

Colorectal Cancer Endpoints Workshop

November 12, 2003

INTRODUCTORY REMARKS

Dr. Pazdur welcomed everyone in attendance and noted that the purpose of this meeting was to have a wide-ranging discussion about the positive and negative aspects of various endpoints for trials of drugs to treat colorectal cancer. By statute, FDA can take advice related to oncologic drugs only from the Oncology Drugs Advisory Committee (ODAC).

The meeting began with presentations by FDA staff on the regulatory background to the issue of endpoints in trials of cancer drugs. The panel then heard presentations on specific issues relating to endpoints for drug approvals in the colorectal cancer setting. In some cases, speakers framed their presentations as arguments for a particular position. Members of the panel then debated the issues raised by the speaker and offered alternative viewpoints.

1. REGULATORY BACKGROUND

1.1. Regulations and Endpoints (Speaker: Dr. Grant Williams)

Introduction

Drug approval in the United States requires adequate and well-controlled studies demonstrating that a drug is both safe and effective for the indication for which approval is sought. The safety requirement comes from the Federal Food, Drug, and Cosmetic Act of 1938; the efficacy requirement, from a 1962 amendment to that Act.

There are two routes to new drug approval in the United States. The traditional route? regular approval, also called full approval? requires the demonstration of *either* clinical benefit *or* an effect on an established surrogate for clinical benefit. Clinical benefit is usually considered to be tangible benefit of obvious worth to the patient, such as prolongation of survival or relief of pain.

FDA has sometimes accepted surrogates for clinical benefit as the basis for regular approval, usually after much clinical experience with the surrogate and widespread acceptance of it by both patients and physicians. For example, reductions in blood pressure and cholesterol are accepted surrogates for clinical benefit in the heart disease setting. On occasion, however, assumptions of clinical benefit based on a surrogate have later been proven wrong.

The second route to drug approval is accelerated approval (AA), which can be based on a surrogate endpoint that is considered reasonably likely to predict clinical benefit. AA is discussed at greater length below.

Usually more than one trial is needed for drug approval. This requirement is based on the definition of “substantial evidence of effectiveness” in the amended Food Drug and Cosmetic

Act and on the fact that the word “trials” is plural in that definition. However, FDA has recognized that in some cases results from a single trial may be sufficient for approval. Approvals based on single trials have been granted on occasion for many years, but this practice was written into law in the FDA Modernization Act in 1997.

FDA’s guidance document *Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products* (May 1998) stated that a single trial may suffice “generally only in cases in which a single multicenter study of excellent design provided highly reliable and statistically strong evidence of an important clinical benefit, such as an effect on survival, and a confirmatory study would have been difficult to conduct on ethical grounds.”

Evidence from a single trial may be sufficient for approval of additional marketing indications for previously marketed cancer drugs (*FDA Approval of New Cancer Treatment Uses for Marketed Drug and Biological Products*, December 1998). Evidence of drug effectiveness in different stages of the same cancer or evidence from closely related cancers may provide sufficient evidence for approval based on a single trial.

Regular Approval Endpoints in Oncology

Survival and improvement in tumor-related symptoms are accepted clinical benefit endpoints supporting regular drug approval. In selected settings, disease-free survival, complete response rates (e.g., acute leukemia), and partial response rate (e.g., hormonal treatment of breast cancer) are established surrogate endpoints supporting regular approval.

The Division of Oncology Drug Products recently evaluated the basis of approval for drugs approved by the division since 1990. This analysis showed that survival was the approval endpoint for a minority of approvals; 73% (48/66) of all approvals were not based on survival. When AAs are excluded, 67% (37/55) of all approvals were *not* based on survival.

The following are examples of cases in which improvement in tumor-related symptoms was the primary basis for regular approval:

- Mitoxantrone was approved for use in patients with symptomatic prostate cancer metastases on the basis of improvement in patients' bone pain.
- Two bisphosphonate drugs (pamidronate and zoledronate) were approved on the basis of a composite bone- morbidity endpoint, skeletal-related events.
- In several clinical settings, improvement in tumor-related symptoms plus objective tumor responses provided mutually supportive evidence that led to drug approval. In diseases with cutaneous manifestations, such as Kaposi's sarcoma and cutaneous T-cell lymphoma, improvements in cosmesis, cutaneous signs, and cutaneous symptoms have provided such evidence.
- In cancers obstructing esophageal or bronchial passages, approvals have been based on both improvement in symptoms of luminal obstruction and objective responses of intraluminal tumors. Such evidence supported the approval of photodynamic therapy for palliation of obstructing esophageal and endobronchial cancers.

Accelerated Approval

AA can be granted for drugs that treat serious or life-threatening diseases when the new drug appears to provide benefit over available therapy. AA can be granted on the basis of a surrogate endpoint that is reasonably likely to predict clinical benefit. After receiving AA, the applicant is required to perform a post-marketing study to confirm that treatment with the drug does indeed provide clinical benefit. If the post-marketing study fails to confirm clinical benefit, or if the applicant does not show due diligence in conducting the required study, the AA regulations describe a process for rapidly removing the drug from the market.

It is important to note that the quality and amount of evidence required for AA is not different than that required for regular approval. The applicant must show substantial evidence of the measured effect from well-controlled clinical trials. Borderline evidence is not acceptable. The difference is that the evidence may focus on a surrogate endpoint that is only *reasonably likely to predict benefit* rather than on an accepted clinical benefit endpoint.

Response rate (RR) has been the primary surrogate endpoint supporting AA for cancer drugs. To satisfy the AA requirement that a drug provide a benefit over available therapy, most sponsors have designed single-arm studies in patients with refractory tumors. In this setting? where, by definition, no available therapy exists? an objective response rate (ORR) of acceptable strength and duration has provided evidence of benefit over available therapy and thus have been the basis for AA. However, uncontrolled studies are limited in their ability to evaluate endpoints such as time to progression (TTP), quality of life (QOL), and survival.

In non-refractory disease settings, AA can be achieved by the demonstration of an improvement in a surrogate endpoint compared to a standard drug in a randomized trial. This approach allows drug activity to be tested in less refractory tumors and provides a toxicity comparison relative to standard therapy. It also allows the use of designs that may be more sensitive to the detection of added benefit, such as the so-called “add-on” design (A vs A+B), and the use of other endpoints such as TTP or tumor-related symptoms. Randomized trials also allow the determination of an individual drug’s contribution to a combination regimen (A vs B vs A+B). A study with this design supported AA for oxaliplatin in the treatment of colon cancer.

AA was the subject of a special ODAC session in March 2003. It was noted that the AA program has led to the approval of 19 New Drug Applications or Biologic Licensing Applications (involving 16 drugs) for new treatment indications. However, some problems were noted with the conduct and completion of post-marketing studies. The consensus of ODAC members was that post-marketing studies should be part of the drug development plan, consistent with the AA regulations, which state that phase 4 trials are generally expected to begin before drug approval is granted. One strategy that may ensure completion of the post-marketing study is to apply for AA on the basis of an interim analysis of a surrogate endpoint (such as RR or TTP) in a randomized trial, with clinical benefit to be confirmed upon the trial’s completion.

In conclusion, Dr. Williams said, an important question that would be the basis for much discussion at this meeting was whether TTP should be considered an accepted surrogate for clinical benefit in any colorectal cancer setting.

1.2. Past Approvals for Colorectal Cancer Drugs (Speaker: Dr. Amna Ibrahim)

Overview

Fluorouracil (5FU) was the first drug approved for the treatment of colon cancer; this approval, in 1962, predated the era of controlled clinical trials in oncology. After a long gap, levamisole was approved for adjuvant use in combination with 5FU in 1990 (Table 1).

Leucovorin (LV) was approved in 1991 for first-line therapy in combination with 5FU. Although reports in the literature have described results that support the use of 5FU/LV for adjuvant therapy, the FDA has not received an NDA submission for this indication.

Irinotecan (CPT-11) initially received AA for the treatment of recurrent colorectal cancer in 1996; this was followed by full approval for the same indication in 1998. Subsequently, in 2000, irinotecan was approved for first-line use.

Capecitabine, approved in 2001, is the only agent that has been approved for use in the first-line colon cancer setting on the basis of non-inferiority analyses. Most recently, in 2002, oxaliplatin received AA for use in combination with 5FU/LV in the treatment of recurrent colorectal cancer.

Survival was the endpoint supporting all regular approvals; randomized trials demonstrating superiority in survival led to all but one of these approvals. Two drugs received AA for use in previously treated populations. One of these approvals was supported by RR (RR) in single-arm trials and one by superiority in both RR and TTP in the interim analysis of a randomized trial.

Agents for Adjuvant Therapy

Levamisole (LEV) was approved in combination with 5FU in 1990 on the basis of the results of two trials. Following surgery, patients were randomized to no further therapy, LEV alone, or 5FU plus LEV. The follow up period was 2 to 5 years. A reduction in the mortality rate of about 30% was observed in the 5FU/levamisole arm. Although the contribution of LEV to the regimen was not demonstrated, this was the first adjuvant regimen to show a survival benefit and LEV was approved on the basis of these results.

