Changes in Gleason Scores for Prostate Cancer: What Should We Expect From a Measurement?

Jorge Ramos, DO, Edward Uchio, MD,†‡ Mihaela Aslan, PhD,*§ and John Concato, MD, MS, MPH*§||*

Background: Men diagnosed with prostate cancer receive therapy based on various clinical characteristics, including the Gleason score, a measurement (range, 2–10) describing a tumor's histological appearance. An upward shift has occurred in the distribution of Gleason scores during the past decade; this change was influenced by reports suggesting that lower scores (range, 2–4) should not be assigned to biopsy specimens.

Methods: We (1) compared Gleason scores from 1994–1995 and 2004–2005 at the same institution, (2) reviewed representative articles examining changes in Gleason scores during the last 2 decades, and (3) assessed the implications of a change in histological measurements.

Results: Among men diagnosed with prostate cancer at VA Connecticut, Gleason scores 2 to 4 were reported for 11.4% (19/167) of specimens in 1994–1995 but only 0.4% (1/260) of specimens in 2004–2005; this difference persisted after adjusting for age, clinical stage, and prostate-specific antigen ($P < 0.001$). Similar results were evident in previous publications on this topic. A change in criteria for a clinical measurement may have unintended consequences, including problems of inconsistency across "time" and "place."

Conclusions: Recent shifts in Gleason scores have led to fewer patients being diagnosed with low-grade prostate cancer; this change can have adverse impacts in clinical care and research.

Key Words: prostate cancer, histology, Gleason score, measurements, patient-oriented research, clinical epidemiology

(*J Investig Med* 2010;58: 625–628)

Treatment for prostate cancer involves selecting therapy after evaluating patient-related and tumor-related characteristics that affect prognosis. For example, a man's age and burden of comorbidity are primary considerations when choosing a treatment. In addition, a Gleason rating provides prognostic information based on histological appearance of biopsy specimens.[1,2] This classification, based on 5 histological variants, yields an overall score ranging from 2 to 10 as the combination of a primary and secondary pattern.[1,2]

In the past decade, a shift toward higher (worse) values has occurred regarding the distribution of Gleason scores, especially among biopsy specimens. This shift was encouraged by reports suggesting that Gleason scores of 2 to 4 should "not be assigned" for biopsies or are "no longer viable entitites."[3,4] Reasons given for this approach include possible undergrading of tumors, interobserver variability of readings, and a potential adverse impact on patient care.[3]

When assessing therapeutic options for a man with prostate cancer, or when evaluating corresponding research publications, it is important to consider the unintended consequences of a measurement shifting over time. To examine specifically the issue of Gleason scores for biopsy specimens for men diagnosed with prostate cancer, we (1) used data on prostate cancer to compare Gleason scores among US veterans compiled 10 years apart at the same institution, (2) surveyed the literature to identify representative articles with a focus on changes in Gleason score during the last 2 decades, and (3) commented on the clinical and research implications of a change in histological measurements.

METHODS

In a quantitative analysis (and with institutional review board approval), we compared data from prostate biopsies obtained during routine clinical care at the VA Connecticut Healthcare System during 1994–1995 and 2004–2005. Data from 1994–1995 were obtained during an observational study of prognosis in prostate cancer,[5] and data were also collected from 2004–2005 to assess a possible shift over 10 years; all information is therefore from the prostate-specific antigen (PSA) era. Baseline characteristics of age, clinical stage, and PSA were compared with a χ^2 test for linear trend. For descriptive purposes, Gleason scores were classified into 4 groups (2 to 4, 5 to 6, 7, and 8 to 10). Our emphasis was on the association of *earlier* versus *later periods*, with *Gleason scores 2 to 4* versus "*all other*" categories. Accordingly, a Fisher exact test evaluated an unadjusted relationship of period and Gleason score, and a logistic regression adjusted the same association for age, clinical stage, and level of PSA.

