1 to 8.6  | 9                 | 542.5        | 55  | 16.6                                    | 2.4           | 1.0 to 5.6  |
| 8                   | 7                   | 482.3        | 51  | 14.5                                    | 7.4           | 2.5 to 22.0 | 4                 | 435.3        | 47  | 9.2                                     | 1.3           | 0.4 to 4.0  |
| 9-10                | 15                  | 383.7        | 45  | 39.1                                    | 19.1          | 7.4 to 49.2 | 8                 | 240.7        | 28  | 33.2                                    | 4.9           | 2.0 to 11.9 |
| Total               | 37                  | 7,430.6      | 693 | 5.0                                     | —             | —           | 37                | 7,430.6      | 693 | 5.0                                     | —             | —           |

Abbreviation: PCa, prostate cancer.

\*From Cox proportional hazards model adjusting for age at diagnosis.

prognostic and research purposes. Given that standard PCa nomograms are most often based on a particular clinical series graded by a single pathologist or one team of pathologists,<sup>18</sup> it is reassuring that GS assigned in a variety of settings do as well as standardized scores in discriminating lethal from indolent cancers when other clinical covariates are considered.

Our findings underscore the difficulty in identifying PCa patients who should be treated with prostatectomy. All men with standardized GS 6 tumors at prostatectomy survived, but many of these men likely would have survived without intervention.<sup>32</sup> By contrast, one third of men with standardized GS 9 to 10 tumors at prostatectomy developed lethal PCa despite surgery, most likely due to micrometastases at diagnosis. If GS at diagnosis is to be used for guiding treatment decisions, our biopsy data suggest that contemporary, standardized scor-

ing is preferable to historical scoring from diagnosing institutions. All 35 men with standardized GS 2 to 6 at biopsy survived, while two of three with standardized GS 9 to 10 at biopsy developed lethal PCa; however, there were no clear trends in mortality rates according to the biopsy GS assigned by diagnosing institutions—a concerning finding given that the original biopsy GS may have factored prominently into disease management.

This study has several strengths. A large sample of patients with 20 years of follow-up allowed the use PCa mortality as a primary end point. In light of data indicating that the risks of dying from PCa and other causes 15 years after biochemical recurrence are virtually equivalent (32% v 33%),<sup>33</sup> PSA relapse cannot be viewed as a substitute for more definitive end points. The availability of data on primary and secondary Gleason pattern from both the initial review and a standardized review allowed us to utilize mortality data to make comparisons of 3 + 4 and 4 + 3 cancers, separately for needle biopsy and prostatectomy specimens. Limitations include missing data on tumor stage and PSA at diagnosis. Nevertheless, our study indicates that contemporary GS from a standardized review can markedly improve prediction of PCa-specific survival compared to the original GS. Further, our study is the first to confirm with PCa-specific mortality data that the predominance of Gleason pattern 4 in GS 7 cancers represents important prognostic information: when it comes to Gleason grading, 3 + 4 does not equal 4 + 3.

**AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST**

The author(s) indicated no potential conflicts of interest.

**AUTHOR CONTRIBUTIONS**

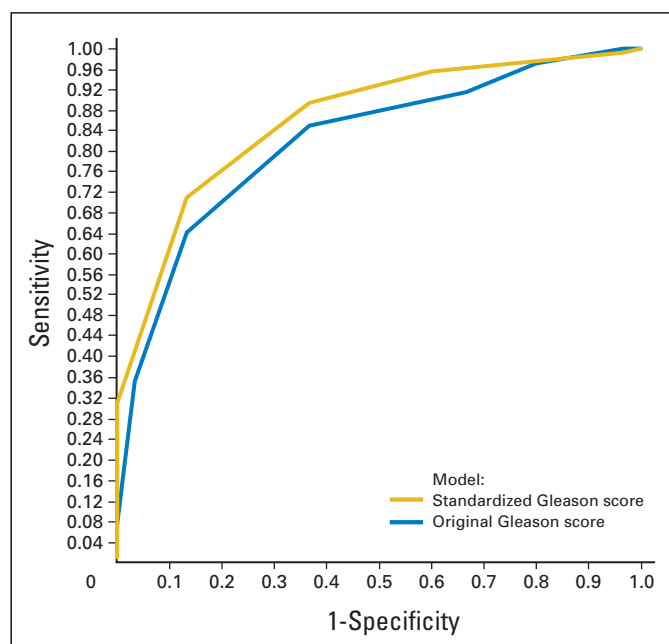
**Conception and design:** Jennifer R. Stark, Meir J. Stampfer, Lorelei A. Mucci