Agents for First-Line Therapy

Combination 5FU/leucovorin (LV) was approved for treatment of advanced disease in 1991. A randomized five-arm study demonstrated an improvement in RR, TTP, and OS (OS) for high- or low-dose LV combined with 5FU. These results remained consistent following a three-arm extension of the initial study. OS was about 12.5 months in both the initial study and the extension.

In 2000, irinotecan was approved for first-line therapy after receiving AA (1996) followed by regular approval (1998) for treatment of refractory colon cancer. Two randomized, multicenter trials compared infusional 5FU/LV plus or minus irinotecan in untreated patients. Each trial enrolled more than 300 patients. Both studies demonstrated an improvement in RR, TTP, and OS for 5FU/LV plus irinotecan. A difference in survival was observed despite the fact that many patients on both study arms received second-line therapy and many patients in the control arm crossed over to the irinotecan-containing regimen.

Capecitabine is to date the only colon cancer drug to be approved on the basis of a non-inferiority analysis. Combined survival data from two open-label, randomized trials of capecitabine vs. 5FU/LV formed the basis of approval. Sufficient historical data existed to allow a reasonably precise estimate of the effect of 5FU/LV on survival. The non-inferiority analysis showed that at least 50% of the 5FU/LV effect was retained by capecitabine. This drug was approved for a restricted first-line indication: “For patients when treatment with fluoropyrimidine therapy alone is preferred.”

Agents for Recurrent Cancer

In 1996 irinotecan became the first chemotherapy agent since 5FU to receive approval for treatment of previously treated, advanced colorectal cancer. Three single-arm studies, with RR ranging from 14% to 21% and response duration of 5.8 months, were the basis of AA for second-line therapy. A survival benefit was subsequently demonstrated in two randomized trials that demonstrated superiority in survival by 2 to 2.5 months as compared with best supportive care and 5FU-based regimens.

Oxaliplatin in combination with 5FU/LV received AA on the basis of improved RR and TTP in an interim analysis of a three-arm randomized trial. Patients in this trial had disease that progressed or recurred within 6 months of treatment with 5FU/LV plus irinotecan (the Saltz regimen). RR of 9% was observed in the oxaliplatin combination arm, compared with 0-1% in the single-agent oxaliplatin and 5FU/LV control arms; additionally, TTP increased by 2 to 3 months in the oxaliplatin combination arm. Three important observations can be made about this trial.

- TTP and the small increase in the RR could be evaluated reliably because the trial was randomized.
- The inclusion of the single-agent oxaliplatin arm definitively showed the contribution of 5FU/LV to the combination regimen.
- The inclusion of the single-agent oxaliplatin arm also demonstrated that oxaliplatin should not be used alone in previously treated patients.

Follow-up of this study did not demonstrate a survival advantage for the oxaliplatin-containing regimen. According to reports in the literature and at scientific meetings, oxaliplatin provides benefits in first-line and adjuvant trials settings. However, these data have not yet been reviewed by FDA.

NDA's Discussed at ODAC but No Approved

Two NDAs were discussed at ODAC meetings but not approved. Some details of these applications remain confidential.

UFT. There were two main problems with the tegafur and uracil (UFT) application. Firstly, the contribution of uracil to the regimen was not demonstrated. For fixed-combination drug products, the regulations require that the contribution of each active component be shown. Secondly, FDA had reservations about the adequacy of the non-inferiority analysis comparing UFT to 5FU/LV. ODAC voted in favor of approval if the sponsor could demonstrate the contribution of uracil to UFT.

Oxaliplatin. Two trials supporting the first-line use of oxaliplatin were presented to ODAC in 2000. These studies compared oxaliplatin plus 5FU/LV to 5FU/LV alone. Increased RR and progression-free survival were shown in the oxaliplatin plus 5FU/LV arm. However, neither study showed that oxaliplatin provided a survival advantage using the protocol-specified primary analysis; one study seemed to demonstrate a trend toward poorer survival in the oxaliplatin plus 5FU/LV arm. ODAC voted no (12-0) when asked whether the analysis persuasively demonstrated a survival advantage for oxaliplatin.

TABLE 1. Summary of Past Approvals for Colorectal Cancer Drugs

Indication	Drug	Year Approved	Type of Approval	Basis for Approval
Adjuvant therapy	Levamisole + 5FU	1990	Regular	Superiority in survival
First-line therapy	5FU	1962	Regular	Superiority in survival
	Leucovorin + 5FU	1991	Regular	Superiority in RR, TTP, and OS
	Irinotecan	2000	Regular	Superiority in survival
	Capecitabine	2001	Regular	Non-inferiority
Therapy for recurrent disease	Irinotecan	1996	Accelerated	RR and/or TTP
		1998	Regular	Superiority in survival
	Oxaliplatin	2002	Accelerated	RR and/or TTP

2. FIRST-LINE THERAPY OF ADVANCED COLORECTAL CANCER

2.1. The Case for Time to Tumor Progression as a Clinical Benefit Endpoint in the First-Line Therapy of Metastatic Colorectal Cancer (Speaker: Dr. Langdon Miller)

Why a New Endpoint is Needed

Dr. Miller said that his presentation would make the case that an objective, non-survival clinical benefit endpoint is needed as the basis for full regulatory approval of new therapies for metastatic colorectal cancer. A new endpoint is needed, he said, because a rapidly increasing number of efficacious therapies has added therapeutic complexity in a disease that languished for decades with only one treatment, 5FU. New therapies have prolonged survival from a median of 12 to 13 months following treatment with 5FU/LV or capecitabine alone to approximately 16 months following treatment with irinotecan/5FU/LV and to more than 19 months following treatment with combinations and sequences of 5FU/LV, irinotecan, and oxaliplatin.

From the perspective of clinical trials design, the success that has been achieved with these multiple therapies has served to confound the relationship between early tumor control effects and long-term survival effects, disconnect early tumor control from long-term survival, and reduce the likelihood that further survival benefit will be observed. As a result, larger sample sizes are needed, more time is needed to accrue patients and acquire mature data, and studies become more costly to conduct. The implications of continuing to rely on the evaluation of survival as the primary measure of clinical benefit in the first-line therapy of metastatic colorectal cancer are the following:

- The value of survival as an endpoint is reduced.
- Drug development in colorectal cancer takes on added risk, time, and expense.
- The conduct of non-inferiority studies or multiple studies becomes impractical.
- Regulatory submissions are delayed.
- Active antitumor therapies may not be definitively studied.

Issues With the Evaluation of Symptom Control as a Clinical Endpoint in Colorectal Cancer

Evaluation of symptom control is complicated by several factors.

- Symptom severity is subjective.
- Disparate types of symptoms in metastatic colorectal cancer complicate interpretation.
- Symptoms are not uniformly present at diagnosis and are often not severe.
- Treatment and disease may induce the same symptoms.
- Relevant symptoms may be missed.
- Instruments may be insensitive to important changes in tumor size. For example, subjective measures of quality of life may not change despite objective tumor shrinkage.

In addition, analysis of symptom progression is often not useful because symptom progression usually follows tumor progression. Thus, the use of symptoms as a primary measure of clinical

benefit in colorectal cancer creates problems with complexity, subjectivity, reliability, and interpretability, all of which make study design and analysis difficult.

Advantages of Time to Progression as a Clinical Endpoint

Time to tumor progression (TTP) offers an objective, reliable, practical alternative to survival and symptom-control endpoints, Dr. Miller said. TTP represents the most common cause of treatment failure, incorporates the value of time, and offers a direct assessment of disease burden that logically correlates with symptom progression and survival.

An analysis of more than 1,000 patients receiving first-line therapy for metastatic colorectal cancer has shown that tumor progression is the most common cause of treatment discontinuation. By incorporating the value of time, TTP categorizes tumor control better than RR.

Changes in median endpoint values suggest that TTP correlates with survival in metastatic colorectal cancer; survival consistently equals approximately TTP plus 8 months. This correlation is preserved across treatment groups and when important prognostic variables are considered, but is altered by factors such as performance status and baseline lactic dehydrogenase (LDH). Testing of the correlation between TTP and survival in a Cox regression analysis, along with the important prognostic factors of baseline performance status and LDH, indicates that TTP is highly prognostic for survival within the population as whole.

TTP provides a direct reflection of drug activity and is not confounded by subsequent therapies, thus offering utility as an endpoint for non-inferiority trials. Relative to survival, the use of TTP as an endpoint would reduce sample sizes, shorten accrual time, shorten the time to acquisition of mature data, and decrease the cost of conducting registration studies.

TTP has several functional characteristics that favor its use as an endpoint. Because it is based on standardized radiographic tumor measurement criteria (the WHO and RECIST criteria), it can be physically described and objectively quantified. It is supported by data available in the primary patient record for FDA audit and can be subjected to central, uniform, blinded review. Further, TTP provides a clear method of presentation and interpretation, its analysis is straightforward and incorporates all available data, and it can be supported by secondary analyses to strengthen understanding of the results.

