

## Presidential Address

# Challenges Facing Clinical Research

William C. Wood, MD

President-Elect Niederhuber, members of the Executive Council, my family and my friends in surgical oncology, and future presidents of this society. . . I begin by stating my humble but profound thanks for the honor of serving as the president of this society. I am honored to be following in the footsteps of people whom I admire and many of whom I count as dear friends. This society has contributed so much to the advancement of care for patients with cancer that is more effective, more humane, and more tolerable than would otherwise be the case. To review the papers presented here over the years is to see the development of scientific care for patients with solid tumors presented by the investigators who made the difference, despite the fact that other societies may be larger or play their trumpets at greater volume.

You may have noticed that I saluted future presidents of the society. In preparation for this address I read many of the addresses of former presidents. I was immediately struck by the fact that virtually all of these began by noting that the prior presidential addresses have just been read. Dr. Copeland was among those who produced tables of the prior presidential address topics and categorized them,<sup>1</sup> as have several other presidents.<sup>2,3</sup> As I considered this, there dawned the sad realization that the future presidents read these addresses because, like commencement speeches, these carefully wrought, elegantly phrased statements leave no imprint on the memory. One's presidential address, like one's funeral, requires that you be present, together with assembled friends and family. The difference is that only for the presidential address do you need a manuscript, and only what you say at your own funeral will be long remembered. Two addresses by surgeons stand out particularly in my mind above all the others I have heard. Unfortunately, what is memorable is their length. Each was so incredibly long

that one passed through all of Elizabeth Kubler-Ross's stages of grief in listening to it: denial, anger, bargaining with God, depression, and finally resignation. . . These talks so impressed me that I promise not to take that road to notoriety.

If one begins by looking backward, the society has heard reports at its annual meetings of remarkable changes in the management of solid tumors over the last 25 years. Some obvious examples come to all of our minds. In breast cancer, the standard operation has changed from removal of the breast to preservation of the breast—with reconstruction for those few women who still require mastectomy. Sentinel lymph node mapping has eliminated the need for axillary dissection with its attendant morbidity for the majority of women with breast cancer. Most exciting is the declining mortality from this disease.<sup>4</sup> In colorectal cancer, the survival benefits of adjuvant therapy are added to the survival advantage from resection of hepatic metastases for a subset of those who initially fail by tumor recurrence. Induction radiation and chemotherapy has made possible sphincter-preserving surgery for the great majority of those with rectal cancer. In the area of sarcomas, we have gone from standard amputation to standard limb-sparing procedures. In melanoma, we have seen a change from wide resection for all to narrower margins. The role of elective regional lymph node dissection was first defined, then the introduction of sentinel lymph node mapping has allowed that procedure to be performed selectively. Vaccines and biological treatments are being introduced.

All of these transformations in the care of solid tumors have been made possible by clinical trials. Most of these clinical trials have been led by members of this society. This entire society has participated in these trials through membership in the National Surgical Adjuvant Breast and Bowel Program (NSABP) and in the other National Cancer Institute Cooperative Groups (CALGB, ECOG, SWOG, NCCTG), with leadership positions with breast, colorectal, melanoma, lung, and sarcoma committees of

---

Received April 18, 2001; accepted April 30, 2001.

Address correspondence and reprint requests to: William C. Wood, MD, Suite B206, Emory University Hospital, 1364 Clifton Road, NE, Atlanta, GA 30322; Fax: 404-727-4716; E-mail: william\_wood@emory.org.

these groups. This has culminated in the development of the American College of Surgeons Oncology Group, led by our former president Dr. Sam Wells.

In considering a topic for today, I was impressed both by the excellence of Glenn Steele's address and by his observation that the opportunity to deliver a presidential address was too singular to be occupied by a lecture that one could give on a more ordinary academic occasion.<sup>2</sup> So, based on 8 years as chair of the CALGB Breast Committee, 8 years as chair of the ECOG Breast Committee, and the last decade as chair of the NCI Breast Cancer Intergroup, I have chosen to address some of the challenges of the design and interpretation of clinical trials, with a couple of incidental observations that the occasion allows. This is pitched at our younger members and is part of a larger plea for rigorous science in clinical medicine and in the university.

