DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

ONCOLOGIC DRUGS ADVISORY COMMITTEE

ENDPOINTS IN CLINICAL CANCER TRIALS

AND

ENDPOINTS IN LUNG CANCER CLINICAL TRIALS

Tuesday, December 16, 2003

8:05 a.m.

Advisors and Consultants Staff Conference Room
5630 Fishers Lane
Rockville, Maryland

PARTICIPANTS

Donna Przepiorka, M.D., Ph.D.
Johanna Clifford, M.S., RN, BSN, Executive Secretary

MEMBERS:

John T. Carpenter, Jr., M.D.
Bruce G. Redman, D.O.
Sarah A. Taylor, M.D.
Otis W. Brawley, M.D.
Stephen L. George, Ph.D.
Bruce D. Cheson, M.D.
Gregory H. Reaman, M.D.
James. H. Doroshow, M.D.
Pamela J. Haylock, RN (Consumer Representative)
Alexandra M. Levine, M.D.
Maria Rodriguez, M.D.

CONSULTANTS (VOTING):

Philip Bonomi, M.D.
David Ettinger, M.D.
Thomas Fleming, M.D.
Bruce Johnson, M.D.
David Johnson, M.D.
Scott Saxman, M.D.

PATIENT REPRESENTATIVES (VOTING):

Michael S. Katz
Sheila Ross

ACTING INDUSTRY REPRESENTATIVE (NON-VOTING):

Antonio Grillo-Lopez, M.D.

GUEST SPEAKERS (NON-VOTING):

Paul Bunn, M.D.
Richard Gralla, M.D.

FDA:

Robert Temple, M.D.
Richard Pazdur, M.D. (by telephone)
Martin Cohen, M.D.
Grant Williams, M.D.
Patricia Keegan, M.D.
Ning Li, Ph.D.

C O N T E N T S

P R O C E E D I N G S

Call to Order

DR. PRZEPIORKA: Good morning to all. I would like to call the meeting to order. This is a meeting that is covering no drug evaluations but, in fact, methods for drug evaluations. I think it is a good time for this talk because there are very new types of drugs coming out for which these issues may be very germane.

I would like to start the meeting by an introduction of the committee members, if we could start with Dr. Grillo-Lopez and just go around. Let us know who you are and where you are from.

DR. GRILLO-LOPEZ: My name is Antonio Grillo-Lopez. This is my first time sitting around this table. I am a hematologist/oncologist. I spent half of my career in industry and half in academia so I am hoping to make some positive contributions here. Thank you.

DR. GEORGE: Stephen George, from Duke University.

DR. CHESON: Bruce Cheson, Georgetown University, Lombardi Comprehensive Cancer Center.

DR. DOROSHOW: Jim Doroshow, City of Hope Comprehensive Cancer Center.

DR. RODRIGUEZ: Maria Rodriguez, M.D. Anderson Cancer Center in Houston, Texas.

DR. BRAWLEY: Otis Brawley, Emory University, Winship Cancer Institute.

MR. KATZ: Michael Katz. I am a 13-year myeloma survivor.

DR. FLEMING: Thomas Fleming, University of Washington.

DR. LEVINE: Alexandra Levine, University of Southern California, Norris Cancer Center.

DR. REAMAN: Gregory Reaman, Children's Hospital and George Washington University.

DR. PRZEPIORKA: Donna Przepiorka, University of Tennessee Cancer Institute.

MS. CLIFFORD: Johanna Clifford, FDA, Executive Secretary to this meeting.

MS. HAYLOCK: Pamela Haylock, oncology nurse and doctoral student in Galveston, Texas.

DR. CARPENTER: John Carpenter, medical oncologist, University of Alabama at Birmingham.

DR. REDMAN: Bruce Redman, University of Michigan Comprehensive Cancer Center.

DR. TAYLOR: Sarah Taylor, University of Kansas Medical Center.

DR. LI: Ning Li, FDA Biometrics.

DR. WILLIAMS: Grant Williams, Deputy Director, Oncology Drug Products.

DR. PRZEPIORKA: Thank you to all.

DR. WILLIAMS: And on the phone, of course, is Dr. Pazdur.

DR. PAZDUR: Hi. I hope you don't hear the dog barking.

DR. WILLIAMS: I was going to say that this was the first time that Dr. Pazdur has ever been speechless--

DR. PAZDUR: And you love that, Grant!

[Laughter]

DR. PRZEPIORKA: Welcome and, Dr. Pazdur, thank you for joining us. We would like to move now to the reading of the conflict of interest statement.

Conflict of Interest Statement

MS. CLIFFORD: The following announcement addresses the issue of conflict of interest with respect to this meeting and is made a part of the record to preclude even the appearance of such at this meeting.

Based on the agenda, it has been determined that the topics of today's meeting are issues of broad applicability and there are no

products being approved at this meeting. Unlike issues before a committee in which a particular product is discussed, issues of broader applicability involve many industrial sponsors and academic institutions.

All special government employees have been screened for their financial interests as they may apply to the general topics at hand. To determine if any conflict of interest existed, the agency has reviewed the agenda and all relevant financial interests reported by the meeting participants. The Food and Drug Administration has granted general matters waivers to the special government employees participating in this meeting who require a waiver under Title XVIII, United States Code Section 208. A copy of the waiver statements may be obtained by submitting a written request to the agency's Freedom of Information Office, Room 12A-30 of the Parklawn Building.

Because general topics impact so many entities it is not prudent to recite all potential conflicts of interest as they apply to each member, consultant and guest speaker. FDA acknowledges that there may be potential conflicts of interest but, because of the general nature of the

discussion before the committee, these potential conflicts are mitigated.

With respect to the FDA's invited industry representative, we would like to disclose that Dr. Antonio Grillo-Lopez is participating in this meeting as the acting industry representative, acting on behalf of regulated industry. Dr. Grillo-Lopez is employed by Neoplastic and Autoimmune Disease Research.

In the event that the discussions involve any other products of firms not already on the agenda for which FDA participants have a financial interest, those participants' involvement and their exclusion will be noted for the record. With respect to all other participants, we ask in the interest of fairness that they address any current or previous financial involvement with any firm whose product they may wish to comment upon. Thank you.

DR. PRZEPIORKA: Thank you. The first item on the agenda then is the opening remarks. Dr. Pazdur, will you be making those opening remarks?

DR. PAZDUR: Why don't we have Dr. Williams do that?

DR. PRZEPIORKA: Dr. Williams?

Opening Remarks

DR. WILLIAMS: Just a few remarks. First of all, we are just very appreciative of all of your presence here today to give us advice. I think we are actually pretty excited about the whole process of getting endpoints out and discussed. For us it is a very difficult problem. We have multiple end of Phase II meetings, multiple different clinical settings and trying to be consistent with the endpoints that we require for drug approval across these many settings is quite a challenge.

This reflects a process that we started about a year ago of looking into endpoints, or even before that internally, and our plan in this process is to have a series of workshops, a series of ODAC meetings on specific clinical settings. We have engaged the National Cancer Institute, AACR and ASCO to help us with picking experts in the field to do workshops on very specific endpoint settings and we plan to follow these with ODAC meetings, and this is the first after these workshops. We had a lung cancer workshop in I think March or April and then this afternoon we

plan to have discussions on lung cancer endpoints.

As we thought about moving toward creating a guideline or guidances we also considered that we should have some sort of a broad discussion to sort of set the foundation, and then also to lay the foundation for a background section of the guidance. So, that is what we are trying to do here this morning. This afternoon we would like some voting on some specific questions. As we go along we will try to determine those that seem appropriate for voting.

But this morning it is more of a broad discussion that we are looking for. What are those principles that we should be evaluating as we move forward to evaluate endpoints? What are those value judgments globally so that we can then apply them to specific instances, specific clinical settings?

So, we look forward to the discussion today. I think it is going to be very interesting and fun. The first talk will be by Dr. Farrell, who will talk about regulatory considerations with endpoints in oncology.

General Regulatory Background

DR. FARRELL: Good morning, everyone.

[Slide]

I am here to discuss regulatory considerations for endpoint used for approval. Requirements for marking approval have been codified and further defined in response to perceived need. Prior to 1938 there were no requirements for marketing approval. As a result of the sulfonamide tragedy, Food, Drug and Cosmetic Act required manufacturers to provide evidence that their product was safe for marketing.

In 1962 Congress, concerned about misleading and unsupported claims being made about marketing products, amended the FDAC to require that manufacturers provide evidence that the product was effective. This was to demonstrate substantial evidence of effectiveness. In the practice the agency has understood that adequate and well controlled investigations or substantial evidence of effectiveness means that efficacy must be demonstrated in at least two adequate and well-controlled trials.

In 1997 Congress passed the Food and Drug Modernization Act which stated that the requirement for substantial evidence of effectiveness could constitute one adequate and well-controlled trial

plus supportive evidence.

[Slide]

There are two basic mechanisms for approval, regular and accelerated approval. The requirement for adequate and well-controlled studies is the same for both mechanisms. The regular approval mechanism provides for approval based on clinical benefit or on an established surrogate for clinical benefit.

The clinical benefit endpoint is usually an endpoint thought of as reflecting quality or quantity of life. In oncology, examples of these endpoints include survival or improvement in a disease-related symptom.

