UNDERGRADING OF PROSTATE CANCER BIOPSIES: A PARADOX INHERENT IN ALL BIOLOGIC BIVARIATE DISTRIBUTIONS

DONALD F. GLEASON*

Paradox: a situation in which relationships appear to be inconsistent or false but, in fact, are correct and true.

A number of authors have noted that the histologic grade or score of the cancer in a resected prostate is often higher than the score of the cancer in the original needle biopsy.[1-3] This is commonly interpreted as "undergrading" of the biopsy. Nevertheless, those authors usually conclude that the histologic score in the biopsy is clinically useful in directing further studies and treatment. Others believe that because of those recurrent discrepancies, grading of prostate needle biopsies is not sufficiently reliable to merit grading them at all.[4]

Substantial random variation is inevitable in any biologic data, but I wish to point out that this undergrading is not a failure of histologic grading itself. The finding of higher scores in resected specimens with low biopsy scores[3] is the unavoidable, predictable result of a paradoxical, systematic deviation from the expected correlation in this type of data. It will always prevail in any bivariate distribution with less than perfect correlation between the x and y values, that is, all biologic bivariate distributions. The problem is not a mysterious discordance between two samples of the same tumor but the mistaken anticipation of perfect correlation between two samples.

The slope constant for predicting y values from x values is rarely the simple reciprocal of the slope constant for predicting x values from y values. Given random variation, both slope constants will deviate systematically away from each other and from the mistakenly expected one-to-one correlation. The size of these systematic deviations will be inversely proportional to the size of the correlation coefficient *r*, which measures how strongly the x and y values are correlated in that set of data.

These deviations guarantee that the observed y values will, on average, be higher than expected at the low end of the range of x values and, conversely, the y values will be lower than expected for the high x values. Midrange, the deviations will decrease and become zero at the mean x, mean y point.

These deviations prevail with distinctly different data pairs, such as heart weight versus body length. The deviations are simply more obvious if the x and y values involve the same variables and values, such as the histologic score of the same tumor in two different samples. One expects a one-to-one correlation but does not find it.

The explanation for these deviations can be ferreted out of any textbook on biostatistics but is rarely emphasized to apply directly to this problem. It was demonstrated very clearly in my old college text, from which Figure 1 is reproduced.[5] The figure shows the scatter of x and y data in the statistician's two-dimensional "surface of distribution." That distribution is plotted on coordinates of relative deviations—units of standard deviation (not labeled)—rather than actual values, to simplify and render more nearly universal the concept being demonstrated. The rectangular crosshatching implies only that both the x and y values are subject to random variation, which causes the x and y data pairs to be distributed in an ellipse.

The shapes of the ellipses vary with the strength of the correlation between the x and y values, which can be expressed by the correlation coefficient *r*, which can range from 0 to 1.0. At the extremes, an *r* value of 0 denotes the complete absence of any correlation; the surface of distribution is a circle (Fig. 1A). An *r* value of 1.0 describes perfect correspondence of x and y, on a straight line through zero (Fig. 1F). Such perfect correlation might be the expected correlation, but it is virtually never seen in biologic systems and is very rare even in the more precise physical sciences.

For correlation coefficients between 0 and 1.0, the surface of distribution is an ellipse with its

*Retired

From the Department of Laboratory Medicine and Pathology, U.S. Veterans Administration, Minneapolis, Minnesota

Reprint requests: Donald F. Gleason, M.D., 6211 Logan Avenue South, Minneapolis MN 55423