Perhaps the most important challenge is that of designing elegant clinical trials that will test concepts rather than simply compare products. Because it is important to select the optimal clinical product or technique, trials addressing such technical matters are not inappropriate. They are simply less exciting than seizing the opportunity of the enormous investment made by patients, clinicians, and the federal government in national clinical trials to address major conceptual issues. A well-designed clinical trial defines our understanding of a concept, whichever way the results of the trial may play out. I cite just two examples, although many are available. The first is an intergroup trial, CALGB 8541 (Fig. 1).<sup>5</sup> Dr. Gianni Bonadonna had reviewed the results of adjuvant CMF for breast cancer and concluded that women who received the full dose of adjuvant chemotherapy did better than those who received less.<sup>6</sup> He concluded that there was a dose response. Others pointed

out that the healthier patients could tolerate the full dose therapy, and the dose may have merely staged the patients, not caused the differing outcomes. Some authors were convinced that quite low doses would be just as effective and cause less toxicity. Only a prospective study could answer this important question.

By a design that took as its standard arm the dose of FAC published by MD Anderson,<sup>7</sup> and compared a 50% increase in what was then standard dose intensity, yet delivered the same overall dose, it was possible to ask if dose intensity increases were as important as they were thought to be at that time. The third arm involved a 50% diminution in dose intensity from standard therapy and only half the total dose of the other two arms. This lower dose intensity was one that analysis of clinical management showed to be often adopted for signs of mild toxicity by many medical oncologists, believing that the tumor would be sensitive or not. This clinical trial accrued 1572 women and showed clearly that delivering a dose below the level that had been shown efficacious in prior clinical trials was accompanied by significant loss of benefit. On the other hand, there was no clear survival advantage to going up 50% in dose intensity. This was a foreshadowing that great increases in dose intensity might be possible, yet not contribute greatly in terms of benefit, as has been subsequently shown. The NSABP B-18 trial is another example of an elegantly designed trial of a concept (Fig. 2).<sup>8</sup> By keeping the chemotherapy exactly the same in both arms, it was possible to address the specific effects of induction chemotherapy versus postoperative adjuvant chemotherapy. This study showed that there was no difference whatsoever in terms of distant metastases and survival. It also showed that it was possible for many more women to have breast-conserving therapy when induction chemotherapy was used, than when surgery was attempted first.

Secondly, effective clinical trials must be feasible. Several national trials have been mounted to address the question of whether localized prostate cancer is better treated with various techniques of radiotherapy or by surgical prostatectomy. These trials have each failed for lack of accrual. The radiation therapists and surgeons involved in these trials could agree that the question was of overriding importance and remains unanswered, yet each group was so certain of the superiority and advantages of the technique that they employed that they found it impossible to present an ethical equipoise to their patients when discussing this trial. Consequently, patients would not consider the randomization. A second example was the series of trials that randomized aggressive adjuvant chemotherapy for breast cancer with super aggressive doses with rescue by stem cell or bone mar-

CAF	C 400	A 40	F 400 X 6	=	TD 2400/240/2400
CAF	C 600	A 60	F 600 X 4	=	TD 2400/240/2400
CAF	C 300	A 30	F 300 X 4	=	TD 1200/240/2400

