FDA Public Workshop on Clinical Trial Endpoints in Prostate Cancer
June 21-22, 2004 – Bethesda, Maryland
Summary

Monday, June 21 – Morning session

INTRODUCTORY OVERVIEW: ENDPOINTS TO MEASURE THERAPEUTIC EFFICACY IN PROSTATE CANCER

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 prostate cancer. This workshop is the third in a series evaluating potential endpoints for drug approvals in the most common cancers. Previous workshops have considered endpoints in lung cancer and colon cancer. Issues highlighted at these workshops are subsequently discussed at meetings of the Oncology Drugs Advisory Committee (ODAC), the FDA's statutory advisory body on issues related to oncology drugs. By statute, FDA can take advice related to oncologic drugs only from ODAC.

Dr. Pazdur noted that time had been allocated on the agenda specifically for questions and comments from the audience. He encouraged those in attendance to make their views known or to contact him via email after the workshop.

The meeting began with a presentation by Dr. Pazdur 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 prostate cancer setting. Members of the panel then discussed the issues raised by the speakers and offered a wide range of viewpoints.

Regulatory Background (Richard Pazdur, MD, FDA)

Dr. Pazdur reviewed the regulatory process by which the FDA makes drug approval decisions. 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. 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, relief of pain, or measurable improvement in tumor-related symptoms. FDA has interpreted the 1962 amendment to require, in most cases, at least two trials for drug approval.

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, provided that post-marketing clinical trials are conducted to substantiate the purported benefit. AA is discussed at greater length below.

A basic dilemma in oncology drug development is that of marginal activity accompanied by high toxicity. To generate confidence that a drug is producing a true treatment effect, subjective bias must be minimized. Blinding of oncology trials is difficult because drugs are often given on different schedules and produce different toxicities; in addition, patients are often reluctant to enter a blinded trial. In the absence of blinding, FDA considers it important to have trial endpoints that minimize the opportunity for subjective interpretation of results by the investigator.

The presence of a true treatment effect can be substantiated by the magnitude of the statistical significance of the finding; by the internal consistency of results in subgroups; and by consistency among secondary endpoints (e.g., an improvement in survival accompanied by an increased response rate and/or a longer time to disease progression). External substantiation (e.g., duplication of results in a second trial) provides additional confirmation that a true treatment effect exists.

Surrogate Endpoints

FDA has defined a surrogate endpoint as "a measurement or sign used as a substitute for a clinically meaningful endpoint that measures directly how a patient feels, functions, or survives. Therapeutic changes on a surrogate are expected to change the clinical endpoint." It is important to note that a correlate is not necessarily a surrogate. The Prentice criteria for a surrogate endpoint require not only that a surrogate endpoint be correlated with the clinical outcome, but also that *the surrogate endpoint fully capture the net effect of treatment on the clinical outcome*.

Meta-analyses of clinical trials data can contribute to the validation of surrogate endpoints. However, the most important aspect of validation of a surrogate endpoint is gaining a comprehensive understanding of the causal pathways of the disease process and of the intervention's intended and unintended mechanisms of action.

FDA has been relatively liberal in its acceptance of surrogate endpoints for drug approvals. Although there are few established surrogates for clinical benefit in oncology, in consultation with ODAC the agency has accepted response rate and time to progression (TTP) as surrogate endpoints in regular approvals of hormonal agents to treat breast cancer. Complete response rates have been accepted as established surrogates for agents to treat leukemias.

With increasing frequency, FDA is accepting surrogates that are reasonably likely to predict clinical benefit as the basis for AA. For example, a 10% response rate in

refractory lung cancer was the basis for AA of gefitinib (Iressa); a partial response rate in refractory multiple myeloma was the basis for AA of bortezomib (Velcade).

Unproven surrogates are those that are used for exploratory or hypothesis-generating purposes. With more data, an unproven surrogate can become a surrogate that is reasonably likely to predict clinical benefit. Similarly, as confirmatory studies uphold the validity of surrogates now regarded as reasonably likely to predict clinical benefit, these endpoints can attain the status of established surrogates.

Although it is widely believed that FDA will accept only overall survival as an endpoint for drug approval, a recent analysis of drugs approved since 1990 showed that survival was the approval endpoint in a minority of approvals; 73% (48/66) of all approvals were not based on survival. When AAs were excluded, 67% (37/55) of all approvals were *not* based on survival (Johnson et al. J Clin Oncol 2003;21(7):1404-1411).

Improvement in tumor-related symptoms has been the basis for approval of a number of oncology drugs.

- Mitoxantrone was approved for use in patients with symptomatic prostate cancer metastases on the basis of improvement in patients' bone pain.
- Approvals of two bisphosphonate drugs (pamidronate and zoledronate) were based on a composite bone-morbidity endpoint (skeletal-related events).
- In several clinical settings, 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 lumenal obstruction and objective responses of intralumenal tumors. Such evidence supported the approval of photodynamic therapy for the 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.

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.

Survival vs. TTP: Pros and Cons

FDA is keenly aware of the pros and cons of survival vs. TTP as trial endpoints. Unlike TTP, survival is 100% accurate and its importance is unquestioned. However, trials that use survival as their endpoint take longer to perform and must enroll larger numbers of patients. In addition, the true survival effect of a treatment may be obscured by secondary treatment and crossover.

However, the use of TTP as a trial endpoint presents several challenging study design issues. TTP is a difficult endpoint to measure; meticulous care must be taken prospectively to ensure that a TTP endpoint has validity. When possible, trials should be blinded. Tumor assessments must be symmetrical on all study arms. Tumor progression must be prospectively defined and prospective methods must be in place for handling missing data. If progression is to be determined radiographically, independent radiology review plays a key role in the analysis and interpretation of trial results. Review of radiographic progression by blinded radiology panels provide credence to the endpoint.

The clinical significance of TTP must also be considered. If a trial is powered to detect a difference in overall survival, a relatively trivial improvement in TTP (e.g., 10 days, 2 weeks) may be statistically significant although its clinical significance is questionable. On the other hand, smaller studies powered to detect only improvement in TTP are likely to be underpowered to detect a difference in survival. Asymmetric assessments and missing data can call into question the reliability and precision of a claimed improvement in TTP. Another issue that continues to be debated is whether a delay in TTP is itself a clinical benefit to the patient or a surrogate for clinical benefit.