Accelerated approval is a mechanism for those products designed to be used for the treatment of serious and life-threatening illness. The mechanism provides for approval based on a surrogate that is deemed reasonably likely to predict clinical benefit. The new therapy must provide an advantage over available therapy, and that can be the ability to treat patients who are unresponsive to or intolerant of available therapy, or it can be a therapy that provides an improvement patient response over available therapy.

[Slide]

The accelerated approval mechanism, as I said, is based on a surrogate endpoint believed to be reasonably likely to predict clinical benefit or it can be based on an effect on a clinical endpoint other than survival or irreversible morbidity. In any case, post-marketing studies are required to determine clinical benefit.

[Slide]

The evidence for accelerated approval should be substantial evidence from well-controlled clinical trials regarding a surrogate endpoint, not borderline evidence regarding a clinical benefit endpoint in a poorly conducted trial.

[Slide]

As I stated before, ideally the substantial evidence should come from more than one adequate and well-controlled investigation. The passage of FDAMA allows us to consider the evidence from one adequate and well-controlled trial plus other supportive evidence. The effectiveness guidance discusses supportive evidence and the characteristics of the single trial.

[Slide]

This slide outlines examples of situations

where extrapolation from existing studies combined with a single clinical trial could support a new indication or new drug application. In pediatrics, if there is bioequivalence in modified-release dosage form, for different doses or for different regimens.

[Slide]

The effectiveness guidance lists the characteristics of a single trial supporting approval. In general these trials should be large, multi-center. The primary results should show consistency across study subsets. This could be thought of as various age categories. The study should be large enough so it could be considered to have multiple studies in a single study, and that could be done through a factorial design. And, the results from secondary endpoints, if positive, could also be supportive for the use of that single trial. The primary endpoints should show statistically persuasive results.

[Slide]

In oncology we have accepted oncology supplemental applications based on a single trial supported by data in a different stage of disease. The FDA has approved cancer drug supplements in an

NDA in an adjuvant setting when there has been a single trial plus supportive evidence in a metastatic setting. One example of this would be Irimidex from the adjuvant treatment of women who are postmenopausal. We have also accepted applications in first-line settings with one trial when there has been supportive evidence based on approval in a refractory setting. An example of that is Gleevec.

In addition, we have accepted applications for the use of products in combination therapy when there has been an approval in a monotherapy setting. An example of that would be Zoloda in combination with Taxotere when Zoloda had already received approval as monotherapy in the treatment of breast cancer.

Theoretically, we could accept an application and approve it based on a single trial in a second cancer if there was already an approval in a closely related cancer.

[Slide]

In summary, the agency has some flexibility in judging what constitutes adequate information to meet its requirements of substantial evidence from adequate and well-controlled

investigations. However, all products must demonstrate that they are both safe and effective. Because oncology is a serious and life-threatening illness we have actually two mechanisms for approval, regular and accelerated approval.

Accelerated approval can be based on a surrogate endpoint with planned completion of a post-marketing study to verify the clinical benefit. Approval can also be based on one trial plus supportive evidence. Endpoints differ for different approval mechanisms. Drs. Dagher and Williams will discuss this issue in greater detail. Thank you.

DR. PRZEPIORKA: Thank you very much, Dr. Farrell. Next, Dr. Dagher will be talking about endpoints for past approvals.

Endpoints for Past Approvals

DR. DAGHER: Good morning.

[Slide]

In the next few minutes I would like to summarize endpoints used for approval of oncology drugs.

[Slide]

This slide provides a summary of endpoints commonly used in the oncology clinical trial

setting. Survival has been considered the gold standard in many settings and provides an unambiguous endpoint that is easily measured. Time to progression may provide several advantages as well as challenges, which Dr. Grant Williams will discuss later this morning. Disease-free survival is an endpoint utilized in the adjuvant setting. Objective tumor response is an endpoint that measures an effect largely related to treatment, independent of the natural history of the disease. Tumor-related symptoms and patient-reported outcomes are quite relevant from the patient's perspective.

[Slide]

For the purposes of regular approval we have considered improvements in survival or tumor-related symptoms as evidence of clinical benefit. In the adjuvant breast cancer setting we have also considered disease-free survival as evidence of clinical benefit.

[Slide]

In some settings, where tumor shrinkage has been associated with symptom benefit or survival, we have considered objective tumor response as an endpoint supporting regular

approval. In leukemias and some solid tumors, such as testicular cancer, durable or complete responses have been utilized for this purpose. In the case of hormonal therapies for breast cancer partial responses have been considered evidence of clinical benefit.

[Slide]

A summary of endpoints and approvals from our Division, published in The Journal of Clinical Oncology, reveals that more than half of the approvals have been based on endpoints other than survival. This applies to all approvals as well as those excluding accelerated approval, a setting in which response rates are often utilized.

[Slide]

The following table, adapted from this publication, illustrates the diversity of endpoints used. For approvals between 1990 and the end of 2002 in the Division of Oncology Drug Products survival was used in 18 of 55 approvals. Response rate, either alone or in conjunction with improvements in tumor symptoms or time to progression, was utilized in 26 approvals. As discussed, improvement in tumor-related symptoms has been used as a basis for approval.

Disease-free survival or other endpoints were used infrequently.

[Slide]

The first two bullets of this slide provide examples where improvement in tumor-related symptoms was the basis for regular approval. In patients with advanced hormone refractory prostate cancer a pain scale was utilized to evaluate mitoxantrone plus prednisone versus prednisone alone.

Photofrin was evaluated for obstructive esophageal lesions. In this case a dysphasia scale was used with supportive evidence for objective tumor response.

In the case of several bisphosphonates approval was based on evaluation of a number of skeletal related events, including pathologic fracture, radiation to bone, surgery to bone or spinal cord compression. In the case of prostate cancer, pain requiring change and anti-neoplastic therapy was also a component of the evaluation.

[Slide]

As Dr. Farrell mentioned, accelerated approval is based on a surrogate endpoint reasonably likely to predict clinical benefit. In

our experience, most of the accelerated approval indications were based on an evaluation of objective tumor response in studies without an active comparator, that is, single-arm studies or those comparing two dose levels of the drug in question. However, randomized trials were conducted in some settings with an active or placebo comparator, allowing for evaluation of time to event endpoints such as disease-free survival or time to progression. Some examples are shown here.

[Slide]

As was also discussed, accelerated approval requires further evaluation of the drug to confirm clinical benefit. Therefore, two strategies have emerged for approaching accelerated approval and subsequent confirmatory evaluation of clinical benefit.

With the first strategy accelerated approval is based on response rate evaluated in single-arm studies of refractory patients and confirmatory studies are conducted in related populations such as those with less refractory disease. This approach has the potential advantage of allowing rapid completion of single-arm studies.

[Slide]

However, accelerated approval may influence the ability to enroll patients for confirmatory studies. Furthermore, it has become more and more challenging to evaluate marginal benefits in more and more refractory populations, and findings in refractory populations may not be relevant to other populations which may benefit from the drug. In fact, evaluation in refractory populations first may lead us to miss an active drug. The single-arm component of the strategy is associated with its own limitations: First, an inability to evaluate time to event endpoints in a non-randomized setting and difficulty in completely assessing the toxicity profile.

[Slide]

The second strategy for accelerated approval depends on evaluation of a surrogate endpoint and an interim analysis of a randomized study, with subsequent evaluation of clinical benefit in the same trial using a final analysis. This approach allows for evaluation of the same population for accelerated approval and regular approval and facilitates completion of a confirmatory study. The randomized setting allows comparison to available therapy and a thorough

evaluation of the toxicity profile.

[Slide]

However, this approach may require more time and patients than single-arm studies and accelerated approval could still influence completion of the study.

[Slide]

In summary, improvements in survival or tumor-related symptoms have been considered evidence of clinical benefit. In some settings durable, complete or partial responses have been considered endpoints supporting regular approval. Finally, objective tumor responses in single-arm trials have been the basis of approval in most cases of accelerated approval. Thank you.

DR. PRZEPIORKA: Thank you Dr. Dagher. We are going to hold questions until the end of the presentations and Dr. Williams will now talk to us about selected issues in oncology trial designs that are pertinent to this morning's topic.

Selected Issues in Oncology Trial Design

DR. WILLIAMS: Well, thank you, Dr. Przepiorka.

[Slide]

Members of the committee, ladies and

gentlemen, what I would like to do is to first review the selected issues in oncology trial design before we go to discussing specific problems and your recommendations for our further deliberations.

[Slide]

Here is the outline of my presentation. I will begin with several difficulties we face in oncology that are well-known to all of you, and I will briefly discuss the non-inferiority trial design and the difficulties we face with this approach. Finally, I will discuss time to progression, expanding upon some of the regulatory issues presented by Dr. Farrell and Dr. Dagher, especially the issues relating to the meaning of clinical benefit and also surrogates for clinical benefit. Then I will discuss the pros and the cons of TTP as an approval endpoint.

