

PRESIDENTIAL ADDRESS

Truth in Labeling: You Can't Fool Mother Nature

JEROME J. DeCOSSE, MD

I SINCERELY THANK the membership for affording me the honor of being your President during this past year. It has been a year of substantial progress, thanks to the work of many. As a result, surgical oncology has become more visible. Indeed the excellent array of papers at this meeting illustrates very well the growing stature and sophistication of our field.

I would like to address some aspects of our scientific communication with one another. It is through our communications with our colleagues and the public that surgical oncology will be accepted, defined and enhanced. The growth of surgical oncology will depend not only upon our creativity, but also upon the accuracy with which we communicate information with one another.

At the present time, we are disturbed by several examples of gross distortion of data coming from scientists in some of the nation's distinguished medical centers. Presumably these acts were openly dishonest and fraudulent ones that we would uniformly condemn. Yet when they are judged in the context of the vastness of biological research and the frailty of human nature (why should doctors be different?), the comparative rarity of these events maintains our confidence.

The issue I wish to comment on is far more common than fraud. It is what Robert Ebert has called "borderline falsification." Perhaps the word "falsification" implies conscious manipulation, and I doubt that is usually the case. The action could also be labeled "data manipulation," "making the conclusion fit the hypothesis," or "data forcing." These defects may be distortions in clear thinking or unconscious, but improper, responses to

pressures of various kinds. Many of these flaws, such as failure to apply proper statistics, conclusions drawn from a sample of inadequate size, or conclusions that have no relationship to the results, are well known. Some are not so evident, some intrude on clinical decision making, and some may be more the weaknesses of a system than of a person.

On the whole, Mother Nature reveals her secrets reluctantly, but the truth once known has a compelling force of its own. That force can be slow. It is also amazing how error can be sustained. Often error is defended by volume. Instead of seeking precision and objectivity we have allowed the adversarial strategies of the law—entirely appropriate among that profession—to pervade our scientific discourse.

Much of what we hear or read turns out to have little or no value. We waste much time. Many reported studies fail the ultimate test, the ability to be reproduced. Mother Nature wins because error cannot be validated. Not only do we waste our time but, worse, for the patient with cancer, false expectations are aroused and then cruelly dashed.

What happens? True, sometimes a good idea is obscured by lack of clarity. More often a bad idea is made to look good by hyperbole. It is terribly important to recognize and discount hyperbole. A healthy cynicism is more enlightening than belief in the tooth fairy.

In many respects, science mirrors society. The media hype of our sensory world seems to be insidiously transposed into our scientific discourse. Madison Avenue is not far from the Island of Kos. There is a converse of Gresham's law in oncology, if not on the television news: in oncology, good news drives out bad news. Bad news tends either not to be assembled or not to be published. Good news gets grants. Some people have taken advantage of that point. How many of us have been asked to explain, "Why are you against your patients having the benefits of laetrile?" How few have asked, "Is it true, doctor, that laetrile doesn't work?"

There is the principle of the expanding denominator.

Presented at the Annual Meeting of The Society of Surgical Oncology, Marco Island, Florida, April 19, 1982.

From the Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, New York.

Address for reprints: Jerome J. DeCosse, M.D., Department of Surgery, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021.

Accepted for publication June 21, 1981.

This problem seems particularly evident in drug trials. Methyl-CCNU comes to mind. Often, the first sample is the best. Many cancer drugs are tested: some initially seem more effective than others by random selection of responsive patients (read luck). A paper or abstract is published giving the good news. As patients continue to accumulate, however, all enter the denominator but fewer enter the numerator, and the response rate drops. Sometimes, these later results are not published.

It is interesting that reports of totally ineffective treatments often describe a response rate. This can be looked upon as "background noise." (What is the definition of a response?) It is worth remembering that up to a century ago, the history of medical treatment was mostly the history of a placebo effect. Be careful about jumping on a bandwagon.

In a similar vein, if a study of an ineffective treatment is performed 20 times, one of 20 trials can be expected to show a difference at a P value of 0.05. Most or all the 19 negative studies will be put in the file while the 20th will be published in a preeminent journal as a positive result. Herein may reside an explanation for those few papers on the nonspecific immunotherapy of human cancer that have suggested benefit.