**FIG. 1.** CALGB 8541 trial. C, Cyclophosphamide; A, adriamycin; F, 5-FU; TD, total dose.

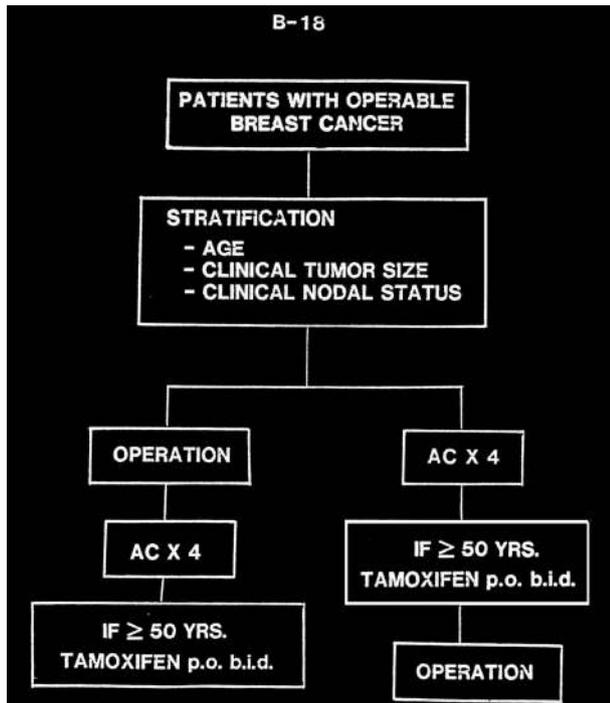


FIG. 2. NSABP B-18 trial

row infusions. There was such enthusiasm for the highest possible dose intensity, despite clues already available that such might be a vain hope, that in major centers throughout this country 10 or 20 women were transplanted for every woman who accepted randomization to these cooperative trials. Only after thousands of women had received bone marrow or stem-cell rescue therapies in the United States was it demonstrated that, at present, we can see no advantage of such super-high-dose therapy. A final example is the question of the danger of blood transfusion in cancer surgery.<sup>9</sup> In retrospective analyses an outcome difference is seen. Does it reflect an immunologic effect of the transfused foreign antigens, the effects of transfused cytokines, or are the transfused a different population or receiving different treatments? Within a given stage, why would one patient be transfused and another not? In that stage the more advanced may be more likely to require transfusion, or the one group may be more invasive, that is, different biologically. The transfused patients may have less skilled surgeons, with less adequate tumor clearance or greater likelihood of tumor spillage. Finally, one group of patients may bleed more easily, that is, be constitutively different or have more angiogenic tumors. Only a prospective, randomized trial can resolve this issue in humans, and it is not feasible because it is unethical.

Finally, a trial must be able to detect an achievable benefit. The dramatic progress made in improved survival of the most common solid tumors has come through a series of small improvements. It is essential that clinical trials be designed with realistic anticipation of the possible improvement to be achieved by the experimental intervention. Further progress will come through detecting moderate effects far more frequently than the occasional dramatic step forward that every investigator hopes to find. We must remember that an improvement of 10% in the survival of breast cancer would save far more lives than the discovery of a cure for Hodgkin's disease. The trials of tamoxifen are an example of the perils of failing to design trials to detect moderate effects (Table 1). Many trials failed to show significant benefit. Only when an overview analysis was performed was the benefit of the most effective agent that we have against breast cancer appreciated.<sup>10,11,12</sup>

Four additional challenges involving bias continue to bedevil clinical trials reported in reputable surgical journals. The first is avoiding bias in randomization. Randomization demands that there be no possible prior knowledge of the next allocation. So-called randomization by odd-or-even unit numbers, even grouped envelopes, fail this test. Secondly, there must be no bias in patient management. Early trials of minimally invasive surgery were complicated by the confidence that patients could be fed earlier, managed with less pain medication, and discharged earlier. While this may have been correct, these end points became self-fulfilling prophecies. When they were applied to the group that had open surgery, it was learned that they, too, could be managed with less pain medication, discharged earlier, and fed sooner than had been commonly done on many surgical services. Thirdly, a bias in follow-up can introduce error in cancer clinical trials that perform analysis based on anything other than death as the end point. Follow-up must be by protocol at fixed intervals for each group in the trial.

