Gleason Scoring System Faces Change and Debate

By Charlie Schmidt

When Donald Gleason, M.D., died at his home in Edna, Minn., on December 28, 2008, he left an unparalleled legacy in the treatment of prostate cancer. The Gleason score, which he devised, remains the most widely accepted indicator of the potential for the disease to grow and spread. Introduced in 1966, it has become a crucial data point that dictates treatment decisions for millions of men worldwide.

But Gleason scores have been trending upward for more than 10 years, leading some experts to worry that they overstate prostate cancer and its severity. “Right now—chiefly because of how Gleason scores are interpreted—prostate cancer is grossly over-diagnosed and overtreated,” said Peter Albertsen, M.D., director of the University of Connecticut’s affiliated program in urology. “But we don’t know how questionable cases might turn out, so when in doubt, we end up treating them all.”

Clinicians now face changes to the Gleason score that could drive patients’ scores higher, possibly compounding the problem. The changes are prompted by evidence showing that even tiny areas of high-grade tumor detected with needle biopsy can significantly worsen a patient’s prognosis. Dubbed “tertiary Gleason components,” these areas reflect more advanced cancer, with a high risk of metastasis. Incorporating them into the Gleason score leads doctors and patients to consider more aggressive treatment options.

This approach has yet to be widely adopted beyond major academic cancer centers. “It’s a new approach, and these things take a long time to make their way into private practice,” explained Peter Scardino, M.D., chairman of the department of surgery at the Memorial Sloan-Kettering Cancer Center in New York. But the recommendation has sparked debate among experts over whether the revised scoring could lead to inappropriate treatment decisions.

The Nuts and Bolts

Gleason began working on the score in 1962, when he was a junior-grade pathologist at the Minneapolis Veterans Administration Medical Center. His initial system, published in Cancer Chemotherapy Reports in 1966, was based on a study sample of 280 patient biopsies. With Gleason’s method, tumor samples obtained from needle biopsy cores, or from the prostate after removal by radical prostatectomy, are graded according to how far they depart from normal cell architecture. More disorganized, or “undifferentiated,” tissues are assigned higher scores, such that pattern 1 tissues are well differentiated and nearly healthy, whereas pattern 5 tissues—at the top of the scale—are poorly differentiated. To derive the score, Gleason simply added the two most common tissue components, with the larger component presented first. Thus, a sample with 70% pattern 3 (dubbed the primary component for its larger area), and 30% pattern 4 (the secondary component), is represented as 3+4=7, whereas a sample containing 70% pattern 4 and 30% pattern 3 is represented as 4+3=7. By convention, tissues of one grade are doubled (e.g., one that is entirely grade 3 is represented as 3+3=6).

Under the Gleason system, patterns of 3 or less denote slow-growing, less worrisome cancers. Patterns 4 and 5 cause more concern. A finding of 4+3=7, which has more grade 4 than grade 3 tumor, is therefore more troubling than a finding of 3+4=7, even though the final Gleason scores are the same. Because of prostate cancer’s slow growth, patients whose Gleason patterns on needle biopsy don’t exceed grade 3 are candidates for active surveillance, or “watchful waiting.” In these cases, doctors withhold treatment on the assumption that the patient will probably die from another cause, unrelated to the cancer. The presence of grade 4 patterns or higher compels increasingly aggressive approaches, depending on the amount of advanced disease and other considerations.

Updated in 1974, with a larger sample of 1,032 men, Gleason scores ranged from a low of 1+1=2 to a maximum of 5+5=10. According to Albertsen, the initial score predated two key developments in the field: the prostate-specific antigen (PSA) test, which revealed more lower-grade tumors, and the introduction of thinner, 18-gauge biopsy needles, which produced less tissue for analysis than the thicker bore needles used previously. As a consequence, pathologists had to grade earlier stage tumors with less material than they once had. Given that—and that needle biopsy specimens were often upgraded after tissues from radical prostatectomy were made surgically available—many started erring on the side of caution. Components once assigned grades of 2 or 3 were now routinely graded as 3 and 4.

In 2005, Albertsen described this phenomenon in the Journal of the National Cancer Institute [J. Natl. Cancer Inst. 2005;97:1248–53]. “It can be hard to tell Gleason grades 1 and 2 apart from 3,” he said. “The result is that 3 is now the lowest score that we see, and it’s also the score that pathologists most often report from needle biopsy.” The steady upward shift in Gleason scores, Albertsen added, leads to a false sense of progress in the treatment of
prostate cancer. In other words, treatment looks better if it succeeds in a Gleason 6 patient than it does in a patient with a lower Gleason score, even if the cancers themselves are identical. Albertsen calls this the “Will Rogers” phenomenon, in reference to the Vaudeville comedian who joked that “when the Okies moved to California, the IQ of both states went up.”