Another possible defect is the presence of selection bias. The clue to the wary is the ellipsis in any publication or address that should specify, "under the conditions of the study (trial) (investigation) we found. . . ." Are the assumptions that undergird the study fully stated? Results are often affected by what is omitted. If the assumptions are wrong, the conclusions are not likely to be good.

Is the sample population representative of the population about which conclusions are drawn? There is a distribution in the biological characteristics or natural history of cancer at any specific site. It has been said that it is possible to find 50 of anything in a country as large as the United States. One percent of cancer at any site will have an aberrant behavior. Conclusions about the 99% must not be drawn from the 1%. Herein resides a possible explanation for the occasional papers about a communicable nature of Hodgkin's disease.

Are the comparisons proper? Ordinarily a clinical study involves a comparison of new information with one or more of four data bases: earlier experience from that institution (historical controls); concurrent clinical experience (preferably randomized trials or matched case controls); the patient as his own control; or the published experience of other institutions (the literature). Each of these has its drawbacks.

Our knowledge about cancer is like an iceberg; seven-eighths of everything can't be seen. Oliver Wendell

Holmes has described science as the topography of ignorance. Failure to control variables, or even to understand what they are, often results in poor inference. The basic purpose of a randomized trial is to provide an even distribution of confounding variables, about some of which our knowledge will be incomplete. Every publication that makes a comparison of this kind should include a comparability table to assure the reader that relevant factors are evenly distributed.

Much has been said and written about historical controls. It is poor form to compare today's surgery with yesterday's radiation therapy. There is also difficulty in comparing treatment in community hospitals with that in referral hospitals. Herein resides another setting for selection bias. To get to the referral hospital the patient must be well enough to travel, hence the sicker patient with the worse outlook is likely to be in the community hospital. Complicate this further by subtle but important differences in criteria for staging. Comparisons of National Cancer Institute or American College of Surgeons regional or national data with those of the university hospital merit cynicism.

The prospective randomized trial provides the best possible conduit for meaningful treatment research in cancer. During the past 30 years, most of the gains in cancer treatment (and they are many) have been verified in randomized trials, and so it will be in the future as multidisciplinary cancer treatment is enhanced. However, the randomized trial also has shortcomings, in part because we have placed undue reliance on statistics. The P value has become the endgame. Many have pointed out that statistical significance is not necessarily biological significance. The utility of a significance level must always be subject to common sense.

The contrary is also true: statistical insignificance does not necessarily mean biological insignificance. This raises a problem that cannot be answered, but only exposed. It involves the importance of small margins in the difficult clinical decisions the oncologic surgeon must make for the cancer patient.

If you found you could perform an operation with a 1% mortality instead of a 2% mortality, you would promptly change your operation. And if you were certain you could achieve a 51% cure rate by a different treatment for cancer rather than a 50% cure rate, you would change your treatment. The problem is that you cannot know with certainty because the sample size needed to reveal that small difference would be virtually impossible to accomplish. Hence, a result showing no difference between two cancer treatments is limited by the conditions of the study, particularly sample size.

The important point is: what do you do with the con-

clusion that the new treatment should replace the old treatment since the results are not significantly different? Suppose that the tumor has a pattern for first recurrence many years beyond the duration of comparison of the two treatments. Trials in breast cancer come to mind. The surgeon realizes that a small margin, if present, for the old or the new treatment cannot be proved or disproved. The surgical oncologist therefore must deal with small margins, recognize that very large trials cannot show small differences, and make clinical decisions that protect the patient from uncertainty.

Uncertainty is where we have legitimate differences of opinion. Addressing uncertainty is what research is all about. It is the purpose of research to convert un-

certainty into the body of knowledge about which we share rational conviction and agreement.

My comments have been addressed to pitfalls within our scientific discourse. Doubtless, much has been omitted. If these remarks seem critical, I would like to place them in a different perspective. A great deal of good has happened in surgical oncology. We are in the midst of a renewal in our field. There is nothing in our biophysical world we cannot learn about. During the coming years we can and must share in generating the gains that will undoubtedly occur to benefit the patient with cancer. The public good and our growth will be served best by care and accuracy applied to our discourse. Above all, we must not promise what we cannot deliver.