In an informal literature review, a PubMed search (through July 2009) was conducted using search terms *Gleason score*, *diagnosis*, and *biopsy*. The search was limited to articles written in the English language; 3754 publications were identified. Publications with titles potentially related to changes in Gleason scores were selected for further review. After this screening process, 242 publications were retained, and the abstracts were retrieved. Six articles[6–11] reporting data on Gleason scores from biopsy specimens, from at least 2 periods, received full review of the text.

Finally, we discuss methodological principles of clinical epidemiology as they relate to the topic of Gleason scores. In particular, measurements of histology should be distinctive to allow different categories to represent different entities, and these measurements should also be consistent to allow the same category to represent the same entity.[12]

RESULTS

Quantitative Comparison

The distribution of reported Gleason scores at VA Connecticut shifted "upward" from 1994–1995 to 2004–2005 (Table 1). Specifically, a Gleason score of 2 to 4 was reported for

From the Departments of *Medicine, and †Surgery (Urology), Yale University School of Medicine, New Haven, CT; ‡Surgical Service (Urology), §Clinical Epidemiology Research Center, and ||Medical Service, VA Connecticut Healthcare System, West Haven, CT.
Received October 13, 2009, and in revised form December 11, 2009.
Accepted for publication January 5, 2010.
Reprints: John Concato, MD, MS, MPH, VA Connecticut Healthcare System, Clinical Epidemiology Research Center (151B), 950 Campbell Ave, West Haven, CT 06516. E-mail: john.concato@yale.edu.
Copyright © 2010 by The American Federation for Medical Research
ISSN: 1081-5589
DOI: 10.231/JIM.0b013e3181d4720c

J Investig Med: first published as 10.2310/JIM.0b013e3181d4720c on 15 December 2015. Downloaded from file:/ on April 20, 2024 by guest. Protected by copyright.

Copyright © 2010 American Federation for Medical Research. Unauthorized reproduction of this article is prohibited.

TABLE 1. Characteristics of Men Diagnosed With Prostate Cancer at VA Connecticut

Characteristic	1994–1995, n = 167	2004–2005, n = 260	P*
Age, yr			
50–59	3 (1.8%)	52 (20.0%)	<0.001
60–69	55 (32.9%)	80 (30.8%)	
70–79	93 (55.7%)	104 (40.0%)	
≥80	16 (9.6%)	24 (9.2%)	
Clinical stage			
I	10 (6.0%)	70 (26.9%)	<0.001
II	131 (78.4%)	182 (70.0%)	
III	13 (7.8%)	2 (0.8%)	
IV	13 (7.8%)	6 (2.3%)	
PSA level, ng/mL			
0–4.0	26 (15.6%)	36 (13.9%)	<0.001
4.1–10.0	54 (32.3%)	158 (60.8%)	
10.1–20.0	32 (19.2%)	42 (16.2%)	
>20	47 (28.1%)	21 (8.1%)	
Unknown (at diagnosis)	8 (4.8%)	3 (1.2%)	
Gleason score			
2–4	19 (11.4%)	1 (0.4%)	<0.001
5–6	62 (37.1%)	110 (42.3%)	
7	35 (21.0%)	90 (34.6%)	
8–10	51 (30.5%)	59 (22.7%)	

*P values determined by χ^2 test for linear trend, except Fisher exact test for Gleason score 2 to 4 versus all other categories; see text for details.

11.4% (n = 19) of specimens during 1994–1995 but was reported for only 0.4% (n = 1) of specimens during 2004–2005; $P < 0.001$. As expected, patients diagnosed in the later period tended to be younger and were more likely to have localized disease as well as a lower PSA. After adjusting (using logistic regression) for age, clinical stage, and PSA level, a statistically significant difference in Gleason scores persisted for VA Connecticut patients in 1994–1995 versus 2004–2005; $P < 0.001$.