[Slide]

During our end of Phase II meetings with sponsors we often ask whether trials can be blinded and we are usually told they cannot. These are the reasons that we are told, first, that there are toxic side effects that are said to unmask both the physician and the patient. Second, the investigators adjust doses based on drug-specific

toxicities and the investigators believe they need to know drug assignment to do this safely. These seem to be very difficult problems, although I think maybe the first point might bear some further discussion--has anyone actually studied the degree of unmasking by side effects of oncology drugs? As we move to new potentially targeted therapies and to oral therapies we should consider whether we can blind more trials.

[Slide]

Placebos are widely used in many areas of drug development. The use of the placebo is seldom feasible in evaluation of advanced cancer. There are some cancer settings where placebo use may be possible. Blinded, placebo-controlled studies might be performed in some early disease settings where no effective treatments exist. In advanced settings the so-called add-on design can allow placebo use comparing drug A plus placebo to drug A versus drug B. In some settings it may be reasonable to continue placebo and drug B even beyond progression. An example of this were the bisphosphonate trials which assessed effects on bone morbidity even after chemotherapy was changed.

[Slide]

So, the unfortunate result of not having blinded, placebo-controlled studies is that we must use controls which are active. If we use a superiority trial design the new drug must beat the active drug, or we can use an add-on design. Not surprisingly, many trials for drug approval are based on drug combinations and add-on designs. Certainly, this can lead to toxic combinations.

The other possibility is to do non-inferiority studies. As I will discuss, these tend to be very large trials and the quality of historical data in oncology is frequently insufficient to support this approach. Again unfortunately, in this setting where blinded, placebo-controlled trials may not be feasible it is very difficult to demonstrate the new drugs are less toxic but have similar efficacy to an approved drug.

[Slide]

The frequent use of drug combinations in oncology also present regulatory challenges. Since marketed approval is for a single drug rather than a combination of drugs, trials supporting regulatory approval need to isolate the effectiveness of the proposed agent. Evidence is

needed showing not only the effectiveness of the combination but also establishing that there is a contribution of the new drug to that regimen.

[Slide]

Now I would like to turn to the topic of non-inferiority. Obviously, I am not a statistician but I will try to share with you what I understand about it. The reason we are not having statisticians do this discussion is because we don't want to be at this a whole day on non-inferiority.

[Laughter]

[Slide]

So, here is the way I see it. First I want to review some non-equivalent words. I don't know if anybody caught the pun in the title here. First of all, we love superiority. We love to hear the word superiority; we love superiority trials. Equivalence is a word you should never say to a statistician, but I was corrected on this, it is all right to say it to a Bayesian.

[Laughter]

Equivalence is something that can never be proven. Because we cannot show equivalence we rule out inferiority by a prespecified margin. We call

this demonstration of non-inferiority. A very important regulatory concept is that proof of non-inferiority does not necessarily prove efficacy, and we will discuss this a bit further. I think the use of these words in our oncology journals can create serious misconceptions. A common problem is the assumption in oncology journals that no statistical difference is the same as equivalence or non-inferiority.

[Slide]

This slide lists the steps needed to perform a non-inferiority analysis. Just the number of steps should suggest the complexity of this process and the potential for error. In this example we are demonstrating that drug B is effective. In order to do this we refer to the effect of drug A observed historically in randomized studies. I think I have these steps out of order; I will stick to the third one.

We then prospectively identify a margin that includes an acceptable fraction of drug A's efficacy. We randomized drug A versus drug B. We prove that drug B is no worse than drug A by that margin. Probably the step that is most often ignored is that we determine that the constancy

assumption is valid. Invalid assumptions at any stage of this process could lead to a false result and this is why non-inferiority studies are not FDA's favorite trial design.

[Slide]

The important constancy assumption is the historically observed drug effect of the active control drug also exists in the current non-inferiority trial and in the population. The problem is that conditions are never the same in historical trials and a current trial. Differences include different populations; differences in supportive care; differences in availability of new drugs that can be taken after failing, including the possibility of crossover. Finally, the designs can be different with different frequency of follow-up. So, any of these could change the sensitivity of the trial to detect the treatment effect. The serious result of violating that constancy assumption could lead to the approval of what has been termed a toxic placebo.

[Slide]

This is another property of non-inferiority trials that Dr. Temple has noted, sloppiness obscures the observations of

differences. For superiority trial designs sloppiness obscures efficacy but for non-inferiority trials sloppiness could lead to a false efficacy claim. Again, this is why we like superiority trials. I think that is a common theme you will be hearing here perhaps.

[Slide]

A critical problem in doing non-inferiority studies in oncology is the paucity of studies that are available to determine the historical effect of the active control drug. We basically strike out at the first step of this process. What we really need is multiple trials showing a consistent, large effect and we need to perform a meta-analysis of those trials which provides us with a dependable effect precisely estimated.

The real situation in oncology, almost without exception, is that we have one or two rather small trials with small effects and with marginal statistical significance. This leads to small historically documented effect sizes; small margins; and very large non-inferiority studies. The process becomes even more complicated when we consider drug combinations and the contribution of

individual drugs to historical effect.

The reason I am presenting this is that I think this is such a complex topic and people don't understand why you don't do a non-inferiority study. I don't think you can say it without trying to go through all these steps, but it is basically just not possible in many of our settings at least using the primary endpoints.

[Slide]

Now I would like to turn to endpoints and surrogates. Dr. Farrell and Dr. Dagher provided an overall review of regulations on oncology endpoints. So, I want to briefly review the history of regulatory standards for efficacy endpoints.

The 1962 amendments to the FD&C Act simply stated that a drug must be shown to have the effect claimed in the label. However, subsequent judicial decisions established that effectiveness meant that the drug must have clinical meaning. In the 1970s marketed applications for cancer drugs were approved primarily based on objective response rates and on rather minimal activity we would say today.

However, based on advice from ODAC in the

late '70s and early '80s, FDA determined that the response rate should generally not be the sole basis for drug approval because the possible benefits associated with tumor shrinkage did not necessarily justify treatment with toxic anti-cancer drugs. Acceptable endpoints for drug approval were improvement in survival or improvement in physical functioning or relief of pain.

As Dr. Dagher discussed, in the 1990s FDA struggled with the difficulty of measuring patient benefit and in some settings found various surrogates to be adequate in specific clinical situations.

[Slide]

There are various definitions for a surrogate. In this context we will use the definition from Dr. Temple. A surrogate endpoint of a clinical trial is a laboratory measurement or a physical sign used as a substitute for a clinically meaningful endpoint that measures directly how a patient feels, functions or survives. Changes induced by a therapy on a surrogate endpoint are expected to reflect changes in the clinically meaningful endpoint.

[Slide]

In various settings for many years FDA has based regular drug approval on surrogate endpoints which were judged by FDA and experts in the field to be reliable indicators of clinical benefit. Examples outside the field of oncology included blood pressure, blood sugar and blood cholesterol.

[Slide]

It may be useful to review where we have used the term surrogate in oncology. In accelerated approval the surrogate need only be reasonably likely to predict benefit. Obviously, this is a lower standard than the usual use of the word surrogate.

We have discussions with statisticians--Dr. Fleming, regarding validated surrogates and we expect to prove quantitatively the relationship between the surrogate and the established endpoint. Unfortunately, in oncology we have very few settings where we quantitatively validate the surrogate. It would be easier to validate surrogates if we had more effective drugs with large effects to compare surrogate and clinical benefit. Finally, we have surrogates that have been used to support regular approval of

cancer drugs in very specific settings, usually based on clinical inference and judgment that these surrogates relate to clinical benefit.

[Slide]

At the recent colon cancer workshop Dr. Fleming reviewed Prentice's criteria for strictly validated surrogates. The surrogate endpoint must be correlated with the clinical outcome. The surrogate must fully capture the net effect of the treatment on that clinical outcome.

[Slide]

In the clinical setting this would involve meta-analyses of clinical trials and a comprehensive understanding of the disease and the intended an unintended effects of drugs. As I stated, where possible this is the kind of evidence we would like for a surrogate endpoint. The question for us today is what should we do with endpoints we have today? What can we use for approval endpoints today and in what settings can we use them? And, what can we do to gather more data for the future?

[Slide]

As we looked at TTP to ask whether it is an acceptable surrogate in various settings, I

propose that the question we should ask should not be whether an improvement in TTP has clinical meaning. I suggest that nobody in the field of oncology really doubts that it is good to delay the growth of cancer. That is not really the question that we need to answer.

[Slide]

The real question is whether you can reliably measure TTP and, if you can, what does it mean? How much delay in progression is worth how much toxicity? With survival we seldom quibble about the size of the effect. Given the low statistical power of our studies, a statistically convincing survival benefit is generally considered to be worth the toxicity of treatment. However, can we say the same for the delay in TTP? That is, when progression is determined by only images on a scan. So, the real question is how do we trade off a TTP benefit compared to drug toxicity?

Another question is the relative value of treatments evaluated by different endpoints. When a well-established survival benefit exists for an approved drug what is the meaning of the claimed TTP effect for an investigational drug? Although two treatments are not required to have equal

efficacy this is, nonetheless, an important consideration for us.

[Slide]

FDA's approach to endpoints for hormonal treatment of cancer illustrates how clinical judgment has played a role in the acceptance of surrogates for regular drug approval. For many years these drugs have been approved primarily based on comparison of response rates with two reasonably large, randomized, controlled studies. TTP and survival were assessed as secondary endpoints. Many hormonal drugs have been approved with this approach. I think that everybody is satisfied that we approved effective drugs through this approach.