**Adding Tertiary Components**

Gleason’s omission of the tertiary component was a reflection of the times. Coming before PSA and other early warning improvements, the score was devised when most cancers were fairly advanced and homogeneous at diagnosis. But when early-grade tumors with more heterogeneous patterns began showing up in oncology labs, researchers took note. In a seminal work, Thomas Stamey, M.D., and John McNeil, M.D., both of Stanford University School of Medicine, showed during the 1990s that the more high-grade tumor detected in the prostate, the worse the prognosis.

Jonathan Epstein, M.D., from the department of pathology at Johns Hopkins Hospital in Baltimore, built on that evidence with a retrospective study of radical prostatectomies. He found that high-grade tertiary components, which he defined as occupying 5% or less of a tumor, correlate statistically with worsening progression rates. Published in the *American Journal of Surgical Pathology* in 2000, that study showed that Gleason score 5–6 cancers—if they contained tertiary patterns of 4 or above—were more dangerous than comparable tumors without tertiary components.

Epstein has since led efforts to revise Gleason scoring accordingly. After his 2000 publication, he worked to replicate his initial conclusions with increasingly larger sample sizes. His efforts culminated in a consensus by international experts, who convened in San Antonio in 2005. Their conclusions, published that year in the *American Journal of Surgical Pathology*, with Epstein as lead author, stated that when considering radical prostatectomy specimens, clinicians should continue adding primary and secondary components while merely commenting on the presence of a tertiary component in their notes.

For needle biopsy specimens, however, which contain less tissue for analysis, the experts recommended a more conservative approach: adding the primary component to the highest grade detected, whether that is secondary or tertiary. Thus, a Gleason score of 3+4=7 with a tertiary component of 5 (which is the most common tertiary grade, according to Epstein), would be represented as 3+5=8.

**Emerging Debates**

That recommendation sparked heated debates over how the revised Gleason scores should be applied. Claiming that the consensus approach is biologically valid, Epstein cites research by Abhijit A. Patel, M.D., from Yale School of Medicine, published in the *Journal of the American Medical Association* in 2007. Patel and colleagues compared Gleason 7 cancers both with and without tertiary components and found that the risk of PSA failure, or posttreatment PSA elevations that might reflect cancer recurrence, was statistically significantly higher among the former. “These men had cancers that were comparable to Gleason 8 or above,” Epstein said. “They were very aggressive and support our view that you should add the most common and the highest grades.”

*continued on page 629*
Gleason Scoring, continued from page 623

But skeptics remain. Scardino claims that the difference in PSA recurrence between these two groups—although statistically real—amounts to just a few percentage points over 8–10 years. “I am not convinced that a Gleason 3+4 cancer with a minor component of 5 will behave exactly like a Gleason 4+4 and that it should therefore be treated similarly,” he said.

The debate crystallizes around whether Gleason 8 patients make good candidates for surgery. Experts disagree. Citing what he describes as routine policy at Johns Hopkins Hospital, Epstein said that, except for some younger patients, surgery isn’t generally appropriate for patients with Gleason 8 scores or higher, including those classified as such by the addition of tertiary components.

“These patients could have high-grade cancer with positive lymph nodes and positive margins, and they might not be curable with surgery,” he said. “What’s more, they could be vulnerable to more systemic disease, so you might want to treat with radiation or some other approach.”

Scardino responds that at other institutions, including Memorial Sloan–Kettering Cancer Center, Gleason 8 patients are seen as “excellent” candidates for surgery, whereas Albertsen insists that the jury’s still out. “We simply don’t have the evidence we need to support surgery over radiation or some other approach,” he said. “Having Gleason 8 just says you have a bad prognosis, and I would treat that cancer with everything I have at my disposal.”

On the other hand, experts do agree that adding a tertiary component to a Gleason score of 3+3 = 6, effectively rendering it a Gleason 7 or higher, has a legitimate effect on treatment decisions. “I think in those scenarios, we’d argue against watchful waiting,” Albertsen said.

Adding tertiary components to the Gleason score will probably become more common, according to experts interviewed for this article. But for that to happen, the tables and nomograms used to predict pathological stage after surgery, or prognosis after radiotherapy, will have to incorporate these patterns explicitly. “If we don’t do that, the consequence is that we’re dropping something with adverse significance,” Epstein said. “We’re underestimating severity in diagnosis by telling our patients that they’re doing better than they are.”

But others countered that with notable exceptions—particularly when it deters against watchful waiting in favor of more active treatment—adding tertiary components offers just a small amount of clinical value. “Nonetheless, I think we would all recommend using it, even if it doesn’t generally make a significant impact on patient decision making,” Scardino said.

At the same time, these experts warn that the upward trend in Gleason score classification has a significant downside with respect to assessing progress.

“The whole grading system has been artificially inflated,” Albertsen said. “And that makes it hard for us to compare historical datasets. We think we’re looking at better survival rates, but we could be looking at a statistical artifact.”

© Oxford University Press 2009. DOI: 10.1093/jnci/djp109