The Oncologist®

Clinical Medicine and Clinical Trials

JOEL E. TEPPER

Department of Radiation Oncology, UNC/Lineberger Comprehensive Cancer Center, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA

INTRODUCTION

Dr. Eli Glatstein has been an important mentor and friend during my 30 years in clinical radiation oncology. He has made countless contributions to our field, and those contributions are being celebrated as part of this special issue of *The Oncologist*. I am indebted to him for his guidance and help in the development of my career generally. I am particularly grateful to Eli for helping me to form my views regarding the role of clinical trials as a guide to clinical medicine. As well as anyone I know, Eli could penetrate the subtleties, the contradictions, the strengths, and the weaknesses of the clinically oriented scientific literature. It is that ability that has made him such a superb teacher and clinical scientist. That ability (along with many others) has had an enormous influence on me and has helped generate the ideas described below. I herein describe some of my thoughts and opinions, and their rationale, on clinical trials. Some of them Eli will agree with, and some he will not. But all reflect his teaching to think carefully about these important issues.

Clinical oncology is rapidly changing. Rapid introduction of newer, more sophisticated diagnostic tools, pharmaceuticals, and therapeutic technologies increases options for our patients. This can lead to confusion because it is often not clear when and if these new treatment options should be used. One of the most important things we need to do as physicians is to learn how to understand and interpret conflicting data and information in the medical literature. Herein, I discuss some of those issues, as they apply to oncology in general and specifically to radiation oncology.

Good clinical trials are important to help the clinician know when and how to employ these new, often expensive, diagnostic/therapeutic options. Nevertheless, many tools reach the clinic without *meaningful* clinical efficacy data. In order to be approved in the medical marketplace in the U.S., devices only have to demonstrate that they accomplish the goals for which they are intended, and that is often a technical, not a clinical endpoint. New drugs have a greater hurdle in that they need to demonstrate some measure of clinical efficacy, although the definition of efficacy is far from straightforward. At times no data are available to demonstrate clinical efficacy, but at times, even when data are available, they are often suboptimal for several reasons.

CHOOSING THE RIGHT ENDPOINTS FOR CLINICAL STUDIES

A central problem in many clinical outcome studies is the definition of benefit. The term is used loosely, and there is

often a disconnect between meaningful clinical benefit and statistical differences between therapies. Modest improvements in median survival, while statistically significant, may not be clinically meaningful because any small gain in survival may be offset by the cost, inconvenience, and toxicity of the additional therapy. Median survival reflects only a single point on the survival curve (i.e., the time when 50% of the patients are alive). Prolongation of median survival often reflects only a transient increase in survival that does not translate into long-term benefits. Even that benefit may accrue to only a small subset of patients.

The most stringent endpoint is survival, and specifically long-term survival. When a larger number of patients survives beyond some long endpoint (5 years, for example), there is a very strong expectation that the patient has indeed been helped by the therapy. The hazard ratio (risk for death in the experimental versus control arms) is a useful measure. However, it does not provide a sense of the *magnitude* of the benefit on an absolute scale, only on a relative scale. Therefore, if the baseline survival is very good, a very small absolute decline in the death rate can lead to a hazard ratio much less than 1.0, erroneously suggesting a large clinical benefit. (For example, reducing the absolute risk for death from 10% to 8% is a hazard ratio of 0.8, or a 20% reduction in the risk for death. Reducing the absolute risk for death from 1% to 0.8% produces the same hazard ratio.) The progression-free rate is another metric of some value. It is less robust than survival because it is heavily dependent on how often one evaluates for progression. Further, improvements in progression-free survival may not lead to improvements in overall survival.

Radiation oncologists deal primarily with issues of local-regional control of cancer. The relationship between local control and survival is not always clear. In breast cancer, many individuals thought that local control was a minor issue, and that improved local control did not translate into longer survival. However, recent data clearly show that improvements in local-regional control do translate into survival benefits, but not in a one-to-one manner. In patients undergoing breast conservation therapy with lumpectomy and radiation, approximately 25% of the improvement in local control afforded by the radiation translates into a survival advantage (i.e., an absolute improvement of 20% in local control produces about a 5% survival improvement) [1].

In some other diseases (e.g., soft tissue sarcomas, rectal cancer), although it is obvious that local control is necessary for cure, improvements in local control have not directly resulted in longer survival [2, 3].

