Stage Migration and Grade Inflation in Prostate Cancer: Will Rogers Meets Garrison Keilor

Ian M. Thompson, Edith Canby-Hagino, M. Scott Lucia

A phenomenon described by several previous investigators—an upward drift in assigned Gleason grade to prostate cancer—is elegantly addressed in the article by Albertsen and colleagues in this issue of the Journal (1). The authors asked the question: are prostate cancers currently more aggressive, or are we just labeling them as more aggressive? Comparing initial biopsy Gleason grades from a decade ago with a contemporary Gleason grade assessment of the same samples, the authors made two important observations. First, there was a substantial shift to higher tumor grades with contemporary evaluation. Second, when survival was stratified by Gleason score, all groups of patients appeared to have improved outcomes, merely by rereading of the biopsy slides and the assignment of a new, contemporary Gleason score.

A similar phenomenon could be the Will Rogers effect, a term that was first coined by Feinstein and colleagues (2); this phenomenon is also known as stage migration (3). An alternative term could be grade inflation, a term that acknowledges the national trend of higher grades, especially in university-level education. Like the observations of Albertsen et al. regarding Gleason grades, an analysis of grades given at the university level in the United States that does not control for year of graduation would suggest that we, as a nation, have increasingly brilliant college students. To our knowledge, the first person to use this term for prostate cancer was Barnett Kramer, MD, MPH, the editor of the Journal of the National Cancer Institute.

What is the cause of grade inflation in prostate cancer? A recent change in the accepted protocol for reporting of biopsy grade may be partially responsible. Prostate cancer is usually multifocal, with prostatectomy specimens often harboring six to eight discrete tumors of various grades. Prostate biopsy, performed to “map” the gland, can thus identify several different tumors. Gleason scoring, which identifies five patterns of decreasing differentiation and increasing aggressiveness, assigns a first number to the predominant pattern in a biopsy core and, if there is another pattern present, a second number to the second most common pattern (e.g., $3 + 3$). Increasingly, more than two patterns are identified in prostate biopsy samples, and some clinicians have suggested reporting a tertiary score (4). However, in recognition of the fact that the most aggressive tumor found in a prostate biopsy, even if it is the smallest component, may determine the prognosis of the patient’s tumor, the current recommendation for assigning Gleason grade is to give the predominant tumor the first score and the highest grade, among the remaining grades, the second score (5). This practice can only lead to increased Gleason scores in some patients.

A second possible cause of grade inflation has been the growing consensus that low-grade tumors should rarely (if ever) be diagnosed (4). This consensus, which has most likely arisen from analyses of Gleason score changes between biopsy and prostatectomy, acknowledges that low-grade tumors are often upgraded at prostatectomy and that the assignment of a low grade to a tumor may lead to a false impression of a biologically inconsequential tumor, potentially losing the opportunity to treat an otherwise curable and lethal tumor (6).

Is the phenomenon identified by Albertsen et al. really grade inflation (an upward shift in grade), or could it reflect reclassification? Reclassification of patients by grade will alter the precision of Gleason score in estimating outcomes. Although the average tumor grade increased from initial to contemporary assessment in the analysis by Albertsen et al., we must recognize that there is a unique bias to this analysis. The lowest Gleason score on the initial reading cannot go lower on a rereading; that is, a Gleason 2 cannot decrease when reread a decade later. Similarly, a Gleason score of 10 (the highest score) cannot increase. This phenomenon is known as spectrum bias. Notably, whereas Albertsen et al. found that all Gleason 2 and 3 tumors increased in grade, the scores of more than half of the Gleason 8–10 tumors did not increase: 72 were unchanged in grade, 68 increased in grade, and 103 decreased in grade. Despite this overall downward grading of initially high-grade tumors, the survival of men in all categories of Gleason score improved. One conclusion that can be reached when examining figure 2 in Albertsen et al. is that contemporary Gleason scoring may more accurately reflect tumor prognosis than grade assignments of a decade ago.