Literature Review

Among 242 "screened" articles, objectives included examining the concordance between Gleason scores from biopsy and prostatectomy specimens, describing longitudinal trends of Gleason scores and various clinical characteristics, or assessing the association of Gleason scores and subsequent health-related outcomes. As another attribute, assessment of biopsy or prostatectomy specimens sometimes included readings obtained during routine clinical care compared with contemporary rereadings of the same slides.

A representative study[6] compared biopsy samples from 1991–1996 and 2002–2006 from the same institution: 17.0% of samples from 1991–1996 (n = 45/265) were scored as 2 to 4, whereas 0% of the 670 samples from 2002–2006 had these scores (Table 2). A study[7] done at a different institution found a similar shift: 12.3% (262/2123) of samples from 1994–1996 and only 0.5% (9/1854) from 1999–2001 received a Gleason score of 2 to 4 (Table 2). Comparable results were also found in another study[8] that regraded specimens from 2 periods; the authors concluded "…the apparent trend toward higher biopsy grades in part may be because of how pathologists interpret these specimens today as compared with 10 years ago."

Three other reports provided graphs showing changes in Gleason scores versus time. Pertinent comments included the following: "The [Gleason score shift] confounds retrospective series spanning the 1990s"[9]; "Our study suggests that a change in practice by the pathologist is a significant factor in this grade migration"[10]; and "These differences [in Gleason scores] should be accounted for when prediction tools or comparisons between the USA and Europe are considered."[11]

Among these reports,[6–11] the stated objectives varied as did the overlap with our current research objective. Only 1 of the reports,[11] however, conducted a multivariable analysis to compare Gleason scores across the different periods. This approach accounts for differences in patient characteristics that might affect the observed pattern of period and Gleason score—such as changes in PSA values at the time of diagnosis, perhaps related to evolving patterns of screening for prostate cancer.

Methodological Considerations

"Measurement" consists of acquiring raw data and expressing the result in a standardized format. When considering Gleason scores in prostate cancer, both of these elements may have changed during a relatively short period.

Large-bore needles have been replaced by thinner (18-gauge) biopsy needles, and the number of biopsies per patient has generally increased. These changes, however, would not necessarily affect the distribution of Gleason scores. The proportion of specimens obtained via transurethral resection of the prostate for benign disease has decreased, but the corresponding impact on the distribution of Gleason scores is also likely to be modest. Of greater significance, judgments used in interpreting raw data have changed. The Gleason system was designed to yield a 2- to 10-point score, but recommendations to limit the spectrum to 5 to 10 constrain the range of results, and a formal "modified instrument" has not been validated.

The ability to distinguish among the 5 classifications (ratings) is an inherent challenge of the Gleason system. As a new challenge, the revised approach may not be applied in a

TABLE 2. Representative Articles Demonstrating an Upward Shift in Gleason Scores

Reference	Earlier Period Gleason Score = n (%)	Later Period Gleason Score = n (%)
Rajinikanth et al.[6]	1991–1996 (n = 265)	2002–2006 (n = 670)
	2–4 = 45 (17.0%)	2–4 = 0 (0%)
	5–6 = 148 (55.8%)	5–6 = 423 (63.1%)
	7 = 48 (18.1%)	7 = 193 (28.8%)
	8–10 = 24 (9.1%)	8–10 = 54 (8.1%)
Sengupta et al.[7]	1994–1996 (n = 2123)	1999–2001 (n = 1854)*
	2–4 = 262 (12.3%)	2–4 = 9 (0.5%)
	5–6 = 1284 (60.5%)	5–6 = 1409 (76.0%)
	7 = 468 (22.0%)	7 = 358 (19.3%)
	8–10 = 109 (5.1%)	8–10 = 78 (4.2%)

*The original citation[7] includes an inconsistency for n = 43 missing Gleason scores; the percentages reported are based on numbers appearing in the publication for each Gleason category.

© 2010 The American Federation for Medical Research

J Investig Med: first published as 10.2310/JIM.0b013e3181d4720c on 15 December 2015. Downloaded from file:/ on April 20, 2024 by guest. Protected by copyright.