TABLE 1. Problems with small trials

Results of individual trials					
Deaths	Women/trial	SIG Bad	NS Bad	NS Good	SIG Good
<100	Few hundred	0	5	12	2
100-250	<1000	0	3	9	1
250-1250	Few thousand	0	0	7	3
Results of overview					
8,000	30,000	Mortality reduction 2-tailed $P = 10^{-14}$			

Adapted from the Early Breast Cancer Trialists' Collaborative Group 1990 Overview.

Finally, in order to avoid bias, it is essential to perform the analysis in terms of the intention-to-treat. It is still not a rare thing to read a publication in which only the patients who actually received a treatment are compared. Patients in one arm of a study after the point of randomization may experience a delay or additional studies. This allows the highest risk patients to be dropped before actually receiving the intended intervention so that the results are of two different populations. Those who actually receive the treatment are already in a selected and more favorable group. Rigorous science will avoid the errors such bias can introduce.

A potential error that all recognize is the avoidance of random error. Many small studies have seen differences that reflect the play of chance rather than the result of an intervention. Studies must be of sufficient size to avoid this common problem. This dictates that most randomized trials must be multi-institutional in order to accrue sufficient numbers to avoid such random errors.

Data-derived subsets are extremely useful for generating hypotheses, but totally worthless for drawing conclusions. Only the use of intention-to-treat and the avoidance of subset analysis, unless they are defined a priori, can avoid such error. Why are a priori subsets different? If a screw falls off an airplane and lands in a bucket in my backyard, it is not astonishing. It had to land somewhere. But if I place a bucket in my backyard tonight to catch a screw that I expect to fall off an airplane, and it does—that would be noteworthy. Finally, such a list of challenges would not be complete without alluding to ethical issues.

Shortly after I moved from Harvard to Emory, I received a call from the National Institutes of Health. The Office of Research Integrity stated that they had submitted a list of ten possible judges to the Canadian government, and that Dr. Larry Norton (this year's president of ASCO) and I had been selected to judge a very sensitive case. It was essential that we leave in about two weeks and go to Canada where we would be involved in a very secret series of case reviews. We could tell no one about this, including our spouses or our deans. I explained that I had just taken on the chair of a large department of surgery, was in the process of very aggressive recruiting, and would be unable to go. They explained that part of my development of Emory's program would probably involve applications for NIH funding and, if that were the case, it would probably be worth my while not to turn them down! I was capable of understanding a threat when I heard one and agreed at once to go. We arrived to find that we were reviewing the charts involving fraudulent data submission that had been detected during careful scientific review on the part of the National

Surgical Adjuvant Breast Program by Dr. Fischer and his colleagues. So secret was our mission that it was 2 years before I managed to get the NIH to pay for my air ticket and hotel costs, which kept getting "lost" each time they were resubmitted for payment! Dr. Norton and I kept this secret until we read about ourselves in the *New York Times* several years later when Congressman Dingell's committee was on a witch hunt against Dr. Fischer and the NCI. Dr. Sam Broder, then head of the NCI, told all of our secrets to that congressional committee. You are all aware of the damage done to the NSABP and all NCI research by the fraudulent reports of that one surgical investigator.

Two years ago, the leaders of the NCI cooperative groups in breast cancer and in bone marrow transplantation met in Washington to consider what the next bone marrow transplant trials should be. If the data from Dr. Bezwoda in South Africa were correct, it seemed unfortunate to spend 5+ years reproducing them. It would be much better to build on his results. If, on the other hand, they were not correct, it was essential to demonstrate that before initiating trials based on his findings. While designing the next group of studies, we agreed that Dr. Ray Weiss and several other investigators would go to South Africa to review these data in order to make certain that they were as clean as possible. The entire scientific community was distressed to learn that they were fraudulent to an amazing degree. As George P. Canellos, Editor-in-Chief of the *Journal of Clinical Oncology*, stated in his editorial of June 2000, "The recent exposure of fabrication of data in a visible trial of adjunctive high-dose therapy for early breast cancer has shaken the foundations of clinical research, which depends heavily on the integrity and accuracy of the investigator(s)."<sup>13</sup> A recent article in *JAMA* stated that about 40% of doctors would lie on insurance forms.<sup>14</sup>