So, what allowed this approach? These are what I believe are the critical factors. We have a long experience with tamoxifen and, despite little data with regard to a survival or TTP benefit, tamoxifen was widely observed to provide benefit to patients. The main indicator of activity was response rate. Given the non-toxic nature of the drugs and similar mechanisms of action, response rates seemed a reliable indicator of clinical benefit in this setting.

[Slide]

Four years ago at ODAC we discussed TTP as an approval endpoint for first-line cytotoxic treatment of breast cancer. The committee was not supportive of TTP for regular approval but did suggest its use for accelerated approval. Prominent in the ODAC deliberations was whether the standard treatment doxorubicin produces a survival effect and, if so, what size is that benefit. Committee members noted that current treatments only produce small TTP effects and they questioned whether there was or was not a correlation between TTP and survival, whether it was reliable. As I note in later discussion, I think this question needs to be carefully evaluated because of the under-powered nature of most of our studies.

Questions were also raised about the reliability of TTP measurement and also a claim that in order to measure TTP accurately frequent scans would be needed. So, the ODAC criticisms were varied and they addressed the data available at the time in the specific cancer setting.

[Slide]

So, I would like to take a closer look at TTP. First of all, what is TTP? The basic

definition is time from randomization to documented progression. However, there are very many different definitions of TTP with a lot of different details, such as how do you handle missing data and how to censor. If TTP is to be used as an important endpoint there should be careful agreement between FDA and the sponsor on the protocol, case report form and the statistical analysis plan. Difficult issues include how to follow the patient for new lesions and how to define and validate progression of non-measurable disease.

[Slide]

I want to mention three TTP-like endpoints that we frequently encounter, time to progression, progression-free survival and time to treatment failure. For TTP the measured event is progression. TTP may be thought of as a measurement of anti-tumor activity. Patients going off study for toxicity and non-tumor deaths are not counted as events. Note that for non-tumor deaths censoring occurs at the last visit where TTP was evaluated. This censoring makes the assumption there is no relation between death and progression, an assumption that might be questioned.

[Slide]

With progression-free survival all deaths are counted as progression events. Dr. Fleming suggested at the recent colon cancer workshop if TTP is being considered as a clinical benefit surrogate, perhaps the deaths should be counted. FDA has often counseled sponsors to keep TTP and death separate however, that is, to measure TTP without the deaths and to measure deaths in the survival analysis. The main concern with including deaths is that patients lost to follow-up will subsequently be counted as progression events at the time of death. In such a scenario sloppy progression to follow-up leads to longer progression times and asymmetric follow-up of such cases could lead to a false result. If deaths are included in the analysis, then careful symmetric follow-up is needed. Perhaps we need analysis rules to deal with patients who have inadequate follow-up.

[Slide]

Time to treatment failure is a composite endpoint measuring time from randomization to discontinuation of treatment for any reason, including progression, treatment toxicity and

death. Because it combines elements of safety and efficacy, TTF is not an acceptable endpoint for documenting efficacy. Time to treatment failure has not supported drug approval.

[Slide]

Let's look more closely at TTP as a potential regulatory endpoint. Here as some of the positive qualities of TTP. TTP is measured in all patients and might, therefore, be a better measure of overall benefit than response. TTP does not require massive tumor shrinkage and might be a better measure for metastatic agents.

From a practical standpoint, progression is often the reason oncologists change therapy. Therefore, an advantage of TTP is that TTP is measured before patients cross over to other therapies. This is of growing importance as we develop more effective drugs. Moreover, because progression often occurs months to years before death much smaller studies may be needed to study TTP than survival and this can vary dramatically with the different diseases.

Finally, some would argue that delaying progression has face validity as an indicator of benefit. The benefit seems obvious because

progression is a necessary step between cancer growth, patient morbidity and death.

[Slide]

But here are some problems with TTP. It has been said that it may not correlate with survival. It is an indirect measure of clinical benefit, sometimes reflecting minor changes on a radiograph. Therefore, small differences in TTP may be of unclear clinical value, especially when one is evaluating toxic treatments.

There are obvious concerns relating to ascertainment bias in unblinded trials, and there are concerns regarding the reliability of a small effect with the kind of trials we have today with monitoring schedules which may vary from patient to patient. Finally, careful assessment of progression at frequent intervals is labor intensive and expensive.

[Slide]

We encounter difficulties in determining the exact relationship between TTP and survival. First of all, there are many different cancer settings so the database for any one setting may not be large and it isn't clear when you can combine data across different cancers. Secondly,

unfortunately, we don't have many treatments that produce large survival effects.

A fundamental difficulty is that there is always more statistical power for the analysis of TTP than survival. On this basis alone even if TTP were a perfect surrogate one would expect some studies to show a statistically positive TTP benefit without a statistically positive survival benefit. Oncology studies are virtually never large enough to rule out a meaningful survival effect and, thus, individually cannot establish a lack of correlation.

Finally, there is the crossover issue. Even if TTP were a perfect surrogate for survival, crossover to other effective therapies could prevent detection of a potential benefit.

In summary, with the trials of the size we usually see in oncology or therapies of only marginal benefit it would be difficult to determine the exact relationship between TTP and survival.

[Slide]

In reviewing these slides from the 1999 ODAC, I came upon this one. Dr. Johnson I thought did a really good job of summing up a comparison of survival and TTP. Survival time is precisely

determined regardless of follow-up. Survival is a known entity. On the negative side, survival takes longer to assess, needs larger trials and its benefit can be obscured by secondary therapy.

[Slide]

TTP is only a surrogate, not a direct measure of clinical benefit. Later today during your deliberations we want to hear your thoughts on the important factors FDA should consider when evaluating TTP as a surrogate for clinical benefit in specific settings. For instance, would TTP be more acceptable in cancer settings where symptoms occur at the time of or soon after progression? What TTP benefit increment would be persuasive? How important is the toxicity of treatment in evaluating a TTP benefit? Finally, to what extent is the benefit of other available drugs important? For instance, what if other drugs produce a substantial survival benefit?

One approach to the problem of TTP measurement has been to convert TTP to a direct measure of clinical benefit by measuring time to worsening of cancer symptoms. For years FDA has suggested this endpoint to sponsors at the end of Phase II meetings. However, sponsors and

investigators have cited several problems with this approach. First, there is the ever-present problem of lack of blinding and potential bias thus the endpoint may not be reliable. Another problem is the usual delay between the time of objective progression and the onset of cancer symptoms. Often alternative treatments are begun before reaching the symptom endpoint. At our colon cancer workshop Dr. Langdon Miller presented data suggesting that in colon cancer there is a fairly long time lag between progression and onset of symptoms. When alternative treatments are begun prior to symptom progression the issue of confounding effects arises, just as it does in analysis of survival.

[Slide]

We must remember a critical difference between analyses of survival and tumor progression. The date of death, represented by the star in this cartoon, will not change regardless of the evaluation schedule or censoring. For progression measurement, however, the date we assign for progression is usually the date of a scheduled visit occurring some time after the actual progression date. It should not be surprising that

assessing progression at longer intervals leads to longer time to progression and that asymmetry in this process could lead to bias.

[Slide]

With measurements repeated over many visits assessment of TTP by traditional methods is difficult and labor intensive. Many problems are encountered by FDA during reviews such as not all lesions being followed, or extra scans being performed, or measurements being missing. So, how do you assure equal measurement? How do you assess the impact of bias? How do you verify progression of evaluable disease by unblinded investigators? These are the difficult issues for review of TTP data.

[Slide]

One approach to making progression assessment practical and reliable would be to consider different progression endpoints. An approach that seems worthy of research is to assess progression at only a single time point. This would considerably decrease the burden in the amount of data collected and eliminate the concern of time-related assessment bias. Scans would need to be evaluated only at baseline and either to

document progression for that time or at the prespecified time to document stable disease.

[Slide]

Progression measured at a single point would be much easier to audit and verify, needing only two sets of scans per patient and time-related bias, as mentioned, would be minimized if not eliminated.

So, I think research into approaches such as this would be of great interest to identify the benefits and problems. In this case you would certainly lose some statistical power, requiring larger studies. There would be concern that you would miss a transient TTP benefit if you hit the wrong point with your single time analysis, and we would lose the information we are used to seeing about other parts of the curve, such as the early effects or the potential benefit of a plateau.

[Slide]

In conclusion, here are some issues you may wish to consider in your deliberations. As FDA proceeds with the workshops and meetings on endpoints for cancer treatment settings, is TTP ready for active consideration as a drug approval endpoint? If so, what are the factors that

determine the acceptability of TTP as a drug approval endpoint? What amount of TTP evidence would be needed to support a TTP claim, such as number of trials, value, magnitude and precision of TTP benefit?

[Slide]

And, can we improve our approach? Do we need research on novel progression endpoints such as a single point analysis? Do we need research on the association between TTP and survival data to validate TTP as a survival surrogate? Should we develop an approach to TTP endpoint definition and censoring methods that are standard? Do we perhaps need a separate workshop just to concentrate on TTP methodology? Can more trials be blinded? Does independent blinder radiologic review improve endpoint assessment? And, can symptoms be incorporated into the endpoint?