Why does grade inflation in prostate cancer matter? There are two reasons. First, although this methodology is utterly invalid, investigators frequently compare results of treatments between two series of patients, whether examining improvements over time with a specific therapeutic modality or making comparisons between modalities. Although we acknowledge that the only valid comparisons are in the context of randomized clinical trials, head-to-head comparisons are frequent because of the paucity of such trials in prostate cancer. If assignment of Gleason score is changing over time, then, given that tumor grade has an overwhelming impact on outcome, it is obvious that any more recent treatment will demonstrate improved outcomes due simply to grade inflation. Thus, if there are any differences in years of diagnosis of patients, such a comparison will be fatally flawed.

We are more concerned that grade inflation is a component of the more insidious phenomena of overtreatment and...
over-treatment of prostate cancer. Currently, about 50% of men in the United States have a prostate-specific antigen (PSA) test annually, and about 75% of men have had a PSA test. Despite a 3%-4% lifetime risk of prostate cancer death, more than 17% of men in the United States will be diagnosed with prostate cancer during their lifetime. By contrast, the lifetime risk of being diagnosed with prostate cancer in the 1970s was about 10%. What has fueled this dramatic increase in diagnosis? Certainly, since 1985, the primary impetus has been PSA testing. We anticipate that the rate of diagnosis will increase further after two publications resulting from the biopsies in the Prostate Cancer Prevention Trial (PCPT) showing a substantial risk of cancer at low levels of PSA (7,8). Additional factors that have increased cancer detection have been the use of automated biopsy “guns,” which have allowed a safer performance of transrectal biopsy; local anesthesia for biopsy; and the concurrent increase in the number of prostate cores obtained at the time of biopsy. Currently, instead of the two to four biopsy samples (cores) of the 1980s, 10–12 cores are most commonly obtained, with some authors recommending as many as 24 cores (9). Previously it was axiomatic that, whereas autopsy studies showed high rates of prostate cancer in men who died of other causes, prostate biopsy simply did not detect these small tumors. With the publication of the results of the PCPT, in which 24.4% of men with an initial PSA level below 3.0 ng/mL were ultimately found to have prostate cancer after 7 years of follow-up, most authorities recognized that clinically insignificant cancers are indeed found with prostate biopsy (10).

One large cohort study has found that more than 90% of men with organ-confined prostate cancer currently opt for treatment (11). With growing data that as many as five of every six men diagnosed with prostate cancer (i.e., a 3% risk of death but a more than 17% risk of diagnosis) may not need treatment and the evidence that treatment adversely affects quality of life, why is it that so many men opt for treatment? One reason may be our risk-averse society. (We put labels on to-go coffee cups that say “don’t spill this on you—this beverage is hot.”) Another reason, however, may be the application of outcomes of watchful waiting for prostate cancers of decades ago to a patient’s tumor today with its current Gleason score. For example, a common reference in counseling patients is a previous report from Albertsen (12). In that study, 767 men diagnosed with prostate cancer between 1971 and 1984 were watched without treatment. The primary determinants of risk of cancer death were age and Gleason score, with risk of cancer death for a Gleason 2–4 tumor being 4%–7% at 15 years compared with 42%–70% for a Gleason 7 tumor (12). Similar conclusions have been reached in other series (13). The application of these historic outcomes to today’s patient almost certainly leads to a greater propensity for active treatment in lieu of surveillance.

Albertsen et al. have successfully drawn our attention to the complexity of interpreting prostate cancer outcome data with their demonstration that comparisons of outcomes among different groups of patients receiving different treatments at different institutions and over different periods will be fraught with very serious errors and are fundamentally unreliable. What is the answer to this morass of data that is faced by a quarter of a million men in the United States annually? The mundane question of “Doctor, what is the best treatment for my cancer?” (a question that will not go away in the next 20 years) will be answered only through clinical trials. One of these trials was recently completed; it demonstrated improved survival and reduced risk of prostate cancer death with radical prostatectomy compared with surveil-

References