The lack of a correlation in these diseases could be related to the fact that patients who are likely to develop local failure are also likely to develop distant disease, although clearly a patient cannot be cured unless they have control of their local disease as well as their distant disease—local failure will eventually lead to uncontrolled tumor. At times local failure does not translate into a survival advantage because effective salvage therapy is available (for instance, amputation for an extremity soft tissue sarcoma or abdominoperineal resection for a rectal cancer), although with the obvious sequelae of the additional therapy. Other times no survival advantage will be noted because patients have not been followed for a long enough time. Survival curves in rectal cancer often just begin to separate at 3 years and are not pronounced until >5 years [4].

Another metric to assess clinical efficacy is quality of life. Because a local recurrence can be morbid, quality of life assessments may be useful for radiation oncologists. A major problem with quality of life assessments, and the associated measure of quality-adjusted life years, is that they are time-consuming and expensive to obtain, and thus are often not incorporated into clinical trial design. Nevertheless, these analyses provide a way to combine survival duration with the palliative and toxic effects of therapy, into one relatively simple metric.

Cost-effectiveness is yet another way to quantify clinical efficacy. With the great interest in decreasing the cost of therapy, measuring the incremental cost divided by the incremental benefit (cost-effectiveness) of one therapy compared with another may hold greater importance in the future. The challenge to the researcher is to determine how best to use these various measures to define the best therapies for our patients.

Radiation oncologists are placed in a vulnerable position because we use expensive technology and because we direct our therapy against local-regional disease and not systemic control. The advances in radiation oncology have been enormous over the past few decades—improved target definition with better imaging from modalities such as fluorodeoxyglucose positron emission tomography scans [5], conformal radiation dose delivery, and other three-dimensional treatment delivery tools, and better integration of radiation therapy with surgery and chemotherapy [6]. In some anatomical sites, precise dose localization with dose escalation has provided further improvements such as with radiosurgery [7, 8]. Despite the apparent benefits of these interventions, we need clinical trials to assess objectively the utility of our advanced radiation planning/delivery systems.

INTERPRETATION OF CLINICAL TRIALS

There is a huge amount of literature that is available on virtually any medical topic, but this can at times be more confusing than helpful. Information can be anecdotal and just

reflect an interesting outcome in the experience of a single physician, the data can reflect a series of patients treated in a specific manner or can represent a prospective randomized trial. It is generally believed by physicians that randomized trials represent the best clinical information to determine the value of a therapy. However, when multiple clinical trials are run, too often they disagree in their measurement of outcomes, and this causes great difficulty in determining what represents "truth." Commonly, individual physicians accept as truth those studies that most closely conform to their preconceptions and find reasons to discount other trials.

The disagreements among trials can be that some studies are positive and others are indeterminate or that they clearly disagree. The field of meta-analysis is needed because individual trials either do not have enough power to answer a question or they are frankly discordant. For example, there are data suggesting that postoperative irradiation for breast cancer is detrimental in terms of long-term survival, but other data suggest a survival advantage. Data from the U.S. show a survival advantage for adjuvant chemotherapy in rectal cancer, but a number of European studies have not demonstrated a similar benefit.

Fortunately, clinical trial results are often concordant, and we are comforted that our decisions are made on strong evidence. But what do we do when studies are discordant, especially when the studies seem to be well performed? There are several possible reasons that different trials, addressing similar questions, produce different results.

The Endpoints Are Not the Same
Often the results of trials are simplified to the issue of whether a therapy produces a benefit, but as discussed above, "benefit" can be defined in many ways. With radiation therapy, one can evaluate survival, disease-free survival, local control, symptom control, etc., and each of these endpoints can be measured at multiple different time points. Short-term endpoints may provide different results than longer-term endpoints. Some physicians accept the value of obtaining local control as a critically important measure of benefit, whereas others may view survival as the only outcome that truly matters and the only outcome that can be measured with certainty. In general, differences in local control can be detected at shorter follow-up intervals than differences in overall survival. The presence of multiple possible measures of outcome can produce heterogeneity in the results of clinical trials.

The Patients Are Not the Same
Studies are designed with entry criteria that are often fairly broad. An adjuvant therapy trial might have entry criteria

that include patients with relatively early-stage disease as well as those with more advanced disease, but the patients who are actually included in a study could strongly emphasize one group over another. This could result in a treatment that is beneficial for a patient subset being swamped by the lack of value in a larger patient subset. The difference in patient composition can vary greatly among trials.