However, certain caveats must be considered. In particular, the minimum interval between tumor assessments should be less than the expected treatment effect size and the frequency of tumor assessment should be the same across study arms, even when treatment cycles are of different lengths. In addition, conservative censoring rules should limit TTP to time on the first-line study therapy.

Summary

In summary, Dr. Miller said, when properly evaluated, TTP satisfies several critical requirements as a drug approval endpoint.

- It directly evaluates changes in disease burden.
- It correlates with other outcomes (in particular, survival).
- It is not confounded by subsequent therapies.
- It offers utility as an endpoint in non-inferiority trials.
- It can be objectively quantified, reviewed, and audited.
- It offers clear interpretation and straightforward analysis.
- It conserves patient resources and hastens drug development.

2.2. Design Issues in Colorectal Cancer Trials: Surrogate Endpoints and Non-Inferiority Trials (Speaker: Dr. Thomas Fleming)

Criteria for Study Endpoints

Study endpoints must be sensitive and measurable or interpretable, but they must also be clinically relevant, Dr. Fleming said. Primary endpoints should unequivocally reflect tangible benefit to patients. Endpoints such as OS and reduction in disease-related symptoms clearly meet the criterion of clinical relevance.

In the hope of reducing the cost of and time involved in conducting clinical studies, greater attention is now being paid to the use of surrogate endpoints such as tumor burden outcomes (TTP, ORR) and biomarkers. The typical approach is to determine whether a treatment effect on the surrogate predicts an effect on a clinical endpoint. However, although an effect on a surrogate clearly establishes biological activity, it does not necessarily establish clinical efficacy.

The disease process may causally induce an effect on both the surrogate endpoint and the true clinical outcome. It should not be surprising, therefore, that the surrogate endpoint is frequently correlated with the clinical outcome. However, if the surrogate does not lie in the causal pathway of the disease process, an effect on the surrogate will not reliably predict an effect on the clinical outcome.

The disease process often influences the clinical outcome via several pathways. If an intervention affects the pathway that is mediated through the surrogate but does not affect other important pathways, false positive conclusions may be reached about the relationship between the surrogate and the clinical outcome. Conversely, if an intervention affects an important pathway that is *not* mediated through the surrogate, a false negative conclusion may be reached.

Even if the intervention has the desired effect on the pathway mediated through the surrogate, it may have other unintended effects that negatively affect the clinical outcome. Experience with the antiarrhythmic drugs encainide and flecainide is a classic example of such an unintended effect. These drugs were very effective at suppressing ventricular arrhythmias, a known risk factor for sudden cardiac death. As a result, an estimated 250,000 to 500,000 Americans annually were treated with these drugs annually during the 1980s. When a randomized trial was performed, however, it was found that encainide and flecainide had unintended effects that ultimately resulted in a tripling of mortality from sudden cardiac death among patients treated

with these drugs as compared with patients who received a placebo. The trial was halted in April 1989 when this finding was revealed.

Validation of Surrogate Endpoints

In the hierarchy of clinical endpoints, an endpoint may be (1) a true measure of clinical efficacy (that is, not a surrogate); (2) a validated surrogate endpoint; (3) a surrogate endpoint that is reasonably likely to predict clinical benefit (the standard for obtaining AA); or (4) a correlate that is solely a measure of biological activity unrelated to clinical benefit.

Validation of a surrogate requires that the effect of the intervention on the surrogate endpoint reliably predict the effect of the intervention on the clinical endpoint. However, correlation of the surrogate with the clinical outcome is necessary but insufficient to establish the validity of a surrogate endpoint. The surrogate must also fully capture the net effect of treatment on the clinical outcome. Thus, meta-analyses of many trials are needed to validate a surrogate endpoint. Additionally, validation of a surrogate requires a comprehensive understanding of both the causal pathways of the disease process and of the intervention's intended and unintended mechanisms of action.

For these reasons, validated surrogate endpoints are rare in clinical practice. An example of a validated surrogate endpoint in oncology might be durable complete response in a substantial fraction of patients being treated for leukemia or lymphoma.

Controversial Issues with Accelerated Approval

The AA process is intended to provide earlier access to promising interventions that have been shown to have an effect on surrogate endpoints that are reasonably likely to predict clinical benefit. Critically, however, a validation trial must be performed in a timely manner to confirm that the intervention does indeed provide clinical benefit.

Between 1995 and 2000, FDA granted AA to 12 drugs. In March 2003, data were presented to ODAC on eight of these approvals that remain unresolved. Once a drug has received AA, sponsors encounter difficulties enrolling patients into validation trials. As a result, the projected average time to completion of a validation trial for each of these agents is 10 years. In one case, the sponsor enrolled just eight patients per year into the validation trial.

In three cases, although validation trials indicated that the drugs had minimal treatment benefit, the products remained on the market. If there is no realistic expectation that a drug will be withdrawn unless it is demonstrated in a validation trial to have clinical benefit, AA is tantamount to a lower standard for full approval.

Design Issues in Non-Inferiority Trials

The aim of a non-inferiority trial is to show that the experimental drug is as effective as or better than, but not inferior to, an active comparator. Such a trial must enable a direct evaluation of the

clinical efficacy of the experimental agent relative to an active control and contribute evidence that enables evaluation of the efficacy of the experimental agent relative to a placebo.

In a superiority trial, it is insufficient for curves to be separated; there must be reliable evidence of an improvement in the clinical endpoint. Similarly, in a non-inferiority trial, it is insufficient for curves to overlap. The evidence must be sufficiently reliable to rule out the possibility that the experimental therapy is meaningfully less effective than the standard of care.

The International Conference on Harmonization (ICH) has defined a “suitable active comparator” as a therapy “whose efficacy in the relevant indication has been clearly established and quantified...and which can be reliably expected to have similar efficacy in the contemplated [active control] trial.” Historical data must be available to verify that the active control has clinical efficacy of substantial magnitude that is precisely estimated, with estimates relevant to the setting in which the non-inferiority trial is being conducted.

The determination of the margin in a non-inferiority trial is, according to ICH, “based on both statistical reasoning and clinical judgment, should reflect uncertainties in the evidence on which the choice is based, and should be suitably conservative.” The choice of margin should take into account the clinical importance of such factors as efficacy, the safety/tolerance profile, convenience of administration, and the likelihood of resistance or drug/drug interactions. Overly liberal choice of margins may ultimately result in a loss of confidence in the efficacy of an intervention. If non-inferiority trials are to be done successively over time, it is crucial to select rigorous margins, Dr. Fleming said.

2.3. Questions and Comments

Dr. Pazdur asked that during this initial question period, questions should focus on clarification of points made during the speakers’ presentations. Dr. O’Connell began by noting that in the two trials presented by Dr. Miller the correlation between TTP and survival was very striking. He asked whether meta-analyses have been done to correlate TTP and survival in other clinical trials. Dr. Miller responded that one meta-analysis had been based on the published summary results of 29 trials involving a total of 13,000 patients. In this analysis, RR was correlated with both TTP and survival and TTP was correlated with survival; all correlations were highly statistically significant. However, this analysis did not rely on primary patient data.

Dr. Schilsky asked whether the relationships between TTP and survival in the CPT-11 trial presented by Dr. Miller were likely to be generalizable to other types of treatment regimens and whether the magnitude of benefit in TTP would reliably predict a similar magnitude of survival benefit. Dr. Miller responded that in a review of recent registration trials, the relationship between TTP and survival (survival equal to TTP plus about 8 months without second-line therapy and 10 or 11 months with second-line therapy) seemed to hold up well.

Dr. Fleming reiterated that correlating longer TTP with longer survival is a first step. However, many markers that are correlates fail to reliably predict the true clinical effect of the intervention on the endpoint. A demonstration that the relative risk effect for both TTP and survival is

reliably correlated across a wide array of trials would provide much stronger evidence. Dr. Miller responded that he had access only to data on 1,000 patients who were enrolled in the two registration studies of CPT-11 in the first-line setting.

Dr. Pazdur noted that one of the purposes of this workshop was to identify areas in which further methodological research is needed. For example, it might be useful to review data from trials conducted by National Cancer Institute-supported cooperative groups to determine whether TTP and survival are consistently correlated and whether that correlation reliably predicts the clinical outcome.

Dr. Miller said it is simple common sense that lack of tumor progression is a good thing. Oncologists typically consider treatment to be successful as long as the patient does not progress and to have failed when progression occurs. Dr. Fleming noted, however, that from a clinical trials perspective it is important to quantify the degree of impact on tumor burden, the number of patients in whom treatment has the desired effect, and the duration of benefit, as well as to ensure that clinical benefit is achieved without other unintended effects. On many occasions, he added, what appeared to be clinical “common sense” subsequently turned out *not* to be a reliable surrogate for clinical benefit.