Response Rate

Response rate is a unique endpoint in that the treatment is entirely responsible for any observed reduction in tumor size. By contrast, the endpoints of survival and TTP comprise in part the effect of the disease's natural history as well as any observed effect of treatment. However, the duration and magnitude of the response must also be considered. Measurement of response rate excludes patients whose disease is stable or whose level of response does not meet the threshold to be considered a partial response.

Several methods exist for measuring response rate. FDA is less concerned with which method is used than with whether a single method is adopted and uniformly applied.

Palliation and Patient-Reported Outcomes

The credibility of palliation and patient-reported outcomes as endpoints is enhanced when trials are blinded. FDA has advised sponsors to use simple, hypothesis-driven instruments to measure these endpoints and to avoid the use of multiple endpoints. Because many patients enter clinical trials without symptoms, it can be difficult to measure palliation and delay in symptom development.

The use of these endpoints has been most successful in diseases in which symptoms are the hallmark of the disease—for example, bone pain in advanced prostate cancer (the

basis for the approval of mitoxantrone), dysphagia in esophageal cancer (the basis for the approval of photodynamic therapy).

FDA is interested in health-related quality of life (QOL) as a potential trial endpoint because it captures the patient's perspective on the success of treatment. However, the design of trials based on this endpoint poses many challenges. The interpretation of results is complicated by lack of blinding, missing data, and multiple endpoints. The clinical significance of small changes in QOL scores is often unclear. Also unclear is whether the use of QOL as an endpoint provides additional information or simply a more systematic recording of toxicity and symptom data.

Conclusion

Regulatory decision-making is a two-stage process. In stage one, the question to be answered is whether the drug has a convincing effect, which can be adequately characterized, on an endpoint. In stage two, the question is whether the observed effect is clinically relevant. This second question can only be asked if the first question is answered in the affirmative. One cannot discuss the clinical relevance of an uncertain or poorly characterized finding.

Relevance of Prostate Cancer Clinical States to Endpoints and PSA Endpoint Application Methodologies For Prostate Cancer (Howard I. Scher, MD)

The management of prostate cancer differs in several ways from the management of other malignancies, said Dr. Scher. In prostate cancer, unlike other tumors, the untreated history of the disease can span 10 years or more; treatment is often deferred at diagnosis, at recurrence, or at relapse; slowing tumor growth may be equivalent to curing the disease; and comorbidities produce a high risk of non–prostate-cancer death.

At initial presentation patients are classified using the Tumor Node Metastasis (TNM) staging system. However, TNM staging is only relevant to the untreated patient and does not inform trial designs for the patient for whom initial therapy has failed. The clinical states model was developed to address the shortcomings of the TNM staging system. The advantages of using this model are that it describes patients at any point in the disease continuum, is applicable to both treated and untreated patients, and provides a framework in which specific issues related to clinical trial design can be addressed.

The clinical states model that Dr. Scher and his colleagues have proposed begins by considering the patient who presents for a prostate cancer evaluation. Most patients today who undergo biopsies have clinically localized disease. If the disease recurs following local treatment, the sole manifestation of recurrence may be a rising PSA level. Detectable metastatic disease is differentiated on the basis of its sensitivity to hormonal therapy and on the presence or absence of castrate levels of testosterone in the patient's blood.

The model is applied clinically by considering at each patient encounter the manifestations of disease at that point in time and the probability that an asymptomatic patient will experience a clinically significant event within a given time frame; offering

appropriate therapy to eliminate clearcut manifestations of disease or to prevent the occurrence of clinically significant events; and/or deferring treatment if the probability of clinically significant events is low.

Applying the clinical states model to prostate cancer clinical trials, the critical questions are: (1) What is the objective? (2) What are the disease manifestations of the patient group (state) that will receive the intervention? (3) What change in those disease manifestations will be used to assess treatment effects? and (4) How will it be determined whether or not the intervention achieved a clinical benefit?

Objectives will vary according to the patient's clinical state. For the patient who initially presents with a high risk for prostate cancer, the clinical objective is prevention. For the patient with localized disease, the objective is to determine whether the disease is indolent and treatment can be deferred, whether local therapy alone is likely to achieve a cure, or whether a combined therapeutic approach is needed because the patient is at high risk for developing metastatic disease. For the patient with a rising PSA, the objective is to prevent metastasis. For patients with metastases, the objective may be to eliminate symptoms, prevent future symptoms, or prevent death from prostate cancer.

Irrespective of clinical state, assessments should be quantitative and reproducible. Outcome measures, like objectives, vary with clinical state; appropriate outcome measures when rising PSA is the sole manifestation of disease are very different from those that are appropriate when the patient has symptomatic metastases. Outcomes for each disease manifestation should be reported separately and should include both the proportion of patients showing the outcome (degree and proportion) and the durability of the outcome. Global categorizations such as complete response, partial response, or stable disease, which vary by disease manifestation and disease state, should be avoided.

PSA Endpoint Application Methodologies

PSA-based endpoints were originally proposed because of the difficulties involved in assessing outcomes by means of bone scans. Determinations of PSA level are both reproducible and quantitative. Natural history studies have shown that, across the disease continuum, rising PSA values precede other manifestations of progression. The use of PSA-based endpoints expands the opportunity for trial participation beyond the subset of patients who have measurable disease. Because changes in PSA levels can be monitored quickly, PSA-based endpoints may accelerate both the development of promising agents and the discontinuation of inactive ones. The central issue with regard to post-therapy PSA changes is whether the endpoint that is used relates to a specific biologic effect on the tumor.

Of the drugs currently approved for treatment of clinical metastases in castrate disease, none was approved solely on the basis of tumor regression or PSA endpoints. Androgen ablation, bisphosphonates, bone-seeking radiopharmaceuticals, and mitoxantrone/prednisone chemotherapy were approved on the basis of symptomatic relief with no improvement in survival. Bone-seeking radioisotopes and bisphosphonates have been shown to delay or prevent

symptoms. Agents such as GnRH analogs, anti-androgens, bicalutamide, and bisphosphonates have been approved based on reduced toxicities. The first agent to show a survival benefit in this disease state, docetaxel used in combination with prednisone, was approved in May 2004.

The fact that such a range of endpoints have been found to provide clinical benefit in one disease state illustrates that different criteria for approval decisions are necessary depending on the relevant disease state, the question being addressed, the type of drug, and the drug's mechanism of action. No decision criterion can stand alone; all manifestations of disease must be monitored concurrently and, ideally, at fixed intervals.