**Financial support:** Meir J. Stampfer, Edward L. Giovannucci

**Administrative support:** Meir J. Stampfer, Anna S. Eisenstein, Tobias Kurth

**Provision of study materials or patients:** Tobias Kurth

**Collection and assembly of data:** Jennifer R. Stark, Sven Perner, Stephen Finn, Anna S. Eisenstein, Massimo Loda, Mark A. Rubin



**Fig 2.** Discrimination of models including Gleason data from original source versus standardized review of prostatectomy specimens. C-statistics from models of 10-year survival with Gleason score categorized as a 10-level ordinal variable for Gleason scores 2 to 10 (with separate codes for Gleason 3 + 4 and 4 + 3). Standardized Gleason score C-statistic, 0.86; original Gleason score C-statistic, 0.82.

**Data analysis and interpretation:** Jennifer R. Stark, Meir J. Stampfer, Jennifer A. Sinnott, Jing Ma, Michelangelo Fiorentino, Edward L. Giovannucci, Mark A. Rubin, Lorelei A. Mucci  
**Manuscript writing:** Jennifer R. Stark

**Final approval of manuscript:** Jennifer R. Stark, Sven Perner, Meir J. Stampfer, Jennifer A. Sinnott, Stephen Finn, Anna S. Eisenstein, Jing Ma, Michelangelo Fiorentino, Tobias Kurth, Massimo Loda, Edward L. Giovannucci, Mark A. Rubin, Lorelei A. Mucci