So, this ends my presentation. I think what we will do is take questions from our seats and just briefly introduce the questions at the beginning of the question discussion rather than to do it now. How long do we have for questions?

Clarification Questions to the Presenters

DR. PRZEPIORKA: Two hours, just for

clarification or the actual questions? Until the break--about 20 minutes. We have the floor open now for questions for the presenters for this morning.

I have a question for Dr. Williams. Just for a point of clarification, for non-inferiority you are not truly looking for non-inferiority per se in terms of the response but it has to be non-inferior in terms of its treatment effect as well as less toxic to be a real winner in that sort of design.

DR. WILLIAMS: Well, let me start with just non-inferiority in general. It just means that you have met your margin. Okay? Non-inferiority for the FDA means that you have met your margin and that margin means the drug works. It is a separate judgment about whether you are less toxic; I mean about the risks and benefits. But there wouldn't be a direct requirement to be less toxic from our regulations, I don't think.

DR. PAZDUR: I think a lot of people confuse that issue of toxicity and non-inferiority since several applications came in dealing with perceived less toxic drugs and comparing them to a standard drug. But, as Grant said, the toxicity

evaluation is different. Many times what we actually see is not really less toxic drugs but a different spectrum of toxicity, and that is another thing that people have to consider also when they are evaluating toxicity.

DR. WILLIAMS: We have never applied this approach but I know I have heard Dr. Fleming talk about it and we have talked about it before, you could always have the toxicity affect your margin. That means you might be willing to accept less proof of efficacy if you knew it was less toxic. But that would be involved in the judgment process.

DR. PRZEPIORKA: Dr. Temple?

DR. TEMPLE: The grim reality of non-inferiority studies is that we usually set a margin at something like preserving half of what we think the effect of the drug is. That is not very gratifying. I mean, you would hate to lose half of the valuable effect and, yet, if you explore sample sizes it is really not possible to do much better than that. So, in return for getting a drug that might have less toxicity, or is easier to give, or is a different dosage form and things like that, we do the best we can sometimes, as Grant pointed out, there often isn't. So, it is a tremendous problem

to get less toxic or more easily taken drugs. The same problem actually arises when you are looking for drugs that mitigate the side effect of another drug. If you want to show that you preserve the effect of the drug, I can't imagine what size studies would make a convincing case and, as Grant said, there is often very unclear evidence on what the actual beneficial effect of the drug is in the first place. This isn't unique to oncology; it occurs everywhere but it is a major challenge.

DR. PRZEPIORKA: Mr. Katz?

MR. KATZ: Where in this do we account for differences in durability of response? For instance, you could have two treatments that have equivalent TTP but very different duration of response and that would be something that would be very different in terms of patient benefit.

DR. WILLIAMS: Well, I guess it would be a separate judgment. If they had the same TTP, that is one thing but duration of response would relate also to response rate. I have never had considerations where we were looking at TTP as a primary endpoint and we saw differences in response rate and we were making a judgment. But I think, obviously, if you are looking at response rate,

duration of response is always an important consideration and a big judgment call when you have such a long duration. I think the O'Shaunnesy paper had some discussions about that in the early '90s about certain settings with big response rates and long durations of response that we might consider using it as an endpoint for clinical benefit, but it is very much of a judgment call.

MR. KATZ: I guess I was raising it strictly because of, you know, the difference in quality of life between being treated with something constantly over a three-year period between your randomization and progression versus being treated with a blast at the front. That is a significant difference. You know, it is separate from the response rate.

DR. PRZEPIORKA: Dr. Grillo-Lopez?

DR. GRILLO-LOPEZ: I believe that TTP is an excellent endpoint for regular approval even and that, in fact, it is much better than survival. It may not be obvious but survival is plagued by a number of biases that we can discuss during the course of the day. One would tend to state that F is the ultimate endpoint when you are talking about survival but, again, there are a number of biases

when you are looking at death as an endpoint.

But to address your question, I think that one way to address the issue of TTP and its relationship to response is to do an analysis of TTP for responders. When you look at TTP for responders, this is even a better endpoint than duration because the problem with duration of response is that you are looking at two time points, both of which are variable. The duration of response starts from the first day that you see a response, and that can vary depending on when the evaluations are done, and ends with progression of disease which, again, can be somewhat variable. Whereas, TTP at least has a definite calendar date for the onset of TTP.

DR. WILLIAMS: WHO does response duration--or ERTC or somebody--from the time of randomization. That is where they routinely measure response duration but, obviously, there is a longer but perhaps more precise measure.

DR. PRZEPIORKA: Dr. Temple?

DR. TEMPLE: I was just going to comment on duration of response. There certainly have been situations where very long response was considered sort of self-evidently beneficial in some of the

leukemia/lymphoma drugs. In testicular cancer, if you are still alive and have not progressed at a year everybody assumes that you would have been dead. So, there are some of those cases but as an endpoint in clinical trials we have never been successful, to my best knowledge, in incorporating that particular measurement into the overall evaluation. We sort of say if it is too short, that might not be meaningful but I don't think it has been more precise than that except when you get these partial responses that last for a year and everybody is very impressed by that as a likely clinical benefit.

DR. WILLIAMS: That was a big role with IL2, wasn't it, Pat? Long duration response?

DR. KEEGAN: Yes, that was the basis for the approval both in metastatic renal cell and metastatic melanoma. Although there were relatively few responses--I think it was less than a 15 percent overall response rate for either one. The responses were measured in months for partial responders and years for complete responders.

DR. TEMPLE: And the treatments for hairy cell leukemia all sort of had those characteristics.

DR. PAZDUR: And that was for Fludara and for valcane too.

DR. PRZEPIORKA: Dr. Redman?

DR. REDMAN: Dr. Farrell, just for my own clarification because I heard the words being used in the same sentence, in the regulations clinical benefit is not defined as survival?

DR. FARRELL: Right.

DR. REDMAN: It is defined as clinical benefit. What we are trying to discuss is what is a clinical benefit and assuming that time to progression is a surrogate endpoint to survival may be false just by definition.

DR. WILLIAMS: But as I said in my talk, clinical benefit it not in the regs, or at least it is not in the Act. Do you want to say more about it, Dr. Temple?

DR. TEMPLE: It is definitely not in the Act. An important court of appeals case--whether that really changes the law or not is debatable, but Warner Lambert versus Heckler said it is just obvious that the Commissioner needs to consider what the effect is. He doesn't have to approve something silly, like there used to be drugs to increase bile flow. You know, that doesn't sound

like it is very useful. But that is what it is and it has never been defined as a particular thing. In other words, as Grant said, everybody thinks that delayed time to recurrence in adjuvant settings probably is a clinical benefit because, you know, you don't have tumor yet or you don't know you have tumor yet or because it is usually symptomatic. That is okay. If somebody thinks that very delayed time to progression must correlate--there is a lot of judgment in it. There is no rule; nothing is written down.

As Grant said, up until 1985 we used to approve everything based on response rate. We didn't think that was illegal but we concluded it wasn't so good.

DR. WILLIAMS: And looking back at the history of oncology, at the very time that we made this decision the Supreme Court was evaluating Laetrile and the Supreme Court was supporting the FDA that we could demand proof of efficacy in terminal cancer patients. The words used were symptoms, function and survival. So, I mean, it is a collection of sort of legal arguments as sort of the basis I think.

DR. PRZEPIORKA: Dr. Fleming?

DR. FLEMING: In considering the concept of clinical benefit, I think many of us have, across many disease areas, considered direct measures of clinical benefit to be measures that unequivocally reflect measures tangible benefit to patients. So, Grant had put forward examples of those. Obviously, duration of survival; measures that reflect quality of life; disease-related symptoms, those are obvious measures.

Where we struggle is that in any disease area there are targeted mechanisms by which we are hoping to achieve those clinical benefits, and we may be more or less right about those. In oncology we would tend to think those would be most directly measures that reflect disease tumor burden. Time to progression, response rate are, in that regard, measures that we would give considerable attention to. One could argue though that you could shrink a tumor by a certain fraction or delay time to progression by a certain fraction and that doesn't necessarily lead to something that the patient would be tangibly aware of unless, as was pointed out--I think Bob pointed out, if progression is associated with symptomatic disease or disease-free survival, if the delay in the time to having

detection of disease provides a psychological benefit. Those are direct tangible factors.

But the complication that arises here is that time to progression may, in fact, be the intended mechanism by which we hope to achieve clinical benefit but the problem is may you delay progression by two weeks or four weeks without that translating into something that the patient is tangibly aware of in terms of longer survival or improvement in symptoms or quality of life.

DR. REAMAN: For clarification, are we lumping together time to progression and time to recurrence and the issue of stable disease as an endpoint?

DR. WILLIAMS: I am specifically mentioning time to progression. We will talk about disease-free survival during the questions. We have taken a stronger stance, as Dr. Dagher has stated, that with disease-free survival in some settings is a clinical benefit. Disease-free survival in the adjuvant setting I don't think we would say is the same as time to progression. So, our discussion here so far has just been time to progression. If you would like to bring up the other now, but we will certainly discuss it later

too.