Copyright © 2010 American Federation for Medical Research. Unauthorized reproduction of this article is prohibited.

similar manner in different laboratories or even reproducibly in the same laboratory over time. (Although this problem also existed for the original Gleason scale, the magnitude of variability would tend to be greater when changes are made in an established procedure.) Thus, the recent shift in scoring introduced new challenges in ensuring consistency for readings.

DISCUSSION

Our quantitative analysis, as well as articles identified in our literature review, confirms the virtual disappearance of low-grade Gleason scores on biopsies when men are diagnosed with prostate cancer. This "Gleason shift" is due largely to the way specimens are interpreted and classified by pathologists.[8,10,13,14] Our observations and opinions are not entirely new, but we wish to bring attention to potential problems caused by an abrupt change in criteria for a clinical measurement.

As an obvious problem that could arise due to this change, the upgrading of Gleason scores artificially increases the apparent aggressiveness of tumors, even if the underlying biology of the disease remains constant. In the PSA era—when small ("early") tumors are more likely to be detected—well-differentiated tumors (ie, the lower end of the original Gleason scale) should arguably receive more attention and finer distinctions rather than be collapsed into scores 5 or 6 by the recent recommendations.[3,4] In particular, screening with PSA would tend to increase the proportion of low-grade tumors, based on a lead-time effect.[15] The decrease in Gleason scores 8 to 10 (Table 1) is consistent with such a trend toward more benign disease, yet the increase in the proportion of Gleason scores 5 to 7 is consistent with a combining of Gleason scores 2 to 4 into that middle range of scores (designated previously as "moderately differentiated").

As another potential problem, the impact of screening, as well as the effectiveness of therapy, becomes more difficult to discern in research studies when comparability of Gleason scores is unstable across periods. Similarly, clinical decisions regarding secondary treatment become more complicated, with a need for reinterpretations of Gleason scores obtained at diagnosis, based on when biopsy specimens were obtained.

The initial arguments in favor of an upward shift in Gleason scores typically used informal reasoning. For example, a concern was raised that "clinicians may assume that low-grade cancers on needle biopsy do not need definitive therapy"[3]—with an upward shift providing an indiscriminate counterbalance against possible undertreatment. Although speculation was made that "little harm will be done by assigning [a Gleason score 2–4 tumor to] a Gleason score 5 or 6 as proposed by this editorial,"[3] a specific concern is that potentially unnecessary treatments for indolent disease may expose patients to complications, including impotence and urinary incontinence. Ultimately, a patient and his physician are best served by "unmodified" information, with accompanying uncertainty part of the decision-making process.

Dr. Gleason himself commented on the topic of evaluating biopsies for prostate cancer, pointing out that "…undergrading [of biopsy specimens] is not a failure of histological grading itself."[16] Acknowledging that prostatectomy specimens provide valuable information, Gleason made the obvious point that such information is not available when a decision is made to resect the prostate. He noted: "Empirical correlations with survival and other clinical and laboratory observations have demonstrated repeatedly that the biopsy scores are clinically useful and can be accepted and used at face value."[16] After a shift in scores, an entire generation of studies would be needed to validate a revised approach.

Our quantitative analysis is based on data from a single site, our review of articles was not a systematic review or a meta-analysis, and our editorial comments are focused mainly on issues of measurement. In addition, we did not consider broader issues, such as whether realities of the medicolegal environment might pressure pathologists to avoid reporting lower Gleason scores. Other modifications in guidelines[17] for Gleason grading of prostate cancer are also beyond the scope of this report, such as defining the Gleason score as the most common histological pattern and the highest grade pattern (eg, a 3 + 4 = 7 pattern with a tertiary score of 5 is now graded as 3 + 5 = 8) or grading cribriform glands as a "4." Finally, the current work did not seek or provide direct evidence of changes in the management of prostate cancer based on the upward shift in Gleason scores. Despite these limitations, our study addressed an important issue and was done in the PSA era, and the quantitative analysis adjusted for other clinical factors that could have explained an upward shift in Gleason scores. Our main purpose was to bring attention to overarching issues related to a common clinical attribute of men with prostate cancer.