Some therapies initially cause PSA values to rise; in some cases, this initial increase may be followed by stabilization or by a slowed rate of increase. Treatment decisions cannot therefore be based solely on whether PSA values decline. More important is whether the observed pattern of PSA change—decline, normalization, disappearance, or modulation in the rate of increase—translates into clinical benefit. In the state of a rising PSA, prognosis can be assessed using measures such as PSA doubling time. Any recommended PSA endpoint must be shown to change the natural history of the disease by delaying the development of metastases or death from prostate cancer.

Another important issue is the fact that the definition of PSA progression will affect the duration of a treatment effect. For example, the duration of treatment effect will appear to be longer when PSA progression is defined as a 50% elevation from the nadir than when it is defined as the first confirmed rise in PSA value. Progression must therefore be prospectively defined.

In 1993 Kelly et al (JCO) analyzed the association between PSA decline and survival. Patients who achieved a 50% decline appeared to have a survival advantage over those who did not, an association that was sustained following multivariate analysis. In a second analysis (Scher et al, JNCI, 1999) involving a larger group of patients, those with no rise in PSA at 12 weeks appeared to have a survival advantage over those with any rise. In a multivariate analysis, whether PSA declined or not was the most significant factor influencing survival, although other factors such as tumor burden and extent of bone disease were also significant. However, the Prentice criteria for surrogacy were not met because no comparative trial showed a survival benefit. Another analysis found that the association between time-dependent PSA levels and relative risk of death was only 17%; thus, a significant proportion of survival difference remains unexplained simply on the basis of PSA (Verbel et al, Clin Cancer Res, 2002).

Caution must therefore be exercised in the use of PSA as a stand-alone endpoint. Other measures of clinical benefit are now being demonstrated. For example, measuring clinical events at a fixed time point can provide a more precise estimation of treatment effects, is applicable to agents with diverse mechanisms, and circumvents the need for surrogates that have not been validated. This design was applied in the trial of zoledronate vs. placebo, in which a clinically significant reduction in both the number of skeletal-related events and the time to occurrence of a skeletal-related event was seen in the zoledronate-treated group. Changes in PSA values did not differ between the two groups, nor was a survival difference shown.

Conclusion

Dr. Scher concluded by posing the following questions for discussion by the panel:

- Has a post-therapy PSA kinetics endpoint circumvented the difficulties in assessing outcomes in prostate cancer treatment?
- Can and should efficacy be assessed solely on the basis of early PSA changes?
- Should other outcome measures be abandoned?

Statistical Issues in the Validation of Surrogate Endpoints, With a Focus on the Evaluation of Treatments for Prostate Cancer (Stuart G. Baker, ScD)

Dr. Baker began by presenting hypothetical data from a clinical trial in which elevation of PSA levels was a surrogate endpoint. In this hypothetical example, 70% of patients who received Treatment B had elevated PSA, compared with 80% of those who received Treatment A. In terms of satisfying the surrogate endpoint, Treatment B appears to be superior to Treatment A. Does this imply, therefore, that Treatment B is better than Treatment A for the true endpoint of death from prostate cancer?

Suppose, Dr. Baker continued, that in this hypothetical trial the probability of death from prostate cancer was 0 in both groups for patients whose PSA levels remained the same, 0.50 for patients with elevated PSA who received Treatment A, and 0.86 for patients with elevated PSA who received Treatment B. In this scenario, the death rate from prostate cancer in the Treatment B group would be 60%, compared with 40% in the Treatment A group. Therefore, although fewer patients who received Treatment B had elevated PSA levels, more of them died of prostate cancer. The trial result for the true endpoint of death from prostate cancer differs from that for the surrogate endpoint of elevated PSA.

Data from a single trial are insufficient to enable any conclusions to be drawn about the validity of a surrogate endpoint, Dr. Baker said. Validation of a surrogate endpoint requires the collection of data from many previous trials in which the same surrogate and true endpoints were used. These data provide a basis for comparing predicted and observed true endpoints in a new trial. If the observed outcomes match the predicted outcomes in the new trial, the surrogate may be considered valid.

Dr. Baker also discussed how to account statistically for deaths from competing risks (causes other than prostate cancer) when trying to predict the effect of treatment on prostate cancer death. Deaths from causes completely unrelated to treatment can be ignored. However, if the rate of death from treatment-related non–prostate-cancer causes (e.g., infection) differs by treatment group, another surrogate endpoint is required to predict the probability of death from a competing risk—and this surrogate, too, needs validation.

Prostate-Cancer–Specific Mortality – Clinical Issues in its Use in Validation of Surrogates (Peter C. Albertsen, MD)

Accurate assessment of cause of death is important for determining clinical outcomes, calculating cause-specific survival rates, and calculating population-based mortality rates, said Dr. Albertsen. In prostate cancer, overall survival and cause-specific survival are very different entities because the disease has a long time horizon and many patients who have the disease die of unrelated causes. By contrast, in diseases with a short time horizon, overall survival and cause-specific survival are very similar.

Cause of death is usually determined from a patient's medical records, death certificate, or both. Population-based mortality rates are usually derived from death certificates, in which three lines are provided for the recording of an immediate cause of death and up to two underlying causes. Depending on the setting in which a death occurs, the death certificate may be completed by a physician who is unfamiliar with the patient's medical history.

Completed death certificates usually go to the state health department, where a clerk converts the cause-of-death information on each certificate into an ICD-9-CM code. Errors can easily occur as a result of a clerk's misinterpretation of the handwriting of the physician who completed the certificate. The ICD-9-CM codes are then sent to an office in Washington, where they are reviewed by a computer algorithm that determines official causes of death. The number of deaths attributed to prostate cancer is then tabulated and a death rate per 100,000 is determined. In general, therefore, maintained Dr. Albertsen, relying on death certificate data is a poor way to precisely determine cause of death.

In prostate cancer clinical trials, however, the central issue is determining whether a patient died of prostate cancer or of another cause. In large clinical trials there is often a cause of death committee, which comprises members from different specialties who review medical records and determine cause of death by consensus, ideally by following an algorithm. This process is currently being used in both the EORTC and PLCO trials.

Several years ago Dr. Albertsen and his colleagues performed a chart review to determine the extent to which death certificates agreed with the determination reached by medical record review that a patient had died of prostate cancer. They found agreement on the cause of death in 80% of cases when ICD-9-CM coding rules were followed. When disagreements occurred, they were equally likely to favor prostate cancer or another competing medical hazard. Disagreements usually occur if a man has more than one cancer or also has significant heart disease.