DR. PRZEPIORKA: Dr. Cheson?

DR. CHESON: I think what we are going to find here at the end of the day is that the importance of the various endpoints is going to vary considerably by disease. Dr. Temple was citing all these examples about how drugs got approved, single agents, all hematologic malignancies. What has been referred to this morning has been more referable to solid tumors. So, this is going to be really complicated.

I would like to get some input from people like Dr. Fleming, all too often we see that time to progression does not translate into a survival advantage. The cause of that is because the survival measurement is under-powered, or is it because once they progress with a longer time to regression they don't respond to subsequent therapy? What is the explanation for this because we see it all too often?

DR. FLEMING: That is a good question and it is one in general that arises as we consider markers as potential replacement endpoints. Just as a quick, brief response to your question, if we are using time to progression and we are using it

as a measure of the intended mechanism by which we hope to achieve clinical benefit, such as survival, why is it that you may see a time to progression effect and not a survival effect? Part of it may be that it is not fully captured in the entire mechanisms through which these processes are influencing outcome.

A better example I think of that might be if you used objective response rate as the surrogate because it may be that you are under-estimating the true effect on the clinical endpoints, such as survival, because the intervention has a cytostatic component that delays progression without necessarily shrinking tumors.

Of course, the other factor is the clinical endpoint can be influenced by unintended mechanisms so that you may be having a potentially partial beneficial effect mediated through the intended delay in time to progression, but that could be offset by other unintended mechanisms, toxicities etc. which would yield in the end a lesser impressive survival effect.

Typically the marker is more proximal and often the true clinical endpoint is more distal. So, it is not surprising that the nature and

magnitude of the effect on the more proximal measure may be different from the more distal.

The critical issue in validating a surrogate, as we will get to later on, is that it shouldn't be assessed in terms of statistical significance, yes/no. It should be assessed in terms of does a relative risk reduction in the time to progression translate into some definable and predictable relative risk reduction in survival. So, if we reduce progression by a rate of 30 percent, is that a pretty reliable estimate of a reduction in death rate by 20 percent? In fact, if that is true, clearly a study is going to be more adequately powered for progression than survival because you can detect a 30 percent reduction with half the sample size of a 20 percent reduction of death.

DR. PRZEPIORKA: Dr. George?

DR. GEORGE: I would like to talk a little more about the time to progression in symptoms issue. I think we all would tend to agree that conceptually, ignoring the methodologic difficulties, a delay in progression is a good thing. We have a lot of problems with measuring it, and how the design is done and all these things

that contribute to it. But it seems to me that if we are after a clinical benefit, an important clinical benefit is that development of symptoms. So, you have some diseases I suppose where you have the distribution of time to development of symptoms after progression that would be relatively short, in which case you would look to build that probably into the definition somehow. In other diseases you might have a very long time, and that becomes a lot more problematic I think because that would be more variable and longer-term in individuals and then you really have to worry about how it translates into individual patient benefit.

I noticed you briefly talked about some related things, like progression-free survival, and you just kind of briefly touched on them. So, do you have any more comments about this issue?

DR. WILLIAMS: Certainly, we look forward to your deliberations on this matter. Of course, right now this is just questions to the speaker. That is one of the biggest things we would like to know, can you do this or not? If you can't do it, then forget it. And, that is basically the answer we have got from most investigators, we can't do this. But if you can, we would love to see it.

DR. PRZEPIORKA: Just to clarify, I don't mean to put words into Dr. George's mouth but, again, it seemed that you were somewhat negative on the concept of progression-free survival as opposed to time to progression. Would you like to expound on that?

DR. WILLIAMS: Okay, what I should have said was that we have often said don't do progression-free. It has been our approach because we have been disturbed by loss to follow-ups coming in as deaths, you know, prolonging survival. It is a very sloppy business and there is no rule in there about how you deal with that. As a secondary endpoint I think that is quite reasonable but I think, as Dr. Fleming said, if you are really going to try to capture more in this endpoint if it is relevant, then include deaths. I think that is a good thing for you to discuss, is that reasonable to do? But if we do, then we have to do something to make sure those deaths don't mess up our analysis and produce unreasonable results like, you know, three-year progression-free survival and then death, things like that.

DR. PRZEPIORKA: Again, your definition of progression-free survival does not include death?

DR. WILLIAMS: TTP does not include deaths. Progression-free survival includes deaths. That is the terminology I use.

DR. PRZEPIORKA: Dr. Temple?

DR. TEMPLE: With TTP you censor the deaths and don't count them. With progression-free survival your worry is that you gain credit for very great delay in progression because nobody observed you for a long time until you died. It doesn't have an obvious bias, it just gives you a wrong number.

DR. WILLIAMS: Well, both of them produce wrong results. I mean, we like to censor the visit before the death instead of at the death but still, you know, that is being cut off because the patient died. Was really that death unrelated? If it was related, then you have non-informative censoring. So, it is which kind of bad data do you want. So, the real way to do it is to do the trial right and not have these kinds of things.

DR. TEMPLE: Can I pursue a previous discussion with anybody? The practical difficulties of doing time to death in addition to time to progression I don't think have been adequately recognized. Just as a quick example,

which will be statistically incorrect, if you delay progression from six to eight months my quick hazard ratio is 0.75. If you improve survival from 12 to 14 months, the same difference; you can't expect to have a bigger effect. So, your hazard ratio is only 0.86.

Now, the implications of that for sample size are major and I haven't even calculated a crossover. So, if you imagine that the crossover to study drug now reduces your advantage from two months to one month, we are talking about major differences in sample size. I am not sure anybody has actually modeled the difficulty but it is clearly going to be very, very hard just on practical grounds alone. You don't even have to postulate that there is a difference in effect on progression to survival. I am just assuming it is the same but still I am sure the sample size goes up a factor of four with what I just said, but someone can correct that. It is a very substantial problem, not really addressed.

DR. WILLIAMS: But underlying that, Bob, we have had many of these discussions and the issue is do you assume a constant hazard or do you assume a constant increment? I don't know what we should

expect.

DR. TEMPLE: Grant, why would anybody imagine that a two-month increase in time to progression would lead to a four-month increase in survival?

DR. WILLIAMS: I don't know but you heard Tom do it and I think the statisticians continually do kind of assume a constant hazard when they go from one endpoint to the other.

DR. PRZEPIORKA: However, this again begs the question of whether or not one is supposed to be a surrogate for the other, or can you say time to progression is a clinical benefit and we don't have to worry about whether it is a surrogate?

DR. TEMPLE: Right, but one of the tempting reasons to do that is the implication for sample size.

DR. WILLIAMS: Maybe we could hear Tom. What is the assumption and which is valid?

DR. FLEMING: Well, I think the essence of what Bob is saying is what drives interest in looking at replacement endpoints. The example I gave was a 30 percent reduction in progression rate compared to a 20 percent reduction in death rate and that would lead to a doubling in sample size.

DR. TEMPLE: It depends how much delayed death is compared to progression.

DR. FLEMING: Indeed. The example you gave, Bob--you are actually not too far off, it would be a three- to four-fold difference in numbers of events required to detect a 12- versus 14-month difference in survival rather than a six- versus eight-month difference in time to progression. It is what drives a lot of interest in looking at replacement endpoints. It is not just because they occur six months sooner that would cut six months off the regulatory process, but the relative risk that you would expect to see in the endpoint that is the direct mechanism by which you hope to achieve ultimate benefit, and it is more proximal, is typically going to be greater.

There are counter examples, Bob? How could it be that there is a counter example? Because your surrogate may be noisy and may not, in fact, be capturing the essence of the mechanism by which you achieve clinical benefit. So you may, in fact, have as impressive a result on the more distal clinical endpoint. But in general what you say is right, and that is that typically you are going to see a bigger relative risk reduction.

So, the challenge is can we achieve that payoff of a quicker assessment based on a smaller sample size, using Bob's logic, without paying the price of having less reliability? When is this quicker answer reliably telling us what we need to know longer term?

But while I have the mike let me just quickly go back to one of your earlier issues and defend what Grant had indicated I had advocated in the past, which is disease-free survival. Disease-free survival and time to progression are both important markers. Time to progression is censoring the deaths and if one is really trying to get at the mechanism by which I am achieving clinical benefit, a targeted mechanism such that what I really want to look at is the treatment effect on the targeted mechanism of tumor burden and I don't want that assessment to be clouded or complicated by the noise of unrelated deaths, I will censor the deaths and look at time to progression. That would make sense if it is a supportive measure of biologic activity. But if it is a registrational endpoint you want it to be as close as possible to what is really clinically relevant and clinically interpretable.

What is really relevant here would be to say I want to delay the time that I have progression or death. A good thing is to be alive and free of progression. So, those deaths should count. When you censor the deaths, and I think it is important for clinicians to know the game that statisticians are playing, if Grant and I are going along and I die and Grant doesn't and we are in the same arm, I am censored in time to progression but I am not left out. Some people think I am censored and I am taken out. No, I am still in the analysis and we are imputing my time to progression by what Grant's time to progression is.

Now, it is an incredible assumption of informative censoring that because I die I am no definition than Grant. I am probably more frail; I am different and so my time to progression would have been different from his. So, when we look at time to progression I would hope that we would also look at that with tremendous caution because we are censoring the deaths and we are making a major assumption about non-informative censoring that is almost certainly not true.