The universities of the Western world are in the throes of postmodern philosophy. This involves not only a failure to study, teach, and understand epistemology as it has been understood for centuries, but a level of philosophical education that rarely gets beyond that of bumper-sticker slogans. A way of dealing with multiculturalism in the universities of this country and Europe as explored by Peter Berger, Yale's great sociologist,<sup>15</sup> has been reduced to "truth for you may be different from truth for me." Truth is seen to have no objective reality in many undergraduate courses of philosophy.

The law of noncontradiction is essential if we are to do either science or philosophy. If two contradictory statements can both be true, not only has communication been robbed of meaning, but there is no basis for experimental science. When Judeo-Christian presuppositions were the

background for Western thought, noncontradiction was taken for granted. Philosophers saw all ideas as valid or invalid in the mind of God and, to the degree that they had physical implications, capable of being tested scientifically.

Now as an aside, it's interesting that a recent Harris Poll show that 90% of Americans state that they believe in God. Yet it seems to me that many persons in departments of science in colleges and universities are exceptions. This is a particularly interesting anomaly if one considers the philosophical basis for an understanding of causality in the universe.

The universe has always aroused wonder in the mind of man. It has been seen to have only one of three possible explanations. First, it may be purely illusory. While defensible in introductory philosophy courses, this explanation has never been more than an academic exercise in Western thought. Secondly, it may have a cause outside itself, a cause that is self-sufficient as a category. This cause has universally been designated as God. Thirdly, it may be eternally existent—the self-sufficient category itself. This third cause was believed by many for centuries. It has been eliminated as a serious proposal by astrophysics: the expanding universe, stars that still contain hydrogen rather than only helium, the direction of radiation, evidence of increasing entropy, all demand a single point of translation of incredible energy into matter: “the big bang.” An eternally existing universe is acceptable to a philosopher but not to a scientist.

A further aside concerns the unwillingness of undergraduate teachers of biology to acknowledge the loss of the Darwinian model of evolution as an explanation for the increasing complexity of the evolutionary tree. When initially written, the idea was that a small organism swimming about that developed a tail by random chance, which tail could increase its motion though the primordial swamp, was advantaged in seeking food and mates, and selection favored retention of this aberrant flagellum. Today we understand that the development of a posterior extension and the complex that provides its motility involve a huge number of DNA couplet changes scattered in numerous places in the genome, none of which pose any evolutionary advantage at all until they are all complete. Darwin's brilliant observations describe the selection of favorable mutations in the genetic pool. But as an explanation for the incredible complexity of the phylogenetic tree, the replacement of the concept of intelligent design with one of random chance and selection was only tenable before the genetic code was understood. The entire evolutionary-progress hypothesis has expired on the rack of modern genetics, yet it lingers on in the undergraduate classroom. Random chance,

given modern probability theory, offers no explanation for the complexity of nature. Again, I plead for more rigorous science.

I enjoy asking medical students who have spent some time doing research to describe how they would describe the scientific method to a young student. Surprised at the naiveté of the question, they reply that the scientific method involves testing a hypothesis, then rejecting it if the experiment fails. I then ask if that was what happened in their year in the laboratory. The awareness dawns that actually we repeat and tweak the experiment when the outcome fails to meet our expectations. We all recognize our attachments to our favorite hypotheses that make us repeat experiments endlessly, varying the reagents or the species, or the experimental milieu until we finally achieve a result that fits with our hypothesis. Two generations ago, Walter Cannon pointed out that we are totally committed to our own hypotheses but always worry about the validity of our experimental work.<sup>16</sup> Our friends all believe our experimental work but think our hypotheses are ridiculous. Good surgical investigation involves careful design, rigorous science, and honest reporting of results.