Conclusion

An explicit protocol for determining cause of death is mandatory for any clinical trial that relies on cause-specific survival as the endpoint. Two approaches to determining cause of death can be used: review of medical records by a committee blinded to the treatment assignment or use of death certificates processed according to the ICD-9-CM coding rules.

Monitoring Treatment Response of Prostate Cancer: Bone Scan and Beyond (Steven M. Larson, MD, PhD)

Bone metastases from prostate cancer are difficult to monitor using conventional imaging methods, said Dr. Larson. He presented a brief overview of bone-scanning technology. Bone scanning agents are taken up by hydroxyapetite crystals in the bone matrix; technetium-99m is the most commonly used radiopharmaceutical labeling agent. Indications for the use of bone scanning in prostate cancer are, in primary disease, to assess for metastasis when PSA is greater than 20 ng/ml; to detect metastatic disease; to determine prognosis and stratify patients; and to monitor treatment response.

A bone scan provides objective evidence of response, progression, and/or rate of progression. Dr. Larson and his colleagues have developed a bone scan index, a quantitative estimate of the percentage of total skeleton invaded by metastases, and have used the index to show that the extent of skeletal involvement is prognostically significant. However, bone scan findings are nonspecific and commonly lag behind tumor response; the technology is better at showing progression than at showing response.

Positron emission tomography (PET) imaging and other types of functional imaging will play a greater role in monitoring treatment response and disease progression in the future. PET imaging uses biomedical tracers (such as FDG, a form of glucose) to detect chemical signals from a tumor. Advantages of such tumor-specific functional imaging are that it can directly monitor the tumor's biochemical response to therapy and that functional changes precede anatomic changes.

A study comparing PET with bone scanning (Morris MJ et al, Urology 2002 Jun;59:913-8) found that in advanced prostate cancer PET is able to distinguish active from quiescent disease. PET has also been shown to be more accurate than bone scanning in monitoring response in progressive prostate cancer. In a recent study (submitted for publication) of PET as an outcome measure for progressive metastatic prostate cancer treated with chemotherapy, investigators concluded: "Little would be lost in terms of clinical decision-making if PET were used in lieu of the many standard studies presently used to determine treatment effects."

Conclusions

Diagnostic imaging is cancer-state–specific. Bone scanning should be performed in most trials of advanced prostate cancer to stratify patients and document progression. In the future, PET imaging using metabolic markers is likely to provide rapid, accurate evidence of treatment response. Evidence is beginning to accumulate that in advanced prostate cancer, FDG-PET will be a good marker of treatment response. Other more sensitive substrates also show promise.

Clarification Questions

In response to a question by Dr. Rhagavan, Dr. Larson said that flare, a false-positive bone scan finding, is more of a problem during the first 3 months of treatment. In the

recent study submitted for publication, bone scanning detected progression accurately but also identified two false positives within the 12-week study period. Dr. Larson added that the "superscan" reflects a high degree of skeletal involvement. PET scanning, however, often reveals less metabolically active disease in these patients than might be expected. It is probable that much of what is seen in a "superscan" is the equivalent of bone scarring although tumor may no longer be present.

Re-evaluation of Radiographic Outcomes: The Casodex Early Prostate Cancer Trial Program Experience (Kevin Carroll, MSc, AstraZeneca Pharmaceuticals)

The objective of the Casodex Early Prostate Cancer (EPC) Trial Program was to determine the benefit of adding 150 mg of bicalutamide (Casodex) to standard care for patients with early-stage disease. The program, which is ongoing, consists of three double-blinded, placebo-controlled trails involving more than 8,000 patients in 23 countries, powered to detect an improvement in progression-free survival (PFS) of at least 15%. Progression was determined objectively by means of bone scan, biopsy, or other imaging technique. A negative baseline bone scan was a requirement for trial entry. Mechanisms and procedures for radiographic imaging were determined by local practice. Bicalutamide was associated with a reduction of about 40% in the risk of progression.

It was recognized at the outset that the pharmacological effects of bicalutamide could compromise the trials' blinding. The possibility of acquisition bias was addressed in the trial design by requiring twice yearly bone scans in all non-progressing patients, irrespective of clinical indication. FDA requested an independent re-evaluation of locally determined progression outcomes to rule out the possibility of assignment bias.

The program sponsor undertook a retrospective re-evaluation of progression outcomes in more than 1,450 patients, including all 339 patients who had a PFS event determined by a positive bone scan or x-ray. The prospective re-evaluation protocol involved review by three independent expert radiologists who were blinded to randomized treatment, patient origin, original investigator bone-scan assessment, and all clinical data. Patients with a positive bone scan were separately re-evaluated by all three reviewers; patients with a negative bone scan were re-evaluated by two reviewers, chosen at random, with the third reviewer providing a blinded tie-break read if necessary. The re-evaluation took 1.5 years to complete at a cost of $5m.

A high degree of agreement was seen between the re-evaluation assessments and the original investigator assessments. Although inter-site differences in imaging technology introduced variability, there was no evidence of investigator bias. Reclassification rates were similar in the two treatment groups. Analyses based on the re-evaluation outcome measures continued to show a significant effect of bicalutamide.

Conclusions

Large scale re-evaluation of radiographic outcomes in the Casodex EPC program did not change the result, indicating that in randomized, blinded trials, investigator evaluation is a reliable basis on which to compare treatments. This experience suggests that in randomized, blinded trials, independent evaluation of clinical outcomes is unnecessary. If

independent evaluation is required, reviewers should not be asked to review images in isolation but should have access to all relevant clinical data.

Clarification Questions

Dr. Kantoff asked why 20% of bone scans initially read as positive were reclassified as negative in the re-evaluation. Mr. Carroll responded that the independent reviewers reached their conclusions solely on the basis of reviewing images; they did not have access to the other relevant patient data that in normal clinical practice they would have taken into consideration. He added that the rate of variability observed was equal in the two treatment groups.

Dr. Larson commented that it is always easier to have confidence in a negative bone scan than in a positive one. He added that in older men bone scan findings may be complicated by the presence of other degenerative conditions unrelated to prostate cancer. Mr. Carroll noted that reviewers had access to patients' negative baseline bone scans and so should have been able to rule out changes due to other degenerative conditions.