DR. PRZEPIORKA: Grant, I have a question for you. You talked about validated surrogates.

Who is responsible for validating surrogates, the FDA or the sponsors?

DR. WILLIAMS: Well, I really don't think that we use the term as a regulatory term. We are looking for something that is a substitute. In this case I was using validated to refer to the Prentice criteria for strict quantitative analyses. Certainly, our regulations don't have validated surrogate in them. I don't think we really have a regulatory answer for what a validated surrogate is, maybe Bob does.

DR. TEMPLE: No, we don't. But the accelerated approval rule says you know those other surrogates we used to use--blood pressure, blood sugar, the ones we are talking about now are less validated than that. That is really all it says. It gives you a direction and that is quite explicit in the preamble, but it doesn't say the other ones meet the Prentice criteria. I don't think anything has ever met the Prentice criteria because there is too much noise in the system to make a very persuasive case for that. But the contrast is with blood pressure, blood sugar and cholesterol which a lot of people would argue about anyway even though those are widely accepted. But it is a

qualitative, somewhat seat-of-the-pants judgment about whether this is persuasive or not.

DR. PAZDUR: Could I answer Donna's question?

DR. PRZEPIORKA: Sure.

DR. PAZDUR: I think the academic and scientific community have the obligation to validate these surrogates. We could accept or not accept the information that is provided to us but this tends to be a long and complicated process and what we are looking for is basically external validation that these are real, true scientific findings to them base regulatory decisions on.

DR. PRZEPIORKA: In that case I would like to follow-up and I am going to assume that there is no guidance document on what would accept as a validated surrogate. Is there a guidance document available for how to validate a surrogate?

DR. TEMPLE: No, there isn't and when you actually get into it, it becomes extremely difficult. For example, I bet if you looked at all studies over all time, shrinking tumors is probably good; I mean I think it is likely if you had a large enough database. What does that tell you about an individual study where the difference in

tumor response is a small percent? In putting a quantitative thing on these is extremely difficult. I mean, people could try to do that. It would be a massive project but I wonder how much it would help you in each individual case as to whether it was plausible or not. But your question leads to the answer that there really isn't much in the way of guidance on this.

DR. PAZDUR: But to follow-up on Bob's comment, I think this is one of the major problems we have had in oncology, that is, as we try to make some correlation here basically our treatment effects have been so small that it is hard to really impact the subsequent endpoint.

DR. PRZEPIORKA: Dr. Dagher, a question for you. You had gone through the list of all the ways of accelerated approval and obviously they need further follow-up for full approval. Can you tell us has there been any drug that has been approved on accelerated approval but had its post-marketing study turn out to be negative, and what did we learn from that and what did we do with it?

DR. DAGHER: Well, we discussed some of these at the March ODAC last year and I mentioned

that you could have confirmatory benefit either in the exact same population or I used the term related population. The reason I mention that is that it is intuitive that you would expect confirmatory studies to be done in the less refractory populations when you are looking for people for second- or third-line accelerated approval. But we have had settings where we have had evidence of clinical benefit confirmed in related populations.

What do I mean by that? We have some settings where we still had somewhat refractory populations but they were related. For example, the approval for Taxotere was for failure of prior athracycline. Then when we looked at confirmatory benefit, that was a population where there were some patients that had failed prior alkylator therapy. So, if you look at the label, after we did the conversion we now have a slightly expanded population, if you will, to say failure of prior chemotherapy which might have included either athracycline or alkylators. So, that is one situation where you could argue, okay, the population was still somewhat refractory but it is a slightly different population.

In the case of irinotecan, the evidence that was helpful in providing evidence to confirm clinical benefit came, as you know, from two European studies not the studies that were originally intended as the studies that were designated originally as those that would provide clinical benefit. In those studies, you could say those were fairly close populations in terms of the patient populations.

So, basically what we are saying is that you could have confirmation of benefit either in the same population or related populations. In terms of regulatory guidance, the 1996 document on reinventing the regulation of cancer drugs illustrated some concepts. One of the concepts was that clearly we recognize that confirmation of clinical benefit doesn't always necessarily have to occur in the exact same population that we use for accelerated approval. Obviously, the reason for that is that it could be more informative for us that further studies are done in different populations. For example, if you had accelerated approval in a third-line setting one could argue that it would be much more informative to have further studies done in the first-line setting and

evaluate benefit in that setting.

DR. PRZEPIORKA: I think my question was probably addressing more a specific individual study as opposed to a confirmatory trial where a drug received accelerated approval on the bases of a surrogate but in long-term follow-up survival was either not different or, in fact, worse with the new drug. Has that ever occurred?

DR. PAZDUR: Yes. Donna, a recent example of this is oxaliplatin. Although we approved the drug on the basis of an interim analysis of a randomized study which showed an improvement in time to progression and response rate, the survival did not show any advantage. Hence, you know, we knew that this was a high probability because there was a built-in crossover for all patients to receive the drug subsequently.

I think an important aspect is that when we take a look at accelerated approval--and this came out in the March talk--that we really have to take a look at the whole context of the drug development. It is not just one trial, this drug also had positive trials in a first-line study in an adjuvant setting. So, yes, there are examples. I think we have to take a picture of how the drug

fits into the context of other trials going on.

DR. PRZEPIORKA: Dr. Temple?

DR. TEMPLE: Well, the oxaliplatin is a very telling example and certain studies in breast cancer in my opinion came out roughly the same way despite a dramatic effect on disease-free survival. But that is because of the reason we gave before. There is crossover and it is later so it is much harder to win.

There are some examples, I mean there is a near miss, if you like. In the ordinary course of things Iressa probably would have been approved for third-line therapy with a requirement that they go study first-line therapy. Well, we know what happened there. They would have failed utterly. The message I think is, you know, you are not always as smart as you think you are. Drugs don't always work better--

DR. BUNN: [Not at microphone; inaudible]

DR. TEMPLE: I am just talking about the results of the well publicized first-line therapy study that was done, an excellent pair of studies. Nobody criticized the design. Yet, if those studies had been the requirement on an accelerated approval--other studies are now the requirement for

accelerated approval--you would have had a case where you didn't get confirmation but, of course, it was a different disease. So, it is possible. Can I say accelerated approval contemplates that. It contemplates the possibility that we will put a drug into the marketplace that ultimately proves not to be effective. The risk is considered worth it in bad diseases with no good treatment.

DR. PRZEPIORKA: Dr. Fleming, a final question?

DR. FLEMING: I was just following up on what I thought your question was, which is are there examples where an accelerated approval is granted and then a validation study is done and the results are not confirmatory. I think in the March 12 and 13 ODAC committee meeting we had we saw several examples. One of those examples was ethiol in advanced non-small cell lung cancer that was used for chemoprotection against renal toxicity, and where a validation study was done and duration of responses were much shorter with ethiol and survival was shorter, time to progression was shorter. Survival was almost statistically significantly shorter and was, in fact, shorter in the subgroup of ECOG performance status.

That was, in fact, an issue that came to light in that advisory committee, that not all validation studies are going to be positive and it is not as simple as saying, well, with crossovers at progression we are going to dilute survival differences. At times makers don't give a reliable assessment of what the ultimate clinical benefit will be. And, one of the complexities here is when those validation studies are quite unfavorable what happens?

DR. PRZEPIORKA: Dr. Dagher?

DR. DAGHER: Just to follow up, this is why Dr. Pazdur was emphasizing this concept of an overall development plan because we talk about confirming clinical benefit in the exact same population or in different populations, the fact is that you could have for a variety or reasons, as Dr. Fleming mentioned, studies that are "designated" as those that are going to be supportive for approval and, yet, those either aren't completed or when they are completed they don't show the results you expect.

This is why we encourage sponsors to sort of have a broad view of the development plan, meaning that we would like to have, you know,

several trials ongoing or in the process of being developed that could ultimately support that full approval. Like in the irinotecan example I provided, because there were other large randomized studies being conducted, even though they weren't designated as those that would be reviewed for confirmation of benefit because they were ongoing they could provide that evidence. So, when we talk about an overall development plan one of the things we are talking about is having other trials ongoing even if they are not necessarily "designated" at the time of the original accelerated approval as the ones we are going to necessarily review for confirmation of clinical benefit.

DR. PRZEPIORKA: Thank you. I think we are going to stop here for a break and we will come back for the open public hearing and Dr. Temple's comments starting at 9:45.

[Brief recess]

DR. PRZEPIORKA: Is there anyone in the public who wishes to make a comment? Now would be the time. Please come forward to the microphone in the front of the room. Seeing no takers, we will proceed to the discussion of the questions and Dr. Williams I think will give us some introductory

comments.

Introduction of the Questions

DR. WILLIAMS: I don't know if Dr. Pazdur is on the phone; I don't hear a cough. I imagine that is going to be the rest of our Division next week.

I just want to introduce you to the questions, sort of the structure. Why don't you turn to them? This morning there will be just sort of general discussion questions that we want to take general principles from to guide us as we go to specific areas. In the afternoon we will look into the questions on lung cancer and have a few voting questions if it seems that that will be helpful.