There also can be major differences in patient populations related to which patients decide to enter a study. Entry criteria can strongly select for a group of patients who are not representative of all patients with that disease. For instance, if a treatment is viewed as especially toxic, that may result in physicians entering only their highest risk patients into the study, even if patients with lower risk disease are eligible. Entry criteria can also be restrictive in a way that can cause an imbalance in patients actually being placed in a study. A possible example of this is the gastric cancer intergroup trial of adjuvant radiation therapy and chemotherapy after surgical resection. The entry criteria for that study required good nutrition 1 month after surgery. Patients who have more extensive gastric resections are less likely to be doing well nutritionally at this time, so the study could have biased entry to patients who had less aggressive surgical therapy [9].

In addition, there can be differences in patient entry into different clinical trials related to local practice patterns, both regional differences within a country and practice differences across multiple countries. These could relate to the use of different standard operative procedures, different staging techniques (or different interpretation of the same staging technique), or varying use of chemotherapy, etc. In the extreme, there are differing definitions of what pathological findings constitute an invasive cancer versus a non-invasive cancer, or what pathology defines an intermediate or high-grade tumor (such as a sarcoma) versus a low-grade tumor. Formal analyses have been performed on the variation in pathologist subclassification of both sarcomas and lymphomas, and these often demonstrate a discordance rate of approximately one third [10, 11]. These are not trivial differences and can have an enormous influence on treatment outcome and the results of a clinical trial.

Patient Evaluation Is Not the Same
Different tests, ostensibly measuring the same parameter, can produce different results. This can occur either at the time of patient enrollment in a trial or during the conduct of a trial while evaluating response or tumor control. A magnetic resonance imaging (MRI) scan of the liver to evaluate a patient for metastatic disease will produce somewhat different results than a computed tomography scan of the liver, and even different MRI scans will produce differing results

depending on the technical details of the scan. Pathology is usually viewed as the gold standard in many clinical situations. In addition to variation in defining the diagnosis, pathological evaluation can also affect the stage and extent of disease. Pathologists who analyze a large number of lymph nodes after a bowel resection are more likely to find nodal metastases and that finding can dramatically change the tumor stage and thus who enters a clinical trial or how those patients are stratified [12].

There is an additional problem now with more sophisticated studies for evaluating the presence of tumor in a specimen. Immunohistochemistry (IHC) can identify very small clusters of tumor cells, and these evaluations are done to varying extents in different institutions. It is likely that the patients who are node positive in breast cancer studies today are a substantially different population than node-positive patients in earlier studies, with a substantial stage migration from improved pathology and IHC. In addition to altering entry into a clinical trial, this information can affect study outcome directly if one is evaluating the pathological complete response rate, an endpoint that now becomes heavily dependent on the vigilance of the pathologist (who may have little interest in the clinical trial).

The Treatments Are Not the Same

Although we have made much progress in standardizing therapy over the past few decades, there are aspects of care that are very difficult to standardize. Chemotherapy tends to be the easiest of the standard oncologic interventions to make uniform, because there is little technical variation in treatment delivery—a dose of 1 g of a drug into a vein cannot be varied much among centers.

The other extreme is surgical management, for which quality control is usually determined by some combination of reading an operative report and reading a pathology report. Neither document is all that reliable and at best provides secondhand evidence of what surgery was accomplished. There have been a few attempts to standardize surgery in the context of randomized trials, specifically in the American College of Surgery Oncology Group (ACOSOG) studies of laparoscopic surgery, in which surgeons had to both demonstrate an adequate volume of cases and have video recordings of their operations screened for technical expertise [13].

This is difficult and expensive to implement in most situations (especially in studies of adjuvant therapies in which the surgery may have been completed prior to entry into a clinical trial even being considered). The value of an adjuvant therapy can be heavily dependent on what operation was actually accomplished for patients in those studies.

Radiation therapy tends to be in-between the situations of surgery and chemotherapy. Radiation treatment plans can be reviewed and the quality of machine calibration can be determined. However, because radiation oncology is a technically oriented field, there can still be large variations in implementation. Despite attempts at uniformity, the definition of dose delivered is variable. When one evaluates studies of intensity-modulated radiation therapy for prostate cancer, doses can be quoted at isocenter, as a minimal tumor dose, or as some other parameter, and that can produce substantially different dose prescriptions for what appears to the inexperienced reader to be the same treatment. Even if one were to make the dose definition uniform across centers, because different techniques are used to obtain that dose the actual dose delivered to varying portions of the tumor could vary among institutions. In addition, it has been shown that anatomic structures, especially tumor volumes, are contoured differently in different centers [14, 15].