Dr. Albertsen noted that most patients enrolled in the North American trial had early disease diagnosed by PSA screening whereas most patients in the European trials had clinically detectable disease. Further, the results of the North American trial were negative whereas the results of the European trials were positive. He asked whether the number of positive bone scans was higher in the European trials and whether that had any effect on the re-evaluation of the North American patients. Mr. Carroll responded that when the data are broken down by trial, the same pattern is seen in each trial; although there is variability between the investigator and reviewer evaluations, there is no evidence of bias.

Patient-Reported Outcomes in Prostate Cancer (Derek Raghavan, MD, PhD)

The difficulty of assessing response in stable disease is a rationale for the use of patient-reported outcomes as endpoints in prostate cancer trials. However, because prostate cancer tends to afflict an older population, many patients suffer concurrently from other conditions that are common in older age groups (e.g., arthritis, benign prostatic hyperplasia), the symptoms of which may confound the assessment of patient-reported cancer symptoms. Optimal technology for differentiating symptoms with different origins has not yet been defined.

In addition to bone pain, advanced prostate cancer may be characterized by a number of poorly defined constitutional symptoms (e.g., pruritus, asthenia, malaise, weight loss). Urinary obstruction may produce local symptoms (e.g., slow stream, nocturia, frequency, hematuria) and may also be related to renal dysfunction, which may contribute to additional symptoms. Patients may experience bone marrow failure, leading to a constellation of symptoms. Involvement of unusual sites (e.g., liver, lungs, lymph nodes, skin) can also complicate the assessment of patient-related outcomes. Furthermore, the relationship between symptoms and survival varies widely; for example, a patient with pain and a change in performance status is likely to have a shorter survival than one whose only symptom is a change in PSA value.

Although relatively sophisticated tools now exist for assessing global QOL, there are often discordances between measures of tumor-related symptoms (e.g., pain, performance status) and measures of overall QOL. Furthermore, optimal methods of applying these tools have not been defined. For example, a reduction in pain intensity may or may not be significant depending upon the patient's baseline pain level. Other methodologic problems include dealing with missing data (which may represent the sickest patients) and defining the optimal statistical approach to analyze symptomatic changes. Confounding variables (e.g., the influence of a patient's knowledge of his PSA status on his reported QOL) must also be dealt with. The impact of hormonal deprivation on QOL is patient-specific and may be grossly underestimated.

Even objective response rates can vary widely depending on the criteria used to define response. The ability to apply standard criteria to define subjective, patient-reported outcomes remains a major challenge.

A dichotomy often exists between objective response, PSA response, and subjective response. For example, in a recent study of tesmilifene, a biochemical modulator of mitoxantrone (Raghavan et al, Proc ASCO, 2003), 66% of patients reported reduced use of analgesia and 75% had a PSA reduction of greater than 75%. However, no consistent relationship was seen between these responses and patient-reported QOL.

Conclusions

Assessment of symptomatic response may lead to both stage migration and response migration because of lack of quantification. Measures of QOL and symptomatic response are still being developed and validated; although useful, they should not be the sole parameter of outcome assessment. The discordance between PSA response, symptomatic response, and objective response is still poorly understood. In summary, patient-reported outcomes, although a very important part of the global assessment of treatment response, should continue to be regarded by the FDA as a work in progress.

Discussion

Questions

A. Bone Scans
- What are the pros and cons of using bone scan findings as components of an endpoint?
- What definitions for endpoints using bone scan are recommended for
    - Recurrence?
    - Response?
    - Progression?
- When must bone scan findings be verified by other clinical investigations?
- What is the optimal schedule for bone-scan follow-up?
- Do PSA findings help with interpretation of bone-scan findings?

B. PSA
- What are the pros and cons of using PSA and/or its time-dependent derivatives as an endpoint to demonstrate drug effectiveness?
- What PSA endpoints are worthy of consideration for use in oncology trials for
  - Recurrence?
  - Response?
  - Progression?
- Should any PSA endpoints be considered validated in any setting?
- If not, what studies or data are needed to validate PSA as an endpoint in various settings?

Discussion Leaders: Dr. D'Amico, Dr. Scher, Dr. Scardino

Dr. Hirschfeld noted that comments and questions by FDA staff should not be interpreted as statements of FDA policy.

A. Bone Scans

<u>What are the pros and cons of using bone scan findings as components of an endpoint?</u>

Dr. D'Amico summarized the key issues raised about bone scans as follows: Bone-scan findings are a solid endpoint that has repeatedly been shown to lead to a decrease in both QOL and survival. Disadvantages of the use of bone-scan findings as an endpoint include inter-reader variability (20% in the Casodex EPC Trial Program), the need for bone scans to be done at the same intervals in all study arms to assure comparability, the fact that bone scan findings tend to lag behind PSA progression, heterogeneity among patients with positive bone scans, and lack of consensus over whether the radiologists interpreting bone scans should be blinded to clinical data.

Dr. Larson commented that although bone scan findings lag behind the development of bone metastases, a positive bone scan is a reliable indicator of disease progression. He added that PSA velocity is helpful in determining when to perform a PET-FDG scan in a patient in whom recurrence is suspected because of a rise in PSA value.

Dr. Donald Coffey, in the audience, said he is inclined to agree with advocates who support changing the existing research paradigm, in which new drugs are studied and approved first in advanced disease, to one that focuses on early disease. In response, Dr. Scher commented that reasons for focusing trials on patients who have a rising PSA and are at high risk for metastatic progression is that such patients clearly require treatment, greater treatment risks are justified given the severity of the patients' condition, and the presence or absence of a treatment effect can be demonstrated more quickly than is the case in earlier disease.

Dr. Scher also noted that the goal of the Prostate-Specific Antigen Working Group, which he chaired, had been to set standards for obtaining an early estimate of treatment efficacy (Scher et al, J Clin Oncol 2004;22(3):537-556). He observed that a lot of information can be obtained from the PSA value; if the PSA does not change in response to treatment with a cytotoxic drug, it is unnecessary to perform a PET scan to determine

that treatment is ineffective. He added that PET scanning adds great cost to trials and expressed skepticism that the added cost was justified.

Dr. D'Amico noted that the kinetics of PSA (e.g., velocity, doubling time) predict the time interval to a positive bone scan. The question is when the predicted time interval becomes clinically significant.

What definitions for endpoints using bone scan are recommended for (a) recurrence, (b) response, and (c) progression?

(a) Recurrence. Dr. Scher said a baseline bone scan can be extremely helpful in ruling out comorbid conditions. He reiterated the importance of performing bone scans at fixed, consistent intervals and noted that multiple scans are needed in order to perform blinded, randomized comparisons. Bone scan findings must be considered in the context of other clinical information, including but not limited to PSA, hemoglobin, and lactic dehydrogenase values.