In summary, the relatively recent change in assessing and reporting Gleason scores is problematic. It is likely "too late" to reverse this change in clinical practice, although it was implemented without adequate consideration. At the very least, however, researchers and clinicians should be aware of the underlying methodological issues, to interpret data appropriately and to help provide optimal patient care.

ACKNOWLEDGMENTS

The authors thank Karen Anderson, Donna Cavaliere, John Ko, and Diane Orlando for assistance with data collection, database management, and manuscript preparation.

REFERENCES

1. Gleason DF. Classification of prostatic carcinomas. *Cancer Chemother Rep*. 1966;50:125–128.
2. Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J Urol*. 1974;111:58–64.
3. Epstein JI. Gleason score 2–4 adenocarcinoma of the prostate on needle biopsy: a diagnosis that should not be made. *Am J Surg Pathol*. 2000;24:477–478.
4. Berney DM. Low Gleason score prostate adenocarcinomas are no longer viable entities. *Histopathology*. 2007;50:683–690.
5. Concato J, Jain D, Uchio E, et al. Molecular markers and death from prostate cancer. *Ann Intern Med*. 2009;150:595–603.
6. Rajinikanth A, Manoharan M, Soloway CT, et al. Trends in Gleason score: concordance between biopsy and prostatectomy over 15 years. *Urology*. 2008;72:177–182.
7. 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*. 2006;106:2630–2635.
8. 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*. 2002;94:2282–2287.
9. Chism DB, Hanlon AL, Troncoso P, et al. The Gleason score shift: score four and seven years ago. *Int J Radiat Oncol Biol Phys*. 2003;56:1241–1247.
10. Ghani KR, Grigor K, Tulloch DN, et al. Trends in reporting Gleason score 1991 to 2001: changes in the pathologist's practice. *Eur Urol*. 2005;47:196–201.

© 2010 The American Federation for Medical Research

J Investig Med: first published as 10.2310/JIM.0b013e3181d4720c on 15 December 2015. Downloaded from file:/ on April 20, 2024 by guest. Protected by copyright.

Copyright © 2010 American Federation for Medical Research. Unauthorized reproduction of this article is prohibited.

11. Gallina A, Chun FK, Suardi N, et al. Comparison of stage migration patterns between Europe and the USA: an analysis of 11,350 men treated with radical prostectomy for prostate cancer. *BJU Int*. 2008;101:1513–1518.

12. Feinstein AR. *Clinical Epidemiology. The Architecture Of Clinical Research*. Philadelphia, PA: WB Saunders Co; 1985:71.

13. Albertsen PC, Hanley JA, Barrows GH, et al. Prostate cancer and the Will Rogers phenomenon. *J Natl Cancer Inst*. 2005;97:1248–1253.

14. Berney DM, Fisher G, Kattan MW, et al. Major shifts in the treatment and prognosis of prostate cancer due to changes in the pathological diagnosis and grading. *BJU Int*. 2007;100:1240–1244.

15. Concato J. What will the emperor say?: screening for prostate cancer as of 2008. *Cancer J*. 2009;15:7–12.

16. Gleason DF. Undergrading of prostate cancer biopsies: a paradox inherent in all biological bivariate distributions. *Urology*. 1996;47:289–291.

17. Epstein JI, Allsbrook WC, Amin MB, et al. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason grading of prostatic carcinoma. *Am J Surg Pathol*. 2005;29:1228–1242.

J Investig Med: first published as 10.2310/JIM.0b013e3181d4720c on 15 December 2015. Downloaded from file:/ on April 20, 2024 by guest. Protected by copyright.

© 2010 The American Federation for Medical Research

Copyright © 2010 American Federation for Medical Research. Unauthorized reproduction of this article is prohibited.