Dr. Williams commented that some endpoints have been used as the basis for drug approvals although strictly speaking they may not have been validated. Dr. Schilsky noted that although the totality of the evidence suggests that RR is reasonably likely to predict clinical benefit in solid tumors, very little data in the literature confirms that this is the case.

Drs. Pazdur and Fleming both pointed out that the strength and duration of the response to treatment influence the plausibility that a surrogate endpoint accurately predicts clinical benefit. For example, a 50% RR with a prolonged duration of response is more plausible as a surrogate for clinical benefit than a 10% RR with a short duration of response. Dr. Pazdur noted, however, that CPT-11 received AA on the basis of a 15% RR, which included partial responses; additional studies subsequently showed that the drug improved survival in both the first- and second-line settings.

It was noted that disease-free survival (DFS) has been accepted as a valid clinical endpoint in breast cancer studies. Although RR has never been rigorously validated as a valid surrogate for the effectiveness of tamoxifen, it has long been accepted as such. Dr. Marshall said there should not be a different standard for colorectal cancer therapies than that which exists for breast cancer therapies.

2.4. Question-Based Discussion: First-Line Treatment Setting

(Moderator: Dr. O’Connell; Discussion Leaders: Dr. Krook, Dr. Marshall)

Dr. Pazdur said FDA had formulated several questions to focus the panel’s discussion. Dr. O’Connell noted that the purpose of the discussion was not to achieve consensus but to identify the advantages and disadvantages of the endpoints under consideration and identify areas where more knowledge or research is needed.

Validity of Survival vs. TTP as Clinical Endpoints

Question: *Is survival the only acceptable endpoint to support the approval of drugs for first-line treatment of colon cancer?*

Dr. Krook said that, as a clinician, he believes OS is the most important issue to the patient. Survival is also the only unequivocal endpoint. He acknowledged that it is more costly and time-consuming to conduct studies in which survival is the primary endpoint.

Dr. Kelsen agreed that TTP is currently less easily verifiable than survival but added that this situation may change as the accuracy of imaging technologies improves, as more consensus develops about what constitutes disease progression, and as it becomes more common for assessments of disease progression to be performed by independent review bodies.

Dr. Pazdur commented that lack of blinding in oncology studies may lead to ascertainment bias. Dr. Blanke noted that patients on trials often deteriorate or die for reasons that are unclear and that trials lack adequate censoring rules for classifying such events. He added that if TTP were to be adopted as a study endpoint, it would be important to have a standard ensuring that any case in which there was doubt as to whether a change constituted progression would be classified as progression.

Dr. Miller said that patients tend to closely monitor changes in tumor size and carcinoembryonic enzyme (CEA) levels because they know implicitly that such markers are important indicators of disease progression. Dr. Fleming responded that although many markers are valid predictors of subsequent risk, it does not necessarily follow that a treatment-induced change in that marker reliably predicts a treatment-induced change in the clinical outcome.

A discussion took place regarding whether absence of disease progression is a clinical benefit in and of itself. Dr. Fleming said absence of progression would be a clinical benefit if progression was always symptomatic and if, therefore, delaying progression meant delaying the onset of symptoms. Dr. Marshall noted that in both clinical practice and clinical trials, disease progression is a signal to stop the current treatment; he asked why this is the case if absence of progression is not a clinical benefit.

Dr. Williams said most people would agree that progression indicates a drug is no longer working, but it does not necessarily follow that progression predicts the benefit of the drug. Dr. Schilsky said he was not willing to accept that an increase in TTP is always of clinical benefit to the patient. He noted that most patients in the first-line metastatic colon cancer setting are asymptomatic at presentation, whereas treatment often causes significant symptoms. In this scenario, the patient may be said to have benefited from treatment if his or her survival is extended. In the absence of a survival benefit, however, it is debatable whether the patient has benefited from treatment. Dr. Schilsky added that he was willing to accept TTP as a surrogate endpoint for AA.

Dr. Kelsen suggested that it would be useful to see data on the extent of concordance between imaging experts as to whether or not progression exists. Dr. Krook commented that in his experience different standards are used to measure progression in cooperative group trials as compared with industry trials. Ms. Roach said there is a need to develop more objective and consistent ways of measuring TTP. Dr. Miller said he believes TTP can be reliably measured with rigorous practice in radiographic analysis, including a blinded secondary review, and conservative censoring rules.

Dr. Marshall proposed that simply counting the number of patients still on treatment at predetermined time points (e.g., 6, 9, and 12 months) would simplify the measurement of TTP and reduce, although not eliminate, bias. Dr. Williams said another approach would be to measure the number of patients whose disease had progressed at a uniform point in time. Dr. Miller pointed out, however, that this approach could miss an improvement in TTP that would be apparent if measurement occurred at a different time point (e.g., measuring progression at 6 and 9 months would not pick up an improvement in TTP from 4 months to 7 months).

Dr. Pazdur observed that one issue that arises in considering the use of TTP as an endpoint is the magnitude of change in TTP that would be considered significant. In response, Dr. Miller noted that three drugs (5FU, CPT-11, and oxaliplatin) have shown a 3-month improvement in TTP in rigorous randomized trials and one (bevacizumab) has now shown a 4-month improvement. Dr. Fleming commented that data from meta-analyses are needed to show that X interval to progression reliably predicts a survival improvement of Y.

Dr. Pazdur asked whether FDA should continue to require that trials be powered to show a survival advantage even if TTP is the primary endpoint and whether sponsors should be required to submit two trials. He noted that many trials now being submitted to FDA are underpowered; he said he feared that acceptance of TTP as an endpoint would ultimately result in the submission of trials underpowered to show TTP.

Dr. Fleming responded that if there were compelling evidence to validate TTP as a surrogate for survival, collection of survival data would be unnecessary. In the absence of such compelling data, however, sponsors should be required to show survival data. He added that comprehensive meta-analyses are needed of the relationship between treatment-induced effects on progression and treatment-induced effects on survival.

Dr. Marshall agreed that, even if TTP were to be accepted as a primary endpoint, the collection of survival data would remain important. He pointed out that improved TTP could have a negative impact on survival if it resulted in a patient's being unable to benefit from second-line therapy.

Dr. Benson commented that the submission of two trials with TTP as an endpoint would "provide an element of comfort" and would enable patients whose disease progresses to go on to second- and possibly third-line therapies, potentially increasing their OS.

Dr. Pazdur asked whether a drug (or drug combination) should be approved if it is shown to be safe and effective but to result in somewhat shorter median survival than another drug that has

already obtained marketing approval. He noted that the drug approval regulations do not require a drug to be *more* effective than other already approved drugs.

Dr. Krook responded that from a research perspective, the drug should not be approved. However, from a clinical-practice perspective, having a variety of agents to choose from is better for patients. For example, in a particular subgroup of patients it might be appropriate to accept somewhat shorter median survival in exchange for considerably less toxicity. Dr. Schilsky said that if the trials are valid, there is no good reason not to approve the drug, which would then have to compete in the marketplace with other available therapies.

Questions: (a) *Is the demonstration of non-inferiority with respect to survival a viable approach to drug approval in the first-line setting, or are the difficulties too great (e.g., the small, imprecisely defined survival benefits associated with standard therapy)?*
(b) *If the demonstration of non-inferiority with respect to survival is a viable approach, suggest active control treatments for these studies.*

Dr. Krook said the demonstration of non-inferiority with respect to survival is a viable approach if the goal is to identify treatments that are equally efficacious but less toxic than current therapies. However, the selection of an active control is challenging because standard therapy is currently a “moving target.”

Drs. Benson and Pazdur both said that, although searching for less toxic agents is a laudable goal, it is questionable whether non-inferiority trials are the best use of scarce resources in colorectal cancer research. Dr. Pazdur said FDA is concerned that if the control arm of a non-inferiority trial were sloppily done, the treatment effect could be lost and marketing approval could be granted to an agent that is in reality no better than a placebo. Such a concern could be addressed by requiring two trials, but this would increase the burden on sponsors. Another concern is that when the intervention in a non-inferiority trial is a drug currently on the market, a high level of crossover is likely to occur, increasing the difficulty of interpreting the study’s results.

Dr. Pazdur noted that when capecitabine was approved on the basis of non-inferiority to 5FU, FDA was able to draw on 20 years of data concerning the effectiveness of 5FU. However, it is rare for such a wealth of data to be available on an active control regimen. Ms Roach commented that patients generally do not understand the margin issue in non-inferiority trials. For example, most colorectal cancer patients are unaware that capecitabine may be anywhere from 50% to 150% as effective as 5FU. Non-inferiority trials with large margins, coupled with surrogate endpoints, make efficacy very difficult for both patients and their doctors to evaluate.

Dr. Miller said the evolving treatment situation in the first-line colon cancer setting makes it very difficult to conduct non-inferiority studies. Additionally, survival may be so disconnected from the first-line treatment effect of the drug that an agent could appear to be non-inferior when in reality it is inferior. By contrast, TTP is a direct measure of a drug’s effect at the time of administration and is not influenced by the effect of second- or third-line therapies.