## REFERENCES

- Gleason DF: Histologic grading of prostate cancer: A perspective. *Hum Pathol* 23:273-279, 1992
- Andren O, Fall K, Franzen L, et al: How well does the Gleason score predict prostate cancer death? A 20-year followup of a population based cohort in Sweden. *J Urol* 175:1337-1340, 2006
- Gleason DF, Mellinger GT: Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J Urol* 111:58-64, 1974
- D'Amico AV, Whittington R, Malkowicz SB, et al: Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* 280:969-974, 1998
- Albertsen PC, Hanley JA, Barrows GH, et al: Prostate cancer and the Will Rogers phenomenon. *J Natl Cancer Inst* 97:1248-1253, 2005
- Smith EB, Frierson HF Jr, Mills SE, et al: Gleason scores of prostate biopsy and radical prostatectomy specimens over the past 10 years: Is there evidence for systematic upgrading? *Cancer* 94:2282-2287, 2002
- Chism DB, Hanlon AL, Troncoso P, et al: The Gleason score shift: Score four and seven years ago. *Int J Radiat Oncol Biol Phys* 56:1241-1247, 2003
- Chan TY, Partin AW, Walsh PC, et al: Prognostic significance of Gleason score 3 + 4 versus Gleason score 4 + 3 tumor at radical prostatectomy. *Urology* 56:823-827, 2000
- Makarov DV, Sanderson H, Partin AW, et al: Gleason score 7 prostate cancer on needle biopsy: Is the prognostic difference in Gleason scores 4 + 3 and 3 + 4 independent of the number of involved cores? *J Urol* 167:2440-2442, 2002
- Rasiah KK, Stricker PD, Haynes AM, et al: Prognostic significance of Gleason pattern in patients with Gleason score 7 prostate carcinoma. *Cancer* 98:2560-2565, 2003
- Mitchell RE, Shah JB, Desai M, et al: Changes in prognostic significance and predictive accuracy of Gleason grading system throughout PSA era: Impact of grade migration in prostate cancer. *Urology* 70:706-710, 2007
- Sengupta S, Slezak JM, Blute ML, et al: Trends in distribution and prognostic significance of Gleason grades on radical retropubic prostatectomy specimens between 1989 and 2001. *Cancer* 106:2630-2635, 2006
- Berney DM, Fisher G, Kattan MW, et al: Major shifts in the treatment and prognosis of prostate cancer due to changes in pathological diagnosis and grading. *BJU Int* 100:1240-1244, 2007
- Berney DM: The case for modifying the Gleason grading system. *BJU Int* 100:725-726, 2007
- Epstein JI: Gleason score 2-4 adenocarcinoma of the prostate on needle biopsy: A diagnosis that should not be made. *Am J Surg Pathol* 24:477-478, 2000
- Bostwick DG, Chang L: Overdiagnosis of prostatic adenocarcinoma. *Semin Urol Oncol* 17:199-205, 1999
- Partin AW, Kattan MW, Subong EN, et al: Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer: A multi-institutional update. *JAMA* 277:1445-1451, 1997
- Stephenson AJ, Kattan MW: Nomograms for prostate cancer. *BJU Int* 98:39-46, 2006
- Kang DE, Fitzsimons NJ, Presti JC Jr, et al: Risk stratification of men with Gleason score 7 to 10 tumors by primary and secondary Gleason score: Results from the SEARCH database. *Urology* 70:277-282, 2007
- Khoddami SM, Shariat SF, Lotan Y, et al: Predictive value of primary Gleason pattern 4 in patients with Gleason score 7 tumours treated with radical prostatectomy. *BJU Int* 94:42-46, 2004
- Lau WK, Blute ML, Bostwick DG, et al: Prognostic factors for survival of patients with pathological Gleason score 7 prostate cancer: Differences in outcome between primary Gleason grades 3 and 4. *J Urol* 166:1692-1697, 2001
- Amling CL, Bergstralh EJ, Blute ML, et al: Defining prostate specific antigen progression after radical prostatectomy: What is the most appropriate cut point? *J Urol* 165:1146-1151, 2001
- Freedland SJ, Sutter ME, Dorey F, et al: Defining the ideal cutpoint for determining PSA recurrence after radical prostatectomy: Prostate-specific antigen. *Urology* 61:365-369, 2003
- D'Amico AV, Moul JW, Carroll PR, et al: Surrogate end point for prostate cancer-specific mortality after radical prostatectomy or radiation therapy. *J Natl Cancer Inst* 95:1376-1383, 2003
- Pound CR, Partin AW, Eisenberger MA, et al: Natural history of progression after PSA elevation following radical prostatectomy. *JAMA* 281:1591-1597, 1999
- Hennekens CH, Buring JE, Manson JE, et al: Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *N Engl J Med* 334:1145-1149, 1996
- Final report on the aspirin component of the ongoing Physicians' Health Study: Steering Committee of the Physicians' Health Study Research Group. *N Engl J Med* 321:129-135, 1989
- Christen WG, Gaziano JM, Hennekens CH: Design of Physicians' Health Study II: A randomized trial of beta-carotene, vitamins E and C, and multivitamins, in prevention of cancer, cardiovascular disease, and eye disease, and review of results of completed trials. *Ann Epidemiol* 10:125-134, 2000
- Efron B, Tibshirani R: *An Introduction to the Bootstrap*. New York, NY, Chapman & Hall, 1993
- Stamey TA, McNeal JE, Yemoto CM, et al: Biological determinants of cancer progression in men with prostate cancer. *JAMA* 281:1395-1400, 1999
- Stephenson RA: Prostate cancer trends in the era of prostate-specific antigen: An update of incidence, mortality, and clinical factors from the SEER database. *Urol Clin North Am* 29:173-181, 2002
- Bill-Axelsson A, Holmberg L, Filen F, et al: Radical prostatectomy versus watchful waiting in localized prostate cancer: The Scandinavian prostate cancer group-4 randomized trial. *J Natl Cancer Inst* 100:1144-1154, 2008
- Bianco FJ Jr, Scardino PT, Eastham JA: Radical prostatectomy: Long-term cancer control and recovery of sexual and urinary function ("trifecta"). *Urology* 66:83-94, 2005