(b) Response. Dr. Scher said it is unusual for the findings of a 3-month bone scan to lead to a change in treatment in the absence of a rise in PSA or the appearance of new symptoms. It is likely to take at least 4 to 6 months for a significant treatment response to be shown on a bone scan, even though PSA values may have declined dramatically.

(c) Progression. Dr. Scher said that for noncytotoxic treatments in particular, it is imperative to leave no doubt that a treatment is *not* working. He argued that very subtle changes in a bone scan are not a good indicator of the absence of a treatment effect.

General discussion

Dr. Moul asked whether, in the light of the findings of the re-evaluation of bone scans in the Casodex EPC Trial Program, FDA would continue to require independent radiologic review of bone scans. Dr. Pazdur said FDA is actively discussing the circumstances in which independent radiologic review of bone scans should be required. He noted that independent radiologic review may be unnecessary if a trial is blinded. FDA's goal is to have confidence that a finding is true, he said.

Dr. D'Amico said that although re-evaluation changed the outcome in 20% of cases in the Casodex EPC Trial Program between two arms, false positive rates were balanced in all trial arms. In a large randomized trial, inter-reader variability that is consistent in all study arms is an indicator of the role played by individual professional judgment in the assessment of any radiologic test.

Dr. Keegan commented that rates of false positives or false negatives should not be confused with rates of concordance. In fact, she said, in the Casodex EPC Trial Program re-evaluation the rate of false negatives was twice as high in the bicalutamide groups than in the placebo groups, whereas the false positive rate was twice as high in the placebo groups than in the bicalutamide groups. Mr. Carroll disagreed with Dr. Keegan's interpretation of the data.

Dr. Kramer asked whether clinicians were blinded to PSA values in the Casodex EPC Trial Program. He posed the hypothetical example of a patient whose year 1 bone scan was negative but who, 6 months later, had a rise in PSA that triggered an additional, unscheduled bone scan, which was positive. He asked whether differential bias could have been introduced if the clinician was permitted to take the patient off therapy as a result of a positive finding on an unscheduled bone scan. Mr. Carroll responded that all positive bone scans throughout follow-up period, whether scheduled or unscheduled, were reviewed. The re-evaluation was requested because of concern about possible systemic investigator bias, but found no evidence that investigation behavior had biased the trial results.

Dr. Eisenberg noted that recently published findings suggested an increased hazard ratio for death among patients in the "watch and wait" group after median follow-up of about 5.5 years. Mr. Carroll said the preliminary survival analysis has shown no differences either overall or in any of the three trials; however, the survival data are not yet fully mature. In response to a follow-up question from Dr. Eisenberg, Mr. Carroll said the central issue with regard to the re-evaluation of the Casodex EPC Trial Program bone scan findings was whether such re-evaluation was worthwhile in a blinded randomized trial; he said the results of the re-evaluation suggest it is not.

<u>When must bone scan findings be verified by other clinical investigations?</u>

Dr. Scardino noted that there are problems with both the sensitivity and specificity of bone scan findings. He reiterated that the interpretation of bone scan findings is more likely to be accurate when a comparison is being made with a baseline scan and when full clinical information about a patient is available. It is unclear exactly what clinical protocol would verify a bone scan finding, he said. The use of a single variable such as PSA value as an indicator for a bone scan is inadequate because the determination of metastatic disease is multifactorial.

Dr. Larson said magnetic resonance imaging (MRI) is a promising technology for determining disease progression because it can distinguish early defects in the bone marrow before the findings on a bone scan are positive. However, more research is needed to compare the results of bone scans, MRI, and PET. He agreed that it is easier to define progression when a baseline scan is available for comparison.

Dr. Hirschfeld asked whether a new positive bone scan when the patient has no symptoms should be considered progression. Dr. Scher responded that corroborating evidence, such as another diagnostic imaging test or a second bone scan, would be needed to confirm the finding. It would be a mistake to take a patient off therapy solely on the basis of an isolated bone scan finding, he said. When a patient has been on hormonal therapy, it will take at least 6 months for objective change to become apparent on a bone scan. Paradoxically, around the time that the bone scan is showing improvement, the patient's PSA may be rising, which may confound the assessment of therapeutic effectiveness.

Dr. Raghavan asked whether follow-up bone scans should be performed on the same machine as the baseline scan. Dr. Larson said performing bone scans at a consistent site

is very important because of the high level of inter-site variability in techniques, methodology, instrumentation, and experience.

Dr. Kantoff said that at his center a positive bone scan is not considered progression in the absence of an increase in PSA value or symptoms. Dr. Scher asked whether a second bone scan 3 months later showing additional lesions, with no further change in PSA value, would be considered progression. Dr. Kantoff said no, adding that he had never seen such a clinical scenario. He added that a very small change in PSA value (e.g., <1 ng/ml) would constitute validation of a positive bone scan.

Dr. Scher said it is common practice to compare a current bone scan to the most recent previous one, whereas the most accurate assessment of the patient's condition is obtained from reviewing the trend over all scans, including the baseline scan.

Dr. Klein said that as a result of inherent differences in the timing of biological processes, bone scan changes may be "out of sync" with the patient's clinical situation. He suggested that biochemical markers of bone turnover that predict later bone scan changes would permit more timely interpretation.

Dr. Scher said patients should not be prematurely taken off a treatment regimen that they are tolerating, noting that novel agents may need time to show an effect and that it may be unrealistic to expect to see a dramatic response within a period of 2 or 3 months.

Noting that progression can be defined in many different ways, Dr. Williams asked whether it is currently practical to prospectively define a bone scan endpoint that can be followed and measured. Drs. Kantoff and Scher agreed that a bone scan in isolation is insufficient as an endpoint. Dr. Scardino responded that clinical decision-making in prostate cancer cannot be based on any single factor in isolation and that bone scan findings must be understood in the context of the patient's overall clinical situation.

Dr. Williams asked whether it was possible to devise an algorithm for a composite endpoint that included a positive bone scan finding. Dr. Scardino said he believed this could be done. Dr. Klein noted that the answer may be clinical-state–dependent; the interpretation of a change from a negative to a positive bone scan finding in a patient on adjuvant therapy may be more clear-cut than the assessment of the severity of progression in a patient who has already had a positive bone scan.