I began with a backward look at the progress in surgical oncology, but it is exciting to glance ahead. Future clinical trials are on the drawing board that will address the new biology of oncology based on a genomic reclassification of tumors. This reclassification will not be based simply on genomic definitions, but on an integration of the genetic profile of the malignant transformation and the protein changes, on the receptor expression on the tumor cells, on the micro-environmental influences on the tumor, and the macro-environmental influences on the host. This all must be integrated with the phenotypic appearances of the tumor that have taught us so much over the last hundred years since biopsy was introduced. Ultimately, future clinical trials will address the revolution from minimally invasive and robotic surgery to the nano-machines of the future. It is our opportunity to lead the early phases of this revolution in the actual techniques of surgery. We all recognize that open surgery as we know it and practice it today will be needed far less often in the decades to come. Those who develop the minimally invasive and robotic techniques will be well positioned for future leadership.

After generations of rhetoric on the subject, we are actually moving from treatment to prevention in oncology. This will redefine what it means to be a surgeon. We are currently involved in leading many high-risk clinics for colorectal cancer, melanoma, and breast cancer, together with our colleagues in other areas of oncology. The move from treatment to targeted prevention can

either be embraced by surgical oncology or it will gradually exclude us. Finally, the future will involve economic considerations of minimally invasive procedures and the costs of specifically engineered small molecules that may address only 1% of the patients with a specific tumor type. At least at the time of their development and introduction, these will be associated with unaffordable costs. They will, consequently, only be possible as research techniques underwritten by major research budgets. Our role will include being advocates for our patients and our patients' welfare against competing economic interests.

At the dawning of this new millennium it is a great time to be a surgical oncologist. . . and the best is yet to come.

### REFERENCES

1. Copeland EM III. Presidential Address, Surgical Oncology. A specialty in Evolution. *Ann Surg Oncol* 1999;6:424–32.
2. Steele G. Presidential Address, Values in leadership—Lessons learned from patients, students, and colleagues. *Ann Surg Oncol* 2000;7:477–83.
3. Winchester DP. Presidential Address, The Society of Surgical Oncology and the Commission on Cancer: Progress through synergism. *Ann Surg Oncol* 1998;5:438–88.
4. Peto R, Borham J, Clarke M, et al. UK and USA Breast cancer deaths down 25% in year 2000 at ages 20–69 years. *Lancet* 2000;355:1822.
5. Wood WC, Budman DR, Korzun AH, et al. Dose and dose intensity of adjuvant chemotherapy for stage II, node-positive breast carcinoma. *N Engl J Med* 1994;330:1253–9.
6. Bonadonna G, Valagussa P. Dose-response effect of adjuvant chemotherapy in breast cancer. *N Engl J Med* 1981;304:10–5.
7. Buzdar AU, Blumenschein GR, Gutterman JU, et al. Postoperative adjuvant chemotherapy with fluorouracil, doxorubicin, cyclophosphamide, and BCG vaccine: a follow-up report. *JAMA* 1979;242:1509–13.
8. Fisher B, Bryant J, Wolmark N, et al. Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. *J Clin Oncol* 1998;16:2672–5.
9. Foster RS Jr, Foster JC, Costanza MC. Blood transfusions and survival after surgery for breast cancer. *Arch Surg* 1984;119:1138–40.
10. Early Breast Cancer Trialists' Collaborative Group. Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer: an overview of 61 randomised trials among 28,896 women. *N Engl J Med* 1988;319:1681–2.
11. Early Breast Cancer Trialists' Collaborative Group. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy: 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. *Lancet* 1992;339:71–85.
12. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomized trials. *Lancet* 1998;351:1451–67.
13. Canellos GP. The policing of clinical trials. *J Clin Oncol* 2000;18:2353.
14. Bloche MG. Fidelity and deceit at the bedside. *JAMA* 2000;283:1881–4.
15. Berger P. *A Rumor of Angels*. New York: Doubleday, 1990.
16. Cannon W. *The Way of an Investigator*. New York: Norton WW, 1945.