Dr. Fleming said a non-inferiority trial requires an active comparator with a substantial level of efficacy that is precisely estimated. He noted that whereas a poorly done superiority study is likely to underestimate effectiveness, a poorly done non-inferiority trial is likely to do the opposite, necessitating a much higher overall standard for quality in a non-inferiority trial.

Dr. Sargent noted that three large randomized controlled trials have now consistently shown a median survival of 20 months following treatment with oxaliplatin plus infusional 5FU, with CPT-11 as second-line therapy. Dr. Fleming responded that to conduct a non-inferiority trial one needs to know not only what median survival to expect in the active comparator arm but also what median survival *would have been* with a placebo instead of the active comparator. Generally, such information about the placebo effect is inferred from historical data.

Dr. Kelsen agreed that the field is currently too fluid to enable the selection of a single active control regimen for non-inferiority trials. Rather, he suggested proposing general principles to guide the selection of an active comparator? e.g., survival of X months demonstrated in at least two randomized controlled trials.

Dr. Schilsky pointed out that non-inferiority trials are costly and complex to perform, provide no improvement on existing therapy, and carry the risk that the field will have moved on by the time results are available. Requiring sponsors to conduct two non-inferiority trials merely doubles the disadvantages.

Dr. Miller asked whether substituting TTP for survival as the endpoint for non-inferiority trials would be a viable alternative. Dr. Fleming responded that designing a non-inferiority trial with TTP as a surrogate endpoint would be an even greater challenge than designing a non-inferiority trial with a survival endpoint.

Dr. Pazdur said that FDA has in some cases permitted sponsors to submit results of a trial conducted in one setting and of a second trial of the same drug conducted in an earlier setting.

Question: *Drug A has previously demonstrated a survival benefit. Drug B shows a superior TTP to Drug A. In the first-line setting, is TTP an adequate endpoint for full approval? that is, is it a reliable surrogate for clinical benefit? If so, is it a surrogate based on its prediction of delayed morbidity or delayed death?*

Dr. O'Connell noted that this question had already been addressed to some extent in the discussion of the first two questions. Dr. Marshall added that the adequacy of TTP as an endpoint for full approval would depend on the magnitude of the difference shown in TTP, toxicity, and survival.

Dr. Pazdur posed the following hypothetical situation: In a single trial, Drug A shows a 12-week improvement in TTP over Drug B (nominal P value 0.05). Drug A shows no improvement in RR and its toxicity is equal to that of Drug B. Are these data sufficient for approval of Drug A? Dr. Fleming said that in this situation it would be necessary to show confidence intervals that exclude the possibility that survival was meaningfully worse with Drug A than with Drug B.

Dr. Kelsen asked Dr. Fleming to comment on the effect on sample size of the need to provide more compelling evidence to support TTP as an endpoint. Dr. Fleming said three issues affect sample size: statistical power, false-positive error rate, and relative risk. Smaller sample sizes are acceptable in studies that use TTP as an endpoint because the relative risk for progression is greater than the relative risk for survival. Dr. Pazdur observed that powering a trial to detect a survival difference even though the primary endpoint was TTP would increase the likelihood of producing a statistically persuasive result on TTP. Ms. Roach commented that a disadvantage of smaller trials is that toxicities may not be revealed until after a drug is on the market.

Endpoints for Accelerated Approval in the First-Line Treatment Setting

Background: AA may be based on a surrogate endpoint reasonably likely to predict clinical benefit in settings in which a new drug demonstrates an advantage over available therapy.

Questions: Drug B is compared to standard Drug A, which has a small documented survival benefit.

(a) Should AA be considered for Drug B in the following circumstances:

1. Drug B has a superior response rate and/or superior TTP compared to Drug A.
2. Drug B has a response rate and/or TTP that is “non-inferior” to Drug A and Drug B is less toxic than Drug A.

(b) Consider whether, in either of the above circumstances, comparative survival data are needed before granting AA.

Dr. Marshall said RR alone would not provide a compelling justification for AA unless the magnitude of both partial and complete responses was overwhelming, as in the case of imatinib in the treatment of chronic myelogenous leukemia. The presentation of data that showed improvement in both RR and TTP, with subsequent evaluation of survival either in the same trial or a subsequent trial, could provide a viable strategy for obtaining AA.

Ms. Roach noted that the shortcomings of the AA process had become apparent at the ODAC meeting in March 2003, which focused on AA. She advocated using AA only for genuine “breakthrough” drugs, noting that AA is tantamount to full approval unless effective mechanisms are in place to withdraw a drug from the market when subsequent studies fail to confirm that it provides clinical benefit.

Dr. Pazdur observed that the decision of a sponsor to apply for AA on the basis of a RR in the refractory disease setting could ultimately shortchange a drug. An alternative strategy that some sponsors are now following is to conduct a single-arm trial while simultaneously accruing patients for a randomized trial. In this way, AA may be obtained on the basis of the results of the single-arm trial and/or the interim analysis of the randomized trial; final results of the randomized trial will validate (or not) the drug’s clinical benefit. Dr. Pazdur added that FDA is working with sponsors to address the problems that have been identified with the AA process. He emphasized that AA should be considered part of a longer-term drug development strategy, which does not conclude when a drug obtains AA. He noted that an advantage of studying a drug in the first-line setting is the likelihood of achieving a greater impact on the surrogate endpoints.

It was noted that the difference between the criteria for full approval and for AA are a matter of great confusion. Dr. Fleming said that TTP could be the basis for full approval if it were accepted that (a) an improvement in TTP is itself a tangible clinical benefit to patients, or (b) an improvement in TTP is a validated surrogate for clinical benefit. By contrast, TTP could be the basis for AA if it were considered *reasonably likely to predict* clinical benefit. A validation trial would then be required to confirm that the drug does indeed provide clinical benefit. It should be understood that unless the results of the validation trial are conclusively positive, the agent will not remain on the market.

Dr. Schilsky endorsed the strategy outlined by Dr. Pazdur but commented that this strategy would probably be easier to implement if RR, rather than TTP, were the endpoint, given the amount of time it is likely to take to accumulate sufficient progression events and to obtain data on progression events from multiple trial sites.

Dr. Sargent noted that the results of an interim analysis may change the final results of a trial. For example, if the findings of the interim analysis are positive, patients will cross over from the control to the experimental arm of the study. He added that even if TTP were to be accepted as a valid surrogate for clinical benefit in the case of cytotoxic agents, it may not be valid when the agent being studied is a molecularly targeted therapy or other non-cytotoxic agent.

Questions from the Floor

A questioner commented that TTP always seems closely correlated to the timing of CT scans; he asked how much bias this introduces into the calculation of TTP. Dr. Sargent responded that the extent of bias could be minimized by consistent timing of CT scans in both study arms.

A second questioner asked whether Dr. Fleming would agree that the most conservative definition of TTP would include progression *or* death. Dr. Fleming agreed that for TTP to be used as a clinical benefit endpoint, the calculation must include patient deaths. The questioner also challenged the relevance of the encainide/flecainide experience in the oncology setting, noting that it was the toxicity of those drugs that resulted in increased patient mortality. Dr. Fleming responded that there are multiple ways in which a surrogate endpoint can lead to a misleading conclusion about the ultimate effect of an intervention on the clinical endpoint. For example, the use of tumor shrinkage as a marker for the effectiveness of a cytostatic agent would result in false negative conclusions because such agents rarely induce tumor shrinkage, although they may substantially delay disease progression.

3. SECOND-LINE AND SUBSEQUENT THERAPY SETTING

(Discussion Leader: Dr. Berlin)

Questions: *In the second-line setting*

(a) Could prolongation of TTP in a randomized study be sufficient for regular approval?

(b) If not, could prolongation of TTP in a randomized study be sufficient for AA? Note that this study will have failed to show (or was underpowered to show) a significant survival difference (e.g., the recent oxaliplatin AA for treatment of refractory colon cancer).

In the refractory setting (no available therapy), could RR (with an adequate duration of response) demonstrated in one or more single-arm studies support AA? If so, discuss the RR and response duration that would suffice.

Dr. Williams commented that the temporal relationship between disease progression and morbidity or death may be different in the second-line or subsequent therapy setting as compared with the first-line setting. Dr. Pazdur asked whether there is any value in the use of TTP as a surrogate for survival if the time interval between disease progression and death is very short.

Dr. Blanke said the case for TTP as an endpoint is stronger in the second-line or later setting because patients have already progressed at least once and are likely to continue to do so and because previously treated patients are more likely to have symptomatic progression.

Dr. Sargent argued that TTP is less relevant in rapidly progressive disease, not only because of the constraint imposed by measurement of TTP at predetermined intervals but also because in advanced disease systemic factors may contribute to the patient's deterioration as much as tumor progression per se.

Dr. Miller noted that it would be difficult to provide data correlating TTP with survival in the second-line setting because second-line therapy in colon cancer is a relatively recent phenomenon. He said sponsors would be interested in using TTP as an endpoint in the second-line setting in a scenario in which an experimental agent is compared to best supportive care, with control patients permitted to cross over to the experimental therapy when disease progression occurs.