Dr. D'Amico endorsed the comments of other panel members concerning the necessity for a baseline bone scan. He observed that a positive bone scan is usually heralded by a change in PSA value and, in this context, is clinically important. A positive bone scan alone, unheralded by a PSA change, may be clinically unimportant. However, analysis of large trials has shown that, irrespective of PSA change, positive bone scans are closely correlated with death. This suggests that it is uncommon for a positive bone scan *not* to be heralded by a change in PSA value.

Dr. D'Amico added that osteoarthritis may be a confounding issue in an aging male population. In men with less advanced disease, bone scans may be of less utility than other indicators of disease progression.

<u>What is the optimal schedule for bone-scan follow-up?</u>

Dr. D'Amico said the optimal schedule for bone-scan follow-up is likely to depend on the patient's clinical state. For example, patients whose PSA is rising rapidly after surgery or radiation therapy are likely to have a positive bone scan within 12 to 18 months of documented PSA failure. In this clinical state, 6 months might be the most appropriate interval for bone-scan follow-up. For high-risk patients in the adjuvant setting, annual bone scans may be sufficient. In metastatic disease, a 3- to 6-monthly interval may be most appropriate.

Dr. Crawford said that in his experience in the Southwest Oncology Group (SWOG) bone scan findings have sometimes been helpful but at other times have simply caused confusion. In some trials it has been considered important to repeat bone scans after 6 weeks to confirm whether or not apparent new lesions were caused by flare. In newly diagnosed metastatic disease, the concept of minimal vs. extensive bone spread has sometimes been helpful in predicting survival, although an analysis of 10-year survival failed to demonstrate a survival advantage for patients with minimal bone spread.

Dr. Eisenberger said his experience suggests that a single positive bone scan should not be disregarded as a possible indicator of bone metastases. He added that, as a practical matter, when a patient has a slowly rising PSA in addition to a positive bone scan, it is usually difficult to retain him in a clinical trial.

Dr. Hirschfeld asked if it was possible to determine how far "off schedule" protocol-mandated bone scans could be to be considered informative. Dr. D'Amico responded that bias could be introduced if irregularities in the performance of scheduled bone scans were not balanced in all arms of the trial.

Dr. Albertsen said that although the probability of a positive bone scan increases directly with PSA value, there can be wide variation in the relationship between these two variables. For example, although one may have a positive bone scan with a PSA value of 10, it is also possible to have a negative bone scan with a PSA value of 100.

Dr. Scardino said that, in a patient whose PSA is rising after radical prostatectomy, a range of factors—including pre-treatment PSA value, pathologic stage, and Gleason grade—must be considered in attempting to predict the patient's likelihood of having a positive bone scan.

In response to a question from Dr. Moul, Dr. Larson said efforts are underway to develop quantitative bone-scan algorithms, but these tools are currently a work in progress.

Dr. Scher asked panel members whether they could define "bone-scan failure" in a patient without detectable disease at study entry or "bone-scan progression" in a patient with metastatic disease at study entry. Dr. D'Amico suggested that following a trend over time might be a feasible approach. For example, if a patient has a positive bone scan at study entry, it could be required that his next scan show increasing lesions that are confirmed in a repeat scan 3 months later. For a patient who has not had a positive bone

scan at study entry, any positive finding on a subsequent scan would have to be confirmed in a repeat scan 3 months later. Dr. D'Amico noted that this is the same methodology used to monitor patients' PSA values. A flaw in this approach, he acknowledged, is that once a patient has a positive bone scan, he may not be willing to wait 3 months for a confirmatory scan. Dr. Eisenberger commented that the required frequency of bone scans may depend on the nature of the treatment patients are receiving.

B. PSA

<u>What are the pros and cons of using PSA and/or its time-dependent derivatives as an endpoint to demonstrate drug effectiveness?</u>

Dr. Scardino said the advantages of PSA as a potential endpoint are that it is simple to measure, easily repeatable, reproduced fairly accurately in different laboratories, and clearly associated with the time course of disease. Because PSA is easily measured, more patients can be enrolled in trials and trials can be completed more quickly. PSA may be a particularly good endpoint for patients at high risk for metastases. However, changes in PSA values may or may not reflect the effect of treatment and PSA alone is therefore not a valid surrogate for survival, although it may be useful when viewed in the context of other disease features.

Dr. D'Amico said the absolute value of PSA is far less important than its rate of change over time. In a study published in the *New England Journal of Medicine* (*NEJM*) on May 27, 2004, a significant number of prostate cancers—about 15% of which were high-grade—were detected in men who never had a PSA level of more than 4.0 ng/ml, the optimal upper limit of normal (Thompson et al, NEJM 2004; 350:2239-2246). Another recent study found that, irrespective of initial PSA value, an annual PSA velocity of more than 2.0 ng/ml was associated with a significantly shorter time to death from prostate cancer and death from any cause (D'Amico et al, NEJM 2004; 351:125-135).

The lesson to be drawn from these studies, as well as from the wealth of other data that exists on PSA kinetics, Dr. D'Amico said, is that PSA must be viewed as a time-dependent parameter (as are mammograms, for example) rather than as a number. The evidence strongly suggests that the man whose PSA rises from 0 to 10 over a 3-week period is at far higher risk than the man whose PSA rises from 10 to 100 over 10 years.

Dr. George commented that although PSA is a good prognostic factor, a number of subtle statistical issues complicate its validation as a surrogate endpoint. He added that PSA values are dynamic over time; PSA testing in effect takes snapshots of a dynamic process at certain time points. However, patterns of PSA change over time are complex and parameters such as doubling time may not present a full picture of that complexity.

Dr. Albertsen referred to a recently published study by his group (Albertsen et al, J Urol 2004; 171(6):2221-2225), which found that patients who died of prostate cancer had a median PSA doubling time of 0.8 years, whereas patients who did not die of prostate cancer within 10 years of diagnosis had either no post-treatment increase in serum PSA or had a PSA doubling time of longer than 1 year. His concerns with the use of PSA doubling time as an endpoint, he said, are that it is unclear how long patients must be

followed to calculate an accurate doubling time and that small biological variations can profoundly affect doubling time.

What PSA endpoints are worthy of consideration for use in oncology trials for (a) recurrence, (b) response, (c) progression?

Dr. D'Amico noted that in recurrence following primary treatment, PSA doubling time may be worthy of consideration as an endpoint because it has now been associated with both time to a positive bone scan and time to death from prostate cancer. For response, PSA velocity may be worthy of consideration in the clinical state of progressive, hormone-refractory disease, in which some evidence suggests it may be a surrogate for all-cause death. A few studies support the use of PSA kinetics as a surrogate for progression; in addition, strong data support the prognostic significance of PSA kinetics in terms of predicting time to a positive bone scan and time to cancer-specific or all-cause death.