In some anatomic locations, even small variations in the definition of a clinical target volume can result in major differences in treatment delivered. Although tumor targets are likely to be the areas with the greatest contour definition variation, problems also occur with normal tissue. For example, how much of the rectum (from proximal to distal) is contoured and how it is contoured (just the wall or the lumen also) when designing a treatment for prostate cancer with radiation therapy vary widely. Many institutions define the rectum over the range (with a 1- to 2-cm margin) of the rectum that will be in the radiation field. In this situation, there is likely to be a paradoxical effect of showing that the percentage of rectum being treated to high dose is lower when larger fields are used, even though more rectum is being irradiated. There are other examples of similar effects, but it is important to realize that tumor and normal tissue volumes can be retrospectively adjusted to produce the dose distributions that one desires. Mathematical precision does not necessarily imply accuracy or uniformity.

The Statistical Evaluation Is Not the Same

Biostatisticians at times do not agree on the best techniques to use when evaluating a clinical trial. Even simple issues such as whether to do a one-sided or a two-sided statistical test can easily turn a negative trial into a positive one (and vice versa). This does not mean that one approach is correct and the other is wrong, but rather that there is a difference of opinion that can lead to differing conclusions as to the meaning of a statistical result. In addition, although most statisticians agree that subset analyses should be hypothesis generating, subsets are at times analyzed as if they are providing definitive information.

INTERPRETING THE RESULTS OF MULTIPLE TRIALS

Given the large number of ways in which clinical trials dealing with the same topic can vary, it is almost surprising that we see concordance as often as we do. Of course, this is because if the therapeutic difference is large enough (or small enough), the differences (or lack of them) will dominate the results. What does one do when faced with a plethora of contradictory data? One approach that is commonly used is to do meta-analyses, or similar analyses, of multiple randomized clinical trials. This has the benefit of amassing a large amount of data into a single usable format. However, it has the distinct disadvantage that, unless the analysis is carefully constructed, poor data included in the analysis can skew the results. In addition, if one or two large trials dominate the results of the meta-analysis, any of the issues discussed above can have a major effect on the conclusions. Thus, as is true for medicine generally, if good data are used to perform the meta-analysis, the results are likely to be useful and meaningful. However, the analysis is no better than the data that go into it.

We must also be fully cognizant that a statistical difference in results between treatments does not necessarily mean that there is a "meaningful" clinical benefit. There is virtually always a tradeoff between side effects and toxicities of therapy and the benefit obtained. A study of adding a biological therapy to chemotherapy in pancreatic cancer produced an improvement in the median survival time of <2 weeks, a difference that was statistically significant [16]. Many would argue that this difference is too small to be clinically meaningful, but perhaps a more impressive difference would have been noted by evaluating the 1-year survival rate or quality-of-life endpoints.

In the present health care system, and with proposed changes to that system, one must ask whether the benefits of specific therapies are outweighed by the financial costs of the intervention. As discussed at the beginning of this article, there are many measures of benefit, but no generally agreed upon metrics outside the straightforward metric of long-term survival. But survival, as important as it is, is not the only meaningful measure. Determining quality of life parameters is important, but will it ever be possible for us to determine the value of a quality-adjusted life year?

Thus, there is no shortcut to obtain the "truth" in answer to a question, and even with the most careful analysis the answers are often ambiguous. One has to do more than read the conclusions of an article to determine if the conclusions in an abstract are valid. The answers to "how good a study is," and to understand the appropriate conclusions, come from careful reading of the study design and analysis, and understanding the specific questions that are answered in the clinical trial and the answers that are pertinent for the patient who one is seeing at that time.

There is a real risk in this approach. It is all too easy for our own biases to creep into the decision of which clinical trials are done correctly, and which analyses are meaningful. The human mind does a wonderful job of forcing facts to support our preconceptions. Despite our earnest desires to the contrary, we are all a product of our training and our environment. It is too easy for a medical oncologist to decide that a short extension of life span is clinically meaningful. It is too easy for a radiation oncologist to be convinced that improving local control in all situations will produce a benefit to the patient. It is too easy for a surgeon to be convinced that resecting another metastasis will improve a patient's quality of life. All physicians need to develop expertise in their chosen specialty and subspecialty, but physicians who do not have a broad understanding of patients, and the entire diagnostic and therapeutic environment, will not be able to understand the subtleties of many clinical trials, and will not be able to make the proper clinical judgments.