Dr. Pazdur observed that reasons other than lack of efficacy of the study drug may explain why a trial fails to show a survival advantage; for example, the trial may have been underpowered or significant crossover may have diluted the apparent survival benefit.

Several panelists agreed that there is a need for research on the relationship between TTP and survival in the second-line and subsequent therapy setting. Dr. Miller commented that little such data exists; in the two randomized trials that were the basis for full approval of CPT-11 in the second-line setting, neither RR nor TTP was systematically measured. It was noted, however, that the NCI cooperative groups could make available some data on investigator-assessed TTP in the second-line setting.

Dr. Williams said it would be interesting to conduct a study to determine whether investigator assessment of disease progression is biased. Dr. Sargent responded that the literature suggests independent assessment decreases RRs by about 10% and slightly reduces TTP.

Dr. Fleming observed that although it is more plausible that surrogate endpoints would be reliable in the second-line setting, they may also be less helpful because of the probable shorter time interval between progression and death in this setting as compared with first-line therapy.

Dr. Berlin asked whether it is possible to correct for crossover in a meta-analysis. Dr. Fleming responded that censoring patients at the time they cross over will not correct for the crossover effect unless all crossovers are random, which is extremely unlikely. One possible approach might be to analyze survival at an earlier time point (e.g., 1 year), when the data may be less diluted by crossover.

Panelists discussed the pros and cons of permitting crossover in trials. Dr. Fleming said that when there are as yet no data to show that an agent is safe and effective, there is no rational reason for permitting crossover. Dr. Marshall responded that allowing crossover enables more rapid recruitment into trials. Furthermore, it is possible that knowledge will be gained from the crossover group, albeit at the expense of diluting survival data.

Dr. Pazdur again referred to CPT-11, which received AA on the basis of a 15% partial RR yet went on to demonstrate a survival advantage in refractory disease in two randomized trials. He asked whether the emphasis on RR as a surrogate for clinical benefit in trials of second-line therapies was misplaced. Dr. Miller responded that a focus on RR overlooks the treatment benefit obtained by patients whose tumors remain stable or shrink less than 50%, the minimum degree of shrinkage required for classification as a partial response. Dr. Blanke observed that retrospective reviews have shown that stable disease extends survival in patients with colon cancer.

Dr. Pazdur said that if CPT-11's effect on survival was mediated through delayed TTP rather than through RR, this raised questions about other aspects of the drug development process, including but not limited to dose selection. Dr. Miller noted that the RR of 15% had been achieved in patients taking a 125 mg dose; when the trial was modified to permit patients to enter the trial on a dose of 100 mg, patients with lower performance status began to be enrolled and the RR declined to 8%. Thus, performance status predicted what appeared to be a dose-response effect.

A brief discussion ensued concerning the magnitude of RR that would support AA in the refractory disease setting. Dr. Schilsky commented that in this setting, rather than relying exclusively on RR, it was preferable to examine all data on the biological activity of an agent, including RR, rate of stable disease, and TTP. Dr. Fleming agreed, noting that surrogates often fail because they inadequately capture clinical benefit that is mediated through a reduction in tumor burden.

It was noted that in the second-line or subsequent therapy setting patients are usually more symptomatic and any disease progression generally predicts limited survival. For some patients, quality of life considerations may be more important than overall prolongation of survival; however, Ms. Roach pointed out that patients vary widely in the value they place on quality vs. quantity of life. Dr. Williams noted that lower toxicity alone cannot be a basis for drug approval; by law, a demonstration of efficacy is required.

In regard to the role of non-inferiority trials in the second-line setting, Dr. Pazdur commented that the difficulty of arriving at appropriate estimates of effectiveness would likely be compounded. Furthermore, the clinical benefit of any therapy in advanced disease is likely to be limited. Dr. Fleming agreed, commenting that the use of TTP (whether as a clinical endpoint in itself or as a surrogate for survival) in a non-inferiority analysis in the second-line setting would be “treacherous.”

Dr. Benson suggested that a composite endpoint might provide a more comprehensive picture of a drug’s effectiveness in advanced disease. However, Dr. Kelsen pointed out that such a composite endpoint would comprise individual elements that either have not been validated (TTP, quality of life) or are known to be unreliable (RR, performance status). He proposed that a prospective trial might be conducted to test the validity of these endpoints both individually and as a composite.

Dr. Pazdur observed that although drugs often first obtain approval for use in advanced disease, they generally progress rapidly to use in the first-line and adjuvant settings. Thus, two years from now, patients may be presenting with advanced disease who have already been treated with “second-line” agents and are unlikely to obtain additional benefit from further treatment with the same drugs.

4. OTHER ENDPOINTS IN ADVANCED DISEASE

The Potential Use of Biomarkers or Quality of Life Parameters in Colorectal Cancer Drug Approvals (Speaker: Dr. Charles Blanke)

Potential of CEA as a Predictive Biomarker in Colorectal Cancer

Carcinoembryonic antigen (CEA), a serum glycoprotein member of the immunoglobulin gene superfamily, is an intercellular adhesion molecule that promotes aggregation of malignant cells and is also involved in immunity, apoptosis, and metastasis. Serum levels of CEA are elevated in a variety of inflammatory diseases, both malignant (breast, lung, gastric, cervical, kidney, and bladder carcinomas; melanoma; non-Hodgkin’s lymphoma) and benign (cirrhosis, gallstones, emphysema, peptic ulcer disease, diabetes). Specifically, serum CEA is elevated in about 85% of patients with metastatic colorectal cancer.

Most, if not all, published data on CEA predate the advent of modern drugs, Dr. Blanke said. An American Society of Clinical Oncology (ASCO) expert group that reviewed the use of CEA in metastatic disease concluded that CEA was the marker of choice for monitoring colorectal cancer but that data were insufficient to recommend the use of CEA alone to monitor treatment response. The group did recommend that chemotherapy be discontinued when two increases in CEA have occurred, regardless of how well the patient is responding to therapy. This recommendation, however, is not widely known or followed in the colorectal cancer treatment community.

Whereas a CEA response can occur in the absence of a true objective response, a true objective response cannot occur in the absence of a CEA response. Furthermore, radiographic disease progression can occur in the absence of an elevation in CEA, although an elevation in CEA cannot occur in the absence of radiographic disease progression. Interestingly, CEA response correlates with survival even in the absence of objective response.

A further caveat concerns the association of CEA with hepatotoxicity. In a 1993 adjuvant trial, 39.6% of patients receiving 5FU/LEV sustained hepatotoxicity. In 19% of the patients who developed hepatotoxicity, CEA levels rose in the absence of disease recurrence (Moertel et al, Hepatic toxicity associated with fluorouracil plus levamisole adjuvant therapy. *J Clin Oncol* 1993;11(12):2386-90). This phenomenon does not occur when patients receive 5FU alone. Thus, there is a need to clearly define the potential for new drugs or combinations of drugs to produce hepatotoxicity-related increases in serum CEA.

Conclusion. In conclusion, Dr. Blanke said, all biomarkers, including CEA, overestimate response and underestimate progressive disease. Despite the association of a reduction in CEA with improved survival, it is not possible to consistently predict clinical benefit to the patient on the basis of a reduction in CEA. Therefore, the use of CEA as an endpoint in drug approval studies in colorectal cancer is not recommended. ASCO has recommended not using other biomarkers such as CA 19-9, DNA ploidy, flow cytometry characteristics, lipid-associated sialic acid, p53, and ras in the monitoring of colorectal cancer treatment. Finally, no biomarker can be used as a study endpoint until the study drug's effect on liver function has been fully characterized.

Predictive Value of Quality of Life Measurement in Metastatic Colorectal Cancer

Quality of life (QOL) has been defined as “a multidimensional construct that encompasses complete information on the impact of disease or its treatment on a patient's usual or expected physical, psychological, and social well-being.”

It is unclear how symptomatic the average patient with advanced colorectal cancer is, Dr. Blanke said. Although most standard texts state that “signs and symptoms are usually present,” no references are ever given for that statement. Trials tend to be skewed toward patients with fewer symptoms or minor symptoms; in a review of multiple series and trials, fewer than 50% of patients had significant symptoms. With modern early diagnosis, it may be necessary to focus more on delaying symptoms than on treating them.

Baseline QOL is known to be a major prognostic factor in colorectal cancer; it is stronger than performance status as an independent predictor of OS. In general, measuring QOL is a helpful way to balance toxicity and therapeutic benefit. However, the use of QOL as a study endpoint presents several problems.

- Whether palliative chemotherapy truly improves QOL is unknown. A meta-analysis of 13 randomized controlled trials comparing palliative chemotherapy with best supportive care concluded that the data on the palliative effect of chemotherapy on QOL were inadequate to draw firm conclusions (Simmonds, Palliative chemotherapy for advanced colorectal cancer:

systematic review and meta-analysis. Colorectal Cancer Collaborative Group. *BMJ* 2000;321:531-5).