Submitted: October 2, 1995, accepted: November 2, 1995

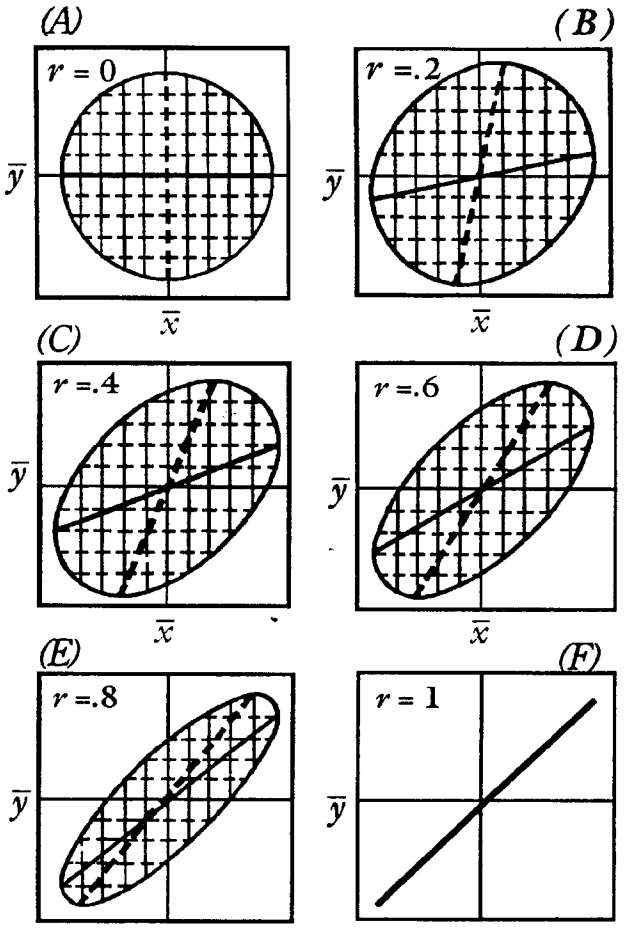

FIGURE 1. *Regression slopes for x and y change as correlation coefficients change, leading to dissociation of the observed and predicted (expected) values. See text. (Reprinted with permission from Treloar.[5])*

long axis congruent with the perfect $r = 1.0$ slope. However, within the ellipse, the lines of best fit for predicting y values associated with randomly identified x values have flatter slopes than the expected relationship, with y values higher than expected at the low end of the ranges and lower than expected at the high end of the ranges (regression of y on x; solid line in Fig. 1B, C, D, E). The actually observed values of y, of course, will vary randomly above and below these lines of best fit.

These deviations from expected are further exaggerated if the x and y values are restricted at either end of their range, as is the case for prostate cancer scores, which can range only from 2 to 10. The ellipse of distribution becomes very distorted but more amenable to simple logical analysis. Thus, if the tumor is score 2 in the biopsy, the tumor might be score 2, 3, 4, 5, or higher in the resected prostate, but there is no score lower than 2. The mean score of a series of resected cancers must always be higher than their biopsy score of 2. If the biopsy score is 3, the same deviations will persist but to a lesser degree. It is these differences that have been interpreted as undergrading of the biopsy in comparison with the resected specimen.

Conversely, if the biopsy score is 10, the score in the resected specimen may be lower but cannot be higher than 10. The mean score in a group of resected prostates will be smaller than their biopsy score of 10. This "overgrading" of the biopsies has gone unnoticed because fewer high-grade tumors are resected. They are often clinically extended and not resected.

Curiously, if one reassigns the resection scores to the x axis and the biopsy scores to the y axis, the systematic deviation appears to reverse itself. The biopsy scores will be higher than the prostatectomy scores in the low range and lower in the high range. This is useless information, defying the logic of the time sequence, but does illuminate the underlying paradox.

That is, we tend to consider that the x variable is the independent variable and expect the y variable to conform closely, but this is not justified. Both x and y variables may be driven by similar forces to move together, but each is also moved separately by random forces (sampling error, and interobserver and intraobserver variation in histologic grading, in the case of prostate cancers). We may accept random variation for the y variable around a perfect one-to-one correlation but forget that the x variable was also subject to random variation before we measured it. It is the random variation in both x and y that distorts the correlation into an ellipse as it deteriorates toward a no-correlation circle.

These disappointing discrepancies are inherent in the model and cannot be adjusted out. 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. Prediction of the score of the resected tumor is irrelevant to those considerations. The Gleason grading system was calibrated on the death rates of 2911 cases in the Veterans Administration Cooperative study,[6] of which about 60% were graded on needle core biopsies alone.

The histologic score and other details of the tumor in the prostatectomy specimen do provide a superior level of predictive information—surgical and pathologic staging—which are very important to the study of cancer biology and very desirable for formulating follow-up treatment. That new information is not available, of course, until the biopsy proves that cancer is present and, along with other clinical information, determines whether or not the prostate should be resected.

It is very likely that the histologic score in the larger resected specimen is better correlated with

the biologic malignancy of the tumor than the score in the smaller biopsy, but it will require a large prolonged study to prove that simple logical assumption. Then it will probably prove to be of relatively minor clinical usefulness, since the major decision (prostatectomy or not) has already been made, with the biopsy score playing an important role.

REFERENCES

1. Catalona WA, Stein AJ, and Fair WR: Grading errors in prostatic needle biopsies: relation to the accuracy of tumor grade in predicting pelvic lymph node metastases. J Urol **127**: 919–922, 1982.

2. Lange PH, and Narayan P: Understaging and undergrading of prostate cancer. Argument for postoperative radiation as adjuvant therapy. Urology **21**: 113–118, 1983.

3. Bostwick DG: Gleason grading of prostatic needle biopsies. Correlation with grade in 316 matched prostatectomies. Am J Surg Pathol **18**: 796–803, 1994.

4. Parker ME: Gleason grading (letter). Am J Surg Pathol **19**: 852–854, 1995.

5. Treloar AE: *Biometric Analysis. An Introduction.* Minneapolis, Burgess Publishing, 1951, p 192.

6. Gleason DF: Histologic grading and clinical staging of carcinoma of the prostate, in Tannenbaum M (Ed): *Urologic Pathology.* Philadelphia, Lea & Febiger, 1977, pp 171–198.