Eli Glatstein did not participate in my formal training as a physician. I took my first job with Eli at the National Cancer Institute (NCI) after completing military service in the Air Force. However, Eli was a true mentor to me as a junior attending physician at the NCI, and did what a mentor is supposed to do—allow an individual to develop his/her skills and interests so that they can succeed in their career. Eli Glatstein was, and is, the consummate clinician, and so much of what is discussed above comes directly from the things that Eli taught me and so many other young faculty members. He always knew that the data were most important, but that data taken out of context were of little value. Eli has never been a fan of meta-analysis because it so often did just that—separated the mathematical analysis of outcomes from the patient and ignored the careful evaluation of the strengths and weaknesses of individual trials. Although Eli should not be blamed for any of the comments above, he should be credited for those portions that are meaningful and helpful, because he was the genesis of so many of these ideas.

REFERENCES

1   Clarke M, Collins R, Darby S et al. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: An overview of the randomised trials. Lancet 2005;366: 2087–2106.

2   Pisters PW, Harrison LB, Leung DH et al. Long-term results of a prospec-

tive randomized trial of adjuvant brachytherapy in soft tissue sarcoma. J Clin Oncol 1996;14:859–868.

3 Peeters KC, Marijnen CA, Nagtegaal ID et al. The TME trial after a median follow-up of 6 years: Increased local control but no survival benefit in irradiated patients with resectable rectal carcinoma. Ann Surg 2007;246:693–701.

4 Radiation therapy and fluorouracil with or without semustine for the treatment of patients with surgical adjuvant adenocarcinoma of the rectum. Gastrointestinal Tumor Study Group. J Clin Oncol 1992;10:549–557.

5 Zakian KL, Koutcher JA, Ballon D et al. Developments in nuclear magnetic resonance imaging and spectroscopy: Application to radiation oncology. Semin Radiat Oncol 2001;11:3–15.

6 Harari P. Concurrent radiation/drug regimens: Current and future horizons. Semin Radiat Oncol 2006;16:64.

7 Carey Sampson M, Katz A, Constine LS. Stereotactic body radiation therapy for extracranial oligometastases: Does the sword have a double edge? Semin Radiat Oncol 2006;16:67–76.

8 Kavanagh BD, McGarry RC, Timmerman RD. Extracranial radiosurgery (stereotactic body radiation therapy) for oligometastases. Semin Radiat Oncol 2006;16:77–84.

9 Macdonald JS, Smalley SR, Benedetti J et al. Chemoradiotherapy after surgery compared with surgery alone for adenocarcinoma of the stomach or gastroesophageal junction. N Engl J Med 2001;345:725–730.

10 Coindre JM, Trojani M, Contesso G et al. Reproducibility of a histopathologic grading system for adult soft tissue sarcoma. Cancer 1986;58:306–309.

11 Classification of non-Hodgkin's lymphomas. Reproducibility of major classification systems. NCI non-Hodgkin's Classification Project Writing Committee. Cancer 1985;55:91–95.

12 Tepper JE, O'Connell MJ, Niedzwiecki D et al. Impact of number of nodes retrieved on outcome in patients with rectal cancer. J Clin Oncol 2001;19:157–163.

13 Weeks JC, Nelson H, Gelber S et al. Short-term quality-of-life outcomes following laparoscopic-assisted colectomy vs open colectomy for colon cancer: A randomized trial. JAMA 2002;287:321–328.

14 Spoelstra FO, Senan S, Le Péchoux C et al. Variations in target volume definition for postoperative radiotherapy in stage III non-small-cell lung cancer: Analysis of an international contouring study. Int J Radiat Oncol Biol Phys 2009 Jun 26 [Epub ahead of print].

15 Li XA, Tai A, Arthur DW et al. Variability of target and normal structure delineation for breast cancer radiotherapy: An RTOG multi-institutional and multiobserver study. Int J Radiat Oncol Biol Phys 2009;73:944–951.

16 Moore MJ, Goldstein D, Hamm J et al. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: A phase III trial of the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol 2007;25:1960–1966.