- Inconsistent methodologies are frequently used to collect QOL data. Further, missing data are ubiquitous because few patients complete all questionnaires. This may lead to bias if patients who are ill or progressing are the ones not completing the questionnaires.
- Questions in QOL instruments may not be specific or sensitive to a particular agent or a particular toxicity. For example, most standard questionnaires used in colorectal cancer include no questions about neuropathy.
- The patients who complete QOL questionnaires in trials may not be representative of the general population with colorectal cancer. Further, the reliability of QOL instruments greatly depends on whether the physician or the patient completes the questionnaire.
- QOL instruments usually generate a composite score, which can make it difficult to evaluate the individual elements that comprise the composite score. For example, a score on a loss-of-appetite scale may be less significant to a patient than an identical score on a pain scale.
- It can be difficult to interpret the magnitude of QOL changes between patients.
- QOL instruments may not be sufficiently sensitive to differentiate disease progression from chemotherapy toxicity.
- Although the timing of QOL assessment is known to affect the results, no standard policy exists concerning when to conduct the QOL assessment if two regimens are given on different schedules in the same trial.
- When a QOL instrument is dominated by psychosocial domains, which are historically less sensitive to treatment-induced effects, changes in QOL scores may not reflect important chemotherapy-induced effects.
- Although QOL reflects both the efficacy and toxicity of a drug, it does not separate these two aspects. In theory, QOL could improve although a drug is ineffective. Conversely, with a toxic drug, QOL could worsen even if the patient had a major response to treatment.

Clinical benefit response (CBR), a composite measure of patient status that encompasses pain, functional impairment, and weight loss, is a possible alternative to QOL. CBR is defined as an improvement in at least one of those three parameters, with no worsening of the others, that is sustained for at least 4 weeks. In pancreatic cancer, CBR correlates well with QOL. However, the use of CBR as an endpoint in colorectal cancer trials presents several problems.

- CBR does not take into account all important disease-related symptoms. Furthermore, it would be difficult to develop CBR parameters specific to colorectal cancer given that there is only fair consensus among experts as to which symptoms (fever, fatigue, pain, weight loss,

diarrhea, abdominal swelling, appetite, constipation, etc.) are important or common. Additionally, in non-pancreatic gastrointestinal tumors, CBR does not correlate well with subjective response or QOL.

- The terminology used to discuss specific symptoms lacks consistency. (For example, is fatigue equivalent to tiredness, sleepiness, or a generic decrease in performance status?)
- The clinical significance of an improvement in CBR depends not only on the magnitude of change but also on the initial severity of the symptom. For example, an improvement from 2 to 1 on a pain scale would qualify as a CBR but is unlikely to be clinically important.
- A good research definition of symptom control is lacking. In the Radiation Therapy Oncology Group's osseous metastasis trial, changing the definition of symptom control changed the study's conclusions.
- No tool is sensitive enough to evaluate all the potentially important aspects of symptom improvement (e.g., time of onset of relief, duration of relief, significance of relief to the patient, severity of the symptoms relieved).

Conclusion. In conclusion, Dr. Blanke said, although baseline QOL is well correlated with important endpoints such as survival, it is not known whether a change in QOL reliably occurs with effective chemotherapy. QOL is insensitive for discriminating the effects of two drugs or regimens. QOL instruments are unable to discern the reasons for either an improvement or a decline in QOL, both of which may be multifactorial. Finally, QOL instruments cannot adequately differentiate between the efficacy and the safety of a study drug.

CBR offers the most promise as a study endpoint, but before it will be useful it must be refined to account for all common major symptoms in patients with colorectal cancer. CBR is not useful for assessing the effectiveness of a drug in asymptomatic patients. If validated, CBR would likely be considered a clinical benefit in itself rather than a surrogate for clinical benefit.

Discussion

Dr. O'Connell commented that CEA is less sensitive and specific than biomarkers used in other forms of cancer (e.g., prostate specific antigen, CA-125). Dr. Benson agreed that the use of CEA as a marker is problematic and that most oncologists will not use it as a determining factor in treatment decisions. Ms. Roach commented that, nevertheless, many patients use CEA as factor in treatment decisions.

Dr. Pazdur said that missing data has been a major problem in trials that have attempted to measure symptom control and QOL. Additionally, the fact that most oncology trials are not blinded reduces the credibility of findings related to symptoms or QOL. Another difficulty is that, in contrast to some other cancers (e.g., prostate cancer, esophageal cancer), colon cancer lacks a predominant or defining symptom that characterizes the disease.

Dr. Schilsky said the biggest single problem in clinical trials is failure to prespecify a QOL hypothesis that identifies a primary endpoint, the magnitude of change in the endpoint that the intervention is expected to induce, and the sample size required to detect that degree of change. As a result, very little QOL data from trials is interpretable. Dr. Pazdur agreed, observing that in many trials submitted to FDA, QOL data are added as an afterthought.

Dr. Miller said sponsors tend to view QOL as subjective, difficult and costly to measure, difficult to reproduce, and generally not very useful in trials of colorectal cancer therapies. Symptomatic patients tend to have lower performance status, which reduces the likelihood that they will respond to an intervention that is targeted to symptoms. On the other hand, in asymptomatic patients, in whom disease progression tends to occur about 4 months before symptom progression, it is difficult or impossible to evaluate the relationship, if any, between symptom progression and first-line therapy. Another difficulty is that when patients discontinue treatment as a result of disease progression, they tend to be evaluated less frequently for the onset of symptoms.

Dr. Pazdur noted that, although CBR was a factor in the approval of gemcitabine in pancreatic cancer, the primary basis for the approval was gemcitabine's demonstrated survival advantage. Dr. Kelsen commented that in pancreatic cancer, unlike colon cancer, severe symptoms develop early in the disease course. He suggested that attempting to measure the impact of treatment on QOL in colon cancer may not be the best use of resources.

Dr. Sargent said he believes QOL is important but is skeptical of the value of multidimensional QOL instruments. A single question? "How is your quality of life today on a scale of 1 to 10?" has been shown to reliably assess the impact of treatment on patients' QOL. Dr. Marshall suggested that resources could usefully be invested in developing better, shorter QOL instruments. Ms. Roach said that as the number of treatment options increases, it becomes more important to prospectively evaluate QOL. Dr. Blanke noted that a forthcoming Intergroup trial in colon cancer will compare a two-item QOL instrument with the standard FACT-C instrument.

5. ADJUVANT SETTING

3-Year Disease-Free Survival vs. 5-Year Overall Survival as an Endpoint for Adjuvant Colorectal Cancer Studies: Data from Randomized Trials (Speaker: Dr. Daniel Sargent)

Dr. Sargent presented preliminary findings from an ongoing meta-analysis to determine whether 3-year disease-free survival (DFS) could replace 5-year OS as an endpoint for studies of adjuvant therapies in colorectal cancer.

Investigators are collecting data from large randomized trials to (1) compare 3-year DFS and 5 year OS for each arm and (2) compare differences in 3-year DFS and 5 year OS between the control and experimental arms of each trial. Individual patient data are used when available. When individual patient data are not available, data are obtained from investigator-furnished summaries. To date, either individual or summary data have been obtained on more than 10,000

patients in 38 treatment arms. Data from 12 trials have been analyzed to date. Of these trials, nine had a no-treatment control arm and three had an active control arm.

Preliminary conclusions. The first preliminary conclusion is that, on an arm-by-arm basis, 3-year DFS appears to be an excellent predictor of 5-year OS. Event rates for 3-year DFS and 5-year OS were virtually identical. This means that the choice of endpoint (3 year-DFS or 5-year OS) would have no effect on sample size.

The second preliminary conclusion is that, as an endpoint for comparison, 3-year DFS may slightly overestimate 5-year OS, which could change the conclusion in a non-trivial proportion of trials. Twelve of 16 comparisons between arms produced the same conclusion for DFS as for OS. In half of the trials, the difference in 3-year DFS was greater than the difference in OS; in four trials, the reverse was true. This means that conclusions about long-term OS that are based on 3-year DFS must be considered subject to confirmation.

Dr. Sargent noted that this analysis is a work in progress. The investigators are actively seeking trials to add to the analysis. They may also analyze the data for other endpoints such as 2-year DFS and 3-year OS.

Discussion

Questions: For colon cancer drugs, does an increase in DFS compared to standard therapy represent clinical benefit and support regular drug approval?

(a) If so, what duration of DFS follow-up is needed for regular approval (3 years, 5 years)?

(b) If so, could AA be granted based on a shorter follow-up period (e.g., 3-year DFS for AA, 5-year DFS for regular approval)?

(c) If not, could a DFS improvement compared to standard therapy support AA? Would a survival advantage ultimately be required for conversion to regular approval?

Dr. O'Connell noted that three of the trials included in Dr. Sargent's analysis had shown a significant benefit in DFS at 3 years but no benefit in OS at 5 years. He asked whether there were any unusual factors about these trials that might explain these results. Dr. Sargent responded that several of the older trials included in the analysis (those that began recruiting patients in the early 1980s) were underpowered to show an OS benefit. In a number of trials, the DFS benefit was statistically significant by a small margin whereas the OS benefit fell short of statistical significance by an equally small margin.