Dr. Hirschfeld asked how one might design a randomized study using a PSA-derived endpoint to test a hypothetical immunotherapy that had been shown to affect PSA levels. Dr. D'Amico responded that one would need to have data from a prior study of the same agent in which the endpoint was survival to quantify the association between the putative PSA endpoint and survival. He added that validation of a surrogate endpoint must be disease-state– and drug-class–specific.

Dr. Scher said that a prognostic indicator should not be confused with a surrogate endpoint. Depending on the drug and the disease state, PSA may be used as an initial screening tool to identify a treatment effect. However, phase III trials must be based on established endpoints that offer clinical benefit. Dr. Scher added that some therapies that have been shown to provide clinical benefit, such as bisphosphonates, have no effect on PSA.

Dr. Yao asked whether PSA in combination with tumor shrinkage might be considered an endpoint that is reasonably likely to predict clinical benefit. Dr. Scher responded that in early-phase trials of cytotoxic drugs, the use of a change in PSA permits a study to be brought to an early end if no effect on PSA values is observed. On the other hand, a decline in PSA values is only the first step in demonstrating clinical benefit; a treatment effect on other disease parameters would also have to be observed.

Dr. Hirschfeld suggested that if a trial was comparing drugs with different effects on PSA values, the use of PSA as part of a composite endpoint as a surrogate for progression would complicate the interpretation of the trial's results. Dr. Scher said the fact that the agents being compared have different effects on PSA values should be taken into account in the trial's design.

Should any PSA endpoints be considered validated in any setting?

Dr. Scher said no PSA endpoints can currently be considered to be validated in any setting.

<u>If not, what studies or data are needed to validate PSA as an endpoint in various settings?</u>

Dr. Scardino said studies are needed that will make it possible to define the level of PSA progression following local therapy that predicts that a patient will develop metastatic disease if left untreated. He noted that many patients survive without disease recurrence for 10 years or more without additional treatment despite a slow but steady rise in PSA values. When he and his colleagues applied six definitions of PSA progression to a large data set from patients treated with radiation therapy or surgery, they found that the results of treatment varied widely (from almost uniform success to a high rate of failure) depending on which definition was used.

Dr. Pazdur observed that the existence of therapies that affect established clinical endpoints is a prerequisite for the validation of a surrogate endpoint. To date, in advanced prostate cancer, only one treatment regimen has been shown to have a beneficial effect on survival. Secondly, the validation of a surrogate requires meta-analysis of multiple trials performed in the same disease setting or preferably in multiple disease settings. Finally, validation of a surrogate requires an understanding of both the intended and unintended effects of the investigational therapy on the pathways of the disease. Dr. Pazdur added that it is not FDA's position that a surrogate must fulfill all of the Prentice criteria to be considered validated, but ultimately the agency must be confident that a surrogate is reasonably likely to predict clinical benefit.

Dr. Scher remarked that there continues to be tremendous variation in prostate cancer outcomes that is not explained by time-dependent changes in PSA values and that there were inconsistencies in the docetaxel results with regard to the treatment effect in relation to the surrogate. He added that more trials are needed to answer these questions.

Drs. Scher and Raghavan commented that much larger proportions of patients with breast cancer than with prostate cancer are enrolled into clinical trials. Dr. Raghavan added than in breast cancer drugs were identified and developed first for advanced disease and later moved to earlier disease setting. Only the enrollment of larger numbers of prostate cancer patients into trials will result in trials that are adequately powered to provide definitive answers, he said.

Dr. Williams noted that much of the data concerning time-dependent PSA parameters is retrospective and questioned whether these parameters could be used as clinical trial endpoints without there being a lot of missing data. He suggested that endpoints need to be developed that measure time-dependent PSA phenomena and that can be used prospectively both to guide treatment and to serve as a primary clinical trial endpoints.

Dr. Coffey, from the audience, referred to a recent study in which, for patients with initial PSA values below 2.0, PSA velocity had predicted which individuals would develop metastatic prostate cancer. He added that a follow-up study currently in press will provide additional data on the utility of PSA velocity as a predictor of disease in patients enrolled in prostate cancer prevention trials. He commented that the level of validation being required for PSA is disproportionate relative to the level of validation that has been accepted for other endpoints and for therapeutic modalities such as surgery and radiation therapy.

Dr. D'Amico noted that in one Southwest Oncology Group study, the proportion of treatment effect explained by PSA change was 68%.

A member of the audience said that few improvements in treatment for early prostate cancer have occurred during the past 40 years. An effective surrogate marker is needed that will correlate with both TTP and survival, he said.

Dr. Robert Temple of FDA, from the audience, recalled that in the 1960s there was enormous controversy concerning whether lowering blood pressure was a good surrogate endpoint. The relationship of blood pressure or cholesterol to heart disease outcomes in untreated patients was well known; the controversy centered on whether that relationship persisted when blood pressure was therapeutically altered. Studies eventually showed that lowering blood pressure had a profound effect on a variety of outcomes. He asked whether there are now enough data to show that reducing PSA levels through therapeutic intervention alters outcomes. If so, he said, PSA might be a good endpoint.

Dr. Scardino responded that one of the dilemmas in prostate cancer clinical trials is that the principal treatment for advanced disease, androgen deprivation therapy, does markedly alter the surrogate (i.e., PSA). Dr. D'Amico, however, disputed whether hormonal therapy confounds the relationship between PSA and outcome. He noted that in a patient with a short PSA doubling time, response time, time to distant treatment failure, and time to death are all short, whereas in a patient with a long PSA doubling time all these parameters are extended.

Dr. Scher asked how much can be learned from retrospective studies and meta-analyses, how many randomized trials are necessary for validation of a surrogate, and how many randomized trials of hormonal therapy in prostate cancer are available for analysis. Dr. Eisenberg commented that trials conducted before the PSA era cannot provide data on the relationship between PSA values and outcomes. Dr. Ellenberg said two randomized trials may be sufficient if the results are clear and consistent and supported by observational data. Although it is necessary to have trials that show an effect, it can also be useful to look at negative trials because a treatment might affect PSA levels but offer no survival benefit. Dr. Raghavan observed that positive trials are more likely to be published and agreed that analyses must include both negative and positive trials.

The workshop was adjourned for lunch.