In response to a question by Dr. Marshall, Dr. Sargent said that although analysis of aggregate data for all 38 arms appeared to show equivalence of DFS and OS, differences appeared when arms within trials were compared. In experimental arms, DFS tended to overestimate OS, whereas in control arms DFS tended to underestimate OS. A paper by Chen et al theorized that, for trials conducted when 5FU was the only available adjuvant therapy for colon cancer, patients who received 5FU when randomized to the experimental arm of a trial may have ultimately had slightly worse OS because they were unlikely to benefit from further treatment with 5FU when their disease recurred.

Dr. Fleming said that Dr. Sargent's data re-emphasize the point that a correlate is not necessarily a valid surrogate and that a correlation may not capture the net effect of an intervention on the clinical endpoint. He added that a different overall net effect might be seen if non-5FU-based treatment regimens were analyzed. Dr. Sargent agreed that his data are most relevant for 5FU-based regimens; it remains an open question whether the findings would hold true for other agents. In response to a question by Dr. Miller, Dr. Sargent said he intends to look at whether the analysis shows a change over time in the interval between progression and death.

Dr. Benson said he was concerned that Dr. Sargent's analysis tended to dilute the differences between individual patients. Patients with stage 2 or stage 3 colon cancer are a heterogeneous group and survival varies widely even among patients with stage 3 disease. Recent retrospective data suggest that the effectiveness of therapy may be strongly associated with the presence or absence of certain molecular markers.

Dr. O'Connell asked whether 3-year DFS might be an acceptable endpoint for AA. Dr. Benson responded that for patients with a poor prognosis DFS may be more significant than OS. He noted that many oncologists are now offering more intense therapy or combination therapy to high-risk patients, although it is not known whether such regimens are beneficial in this population. Dr. Pazdur noted that AA can be granted when there is uncertainty regarding the ultimate clinical benefit of an endpoint.

Drs. Sargent and Benson both said that although the data on DFS are promising, they believe it is premature to embrace DFS as a surrogate for OS. Dr. Schilsky disagreed, saying that delaying disease progression is a benefit to the patient and that he is persuaded that 3-year DFS is a valid surrogate for OS. Dr. Marshall said he was comfortable accepting 3-year DFS as a surrogate for OS, but he believed it remained necessary to track OS to ensure that unknown toxicities or other adverse effects did not eradicate any ultimate survival benefit. Dr. Cohen said most patients would regard an increase in 3-year DFS as a significant clinical benefit.

Panelists acknowledged that the granting of AA on the basis of 3-year DFS would likely result in significant trial crossover, which could obscure the evaluation of OS. It was agreed that if the period of time within which disease recurrence is most likely were known, it would then be possible to determine the extent of DFS most likely to predict improved OS. (For example, if 80% of recurrences occur within 4 years, it would be reasonable to accept 4-year DFS as a predictor of improved OS.)

In response to a question from the floor, Dr. Pazdur said FDA has not given more thought to the issue of how much benefit would have to be retained in a non-inferiority trial in the adjuvant setting.

6. RECTAL CANCER ENDPOINTS

Endpoints in the Adjuvant and Neoadjuvant Setting in Rectal Cancer

(Speaker: Dr. Meg Mooney)

Surgery has been the primary therapeutic modality for rectal cancer for over a century, Dr. Mooney said. However, adjuvant therapy for this disease has evolved considerably within the past 25 years. Postoperative chemoradiation therapy has been the mainstay of adjuvant therapy in the United States since 1990, when this approach was identified by a National Institutes of Health consensus conference as the standard of care for stage 2 and 3 rectal cancer. More recently, clinicians in both the U.S. and Europe have shown interest in neoadjuvant therapy, which offers the potential advantages of improved local control and increased opportunity for sphincter-preserving surgery.

Local control. The 1990 consensus conference recognized that there is a significant risk of local-regional failure as the only or first site of recurrence in patients with curative resected rectal cancer. For stage 3 rectal cancer the risk of local-regional failure may be 50% or more. Local-regional failure, with or without distant metastases, is the major mode of treatment failure and is associated with significant morbidity. Most failures occur within 2 to 3 years and failure is rare after 5 years. When surgery alone is the primary mode of therapy, successful salvage following local-regional failure is rare.

In two early trials of postoperative chemoradiation vs. surgery alone (GITSG 7175 and NSABP R02), patients who received both chemotherapy and radiation had a significantly lower rate of local failure. Another trial (NSABP R01) found a 5-year survival advantage for men treated with chemotherapy. An NCCTG trial found that chemotherapy and radiation conferred a significant 5-year survival advantage.

Two European trials compared a short course of preoperative radiotherapy with surgery alone. In the Swedish Rectal Cancer Trial, patients treated with radiation plus surgery had both a significantly lower rate of local failure and significantly improved 5-year OS. In the Dutch CRC Group Trial, in which surgery was standardized, preoperative radiotherapy reduced local recurrence but did not improve OS compared with surgery alone.

Sphincter preservation. Preoperative and postoperative chemoradiation have been compared in three trials. However, two trials performed in the United States did not accrue sufficient patients to produce significant findings. In the NSABP-R03 trial, despite low accrual, treatment-related toxicity and rates of sphincter-preserving surgery were similar in the two study arms.

Results of the third trial, conducted in Germany, were presented at the American Society for Therapeutic Radiology and Oncology annual scientific meeting in October 2003. Rates of sphincter preservation, anastomotic stenosis, and 5-year local recurrence were all significantly lower in patients who received preoperative chemoradiation.

Pathologic complete response. Numerous small studies have attempted to correlate pathologic complete response (pCR) with increased sphincter preservation, decreased rates of local

recurrence, and improved survival. These studies involved different therapeutic regimens, enrolled patients with different stages of disease, and were dependent on pathologic review as well as on the quality of the initial surgery. Rates of pCR ranged from 9% to 24% among patients treated with preoperative chemoradiation.

Quality assurance in pathologic assessment is one of the biggest problems in evaluating pCR. For example, in the U.S. GI Intergroup Adjuvant Trial INT-0114, rates of 5-year relapse-free survival varied depending on the number of lymph nodes that were analyzed. Additionally, fewer than 10% of pathology reports provided information on the circumferential margin.

Additional trials. Two additional large randomized trials in rectal cancer are planned in the United States. NSABP-R04, scheduled for activation in 2004, will compare preoperative 5FU plus radiation with preoperative capecitabine plus radiation. ECOG E3201, activated in October 2003, will compare postoperative chemoradiation regimens based on 5FU, CPT-11, and oxaliplatin. It is hoped that these trials will enable an evaluation of the advantages and disadvantages of local control, sphincter preservation, and pCR as endpoints in adjuvant and neoadjuvant therapy for rectal cancer, Dr. Mooney said.

Discussion

*Questions: (a) In selected drugs for local therapy (e.g., radiation sensitizers), is local control of rectal cancer a suitable endpoint for either full approval or AA?
(b) Discuss the role of pathological complete RR as an endpoint for AA or full approval in neoadjuvant therapy of rectal cancer.*

Dr. Cohen said local control is an appropriate endpoint and should be measured at 3 years because 80% of local failures occur within that time. He added that the ultimate goal in rectal cancer trials should be to measure colostomy-free survival, as is now the case in therapeutic trials in anal cancer.

Dr. Pazdur commented that sphincter preservation would be more applicable to lower-lying rectal tumors and that this endpoint involves a degree of subjectivity on the part of the surgeon. Dr. Cohen said the difficulty with both sphincter preservation and local control as endpoints is that they rely not only on radiation and chemotherapy but also on surgical expertise. By contrast, pCR is independent of surgical expertise and is thus the most objective endpoint in the neoadjuvant setting. Dr. Schilsky disagreed, pointing out that there is a lack of consensus among pathologists as to what constitutes pCR and how best to assess it.

Dr. Schilsky also noted that uncontrolled rectal cancer in the pelvis is an extremely morbid condition. If an agent were shown to improve local control and reduce the rate of local failure, with acceptable toxicity, that would be sufficient grounds for full approval. He added that sphincter preservation might or might not improve quality of life; a patient with a poorly functioning sphincter might be better off with a colostomy. Dr. Blanke also observed that sphincter function must be considered in addition to sphincter preservation.

Dr. Cohen said the central question in regard to pCR is the biological significance of minimal residual disease. He noted that in anal squamous cancer most patients do not have a recurrence of disease despite the presence of residual cancer cells.

7. QUESTIONS FROM THE FLOOR

A questioner asked what level of evidence would be required for approval of a combination therapy. Dr. Pazdur responded that it would be necessary to isolate the effect of the drug for which approval was being sought in order to demonstrate that it was contributing to the effectiveness of the combination regimen. He added that FDA has generally required that such an effect be shown in human studies.

Dr. Pazdur thanked all participants and the workshop adjourned.