For this morning's session the first question is just on survival. It will be a continuation of what we have had here. The second question is about time to progression. We have had a lot of trouble trying to figure out how to do this. So, what happened is, you know, Dr. Pazdur took all of my little questions and was going to throw them away. Instead, I stuck them in the appendix.

[Laughter]

So, what we need to do is to talk about time to progression but also all of the different factors about time to progression, how important are the different factors? In the appendix I have sort of taken the different factors out to give you a little idea of what we are talking about, if you need to refer to that, things like relationships of time to death; whether patients are symptomatic; the magnitude and precision of the benefit; whether or not there is a benefit out there that has a survival effect for instance, whether that matters; how much does it matter if the endpoint is highly reliable or if it is more fuzzy; toxicity and the design, superiority versus non-inferiority.

I mean, you can come up with all kinds of scenarios but these are the factors that we are often considering when we say is this acceptable or not. So, there is a question here that mentions each of these factors and if you need to think more about them there is the appendix.

Then, there is the question of disease-free survival. We didn't really present on it but there is a little discussion here. Basically the issue is we have accepted disease-free survival in breast cancer, partly

because it is hormonal therapy and I think one of the early defenses was that these patients were more symptomatic at the progression so it is more like delaying symptoms. But others will argue that disease-free survival itself is clinical benefit, that you don't have known cancer and now you do and now you get toxic treatment. So, how you weigh in there I think will be important to us as we move forward.

Those are really the main two questions for this morning. Certainly, if you feel like there are other questions or points that you want to discuss, that is fine. So, I will turn it over to Dr. Przepiorka.

Questions for Discussion

DR. PRZEPIORKA: Thank you. Dr. Williams, just as a point of planning for this discussion and trying to make sure we get everything in, especially that last question which may actually have some importance regarding hematologic malignancies, and recognizing the complexities of the discussion for TTP, would you mind terribly if I took some of these out of order?

DR. WILLIAMS: You are welcome to.

DR. PRZEPIORKA: Thank you. Let's start

with the first question for the committee. Discuss the role of survival as an endpoint. Consider in your discussion the importance of whether existing therapies prolong survival and the potential confounding of survival results by patient crossover or where several subsequent therapies may also affect survival.

We actually discussed this a little bit about four years ago, if I recall. At that time I do recall Dr. Pazdur very pessimistically stating there is no drug that really improves survival in cancer so crossover shouldn't make any difference. But I think in the modern era that is no longer true, or am I incorrect about that? Dr. Grillo?

DR. GRILLO-LOPEZ: Perhaps even before we start discussion we need to make a distinction between survival as a goal and objective and survival as an endpoint. Survival is a goal for all of us here in this room because we are all involved either in patient care or in some way trying to better the lot of patients. You know, I have taken care of cancer patients and survival is very important to me. I am a cancer survivor myself. Survival is very important to me. But it is a word that is very compelling and that has a

lot of emotional baggage behind it. Perhaps because of that we are tempted many times to follow it with the phrase gold standard and perhaps we shouldn't.

Perhaps as you said earlier in our discussions today in considering TTP, and we will hear a lot about the pros and cons of TTP, we have to divorce that from survival as TTP being a surrogate for survival because survival is not a very good endpoint in fact. I love survival as a goal, as an objective. I dislike it intensely as an endpoint because it is subject to so many biases and a lot of people don't recognize that. The most important one may be that patients do get subsequent therapies and those subsequent therapies may or may not be active but there are extremes. There is the patient who chooses to have the best possible care, who takes care of himself, who follows treatment and who happens to respond to subsequent therapies. He will have a longer survival than at the other extreme, the patient who chooses to expedite his demise ultimately, perhaps even through suicide. If you have done enough clinical trials you will have had patients who committed suicide. It can be subtle at times. It

can be as subtle as stopping your medication and no one knows about it but you. But we think it is just jumping under the train; it is not like that. So, it is a very biased endpoint. It has more biases, in my mind, than TTP does.

DR. PRZEPIORKA: Dr. Brawley?

DR. BRAWLEY: I am sorry, are you talking about survival as measured in a randomized clinical trial or are you talking about survival as simply increased time from diagnosis to death as measured through comparing various trials?

DR. WILLIAMS: Randomized trial as a primary endpoint.

DR. TEMPLE: It is not that you couldn't be persuaded by a historically controlled trial but it just almost never happens.

DR. BRAWLEY: I have a second question which is more for Dr. Fleming and Dr. George. I sort of mentioned it to both of them. Are we assuming that increased survival in a randomized clinical trial translates in a decrease in either overall mortality or cause specific mortality?

DR. PRZEPIORKA: Dr. George?

DR. GEORGE: Since I heard my name mentioned--yes, we talked about this at the break.

Well, let's talk about lung cancer since we are going to talk about it this afternoon, I think it is traditional to use overall survival as the primary endpoint even though in many studies, if you look at attribution of cause of death, there are quite a few deaths that are not attributable to the treatment, not attributable to the disease but are from other competing causes of risk. So, I don't think we are assuming that. What we are doing though is we are saying that we don't really know; we can't really trust this attribution, first of all, in cause of death. Secondly, we wouldn't know quite how to interpret, say, a difference in cause specific mortality, say in lung cancer in this case, in the two treatments if there wasn't an overall survival difference because we don't know what the full mechanism of action of the treatments is.

So, I think it is not true that we are assuming anything about the different causes of mortality but what we are doing is saying that the overall survival is the important thing in those kinds of settings.

DR. PRZEPIORKA: Mr. Katz?

MR. KATZ: Well, I think we have to be

careful to talk both about the difficulty and the practicality of each of these measures separately from the validity of these measures as true measures of patient benefit because they are different issues. It seems apparent that we don't really have the capacity since we can't freeze time and we don't have computer models to basically run clinical trials in the blink of an eye, we can't answer the questions adequately.

I think Dr. Cheson said that the punch lines are likely to be different for different disease settings. I agree. But I think the other thing is that the punch line in terms of whether a certain endpoint is really an indicator of patient benefit is likely to be different for different patients because different patients may view overall survival benefit of eight months as something huge, whereas someone else, you know, may value disease-free, progression-free survival and maintaining a constant in terms of their current life styles as a higher benefit. So, I think we ought to view all of these, and is each of them valid to use as a measure and sort of add them as arrows and quivers as opposed to saying which is the best one to use because we have to use a lot of

them I think to get the right result.

DR. PRZEPIORKA: The question not here that I would like to throw out came up with our journal club back at home yesterday. We were reviewing a paper where difference in median survival ended up being 1.2 months but, because there were so many patients, the p value was 0.003. Dr. Williams I believe stated earlier that survival, when considered the endpoint, was easy to measure because when it is significantly different it is acceptable. But here our group looked at a paper and said we still wouldn't change therapy based on that. Any discussion on what is a meaningful increase in survival? Dr. Cheson?

DR. CHESON: Again getting back to what I said before, it is all relative. Whether you are talking lung cancer, whether you are talking follicular lymphoma or let's look at melanoma. We have some interesting drugs there. A difference of two months may be very meaningful. Yet, if you look at that in follicular lymphoma, as you know, we would go "pah."

DR. PRZEPIORKA: I think Dr. Williams asked earlier for discussion of principles and I think he is going to want some rather specific

examples. So, if you would like to discuss what you would consider meaningful survival in a lung cancer patient versus a low grade lymphoma patient he would probably be happy to hear those numbers.

[Laughter]

Just as examples of people who have long lives and short lives.

DR. CHESON: Well, I think also you have to look at whether you are talking front-line therapy or relapse therapy and, as he also mentioned, the risk of the therapy. For follicular lymphoma in the relapse setting I would think four to six months with a new therapy might be something important, whereas that would be only of marginal interest in up-front where some of the newer agents are, hopefully, getting us nine months to a year with additional therapy.

This is a totally moving target, particularly in the hematologic malignancies which, as you know, are far ahead of the solid tumors.

DR. PRZEPIORKA: Yes.

DR. CHESON: Every time we get a new drug approved, the bar just gets set higher and higher. So, what you say today is not going to be relevant in another six months for lung cancer, which I

don't follow. Paul and Bruce can certainly comment much better on what would be a meaningful endpoint. I know when I was still in my former job they were talking about response rates of interest in lung cancer being in the ten percent range. We saw that with Iressa and that would not cut it at all in hematologic malignancies, even in the most aggressive of those. So, it is a totally moving target.

DR. PRZEPIORKA: Any guiding principle you might come up with though? If drug A gives you two years benefit over no therapy and drug B is coming along, how much more benefit would you want to see?

DR. CHESON: It is hard to give an absolute number.

DR. WILLIAMS: Dr. Przepiorka, maybe I could focus that a bit?

DR. PRZEPIORKA: Sure.

DR. WILLIAMS: Because we have not, that I know of, not approved a drug that had a survival effect that we really believed. I mean, you also have to trade off the toxicity. But I think what we would really like to know is when you have a drug with a survival effect out there, how does that affect your acceptance of another endpoint

that isn't survival? A lot of times these survival effects are not so big--one or two months, as you mentioned, and that is what you have, maybe it is a symptom endpoint, maybe it is TTP or another endpoint with another drug. How does that, and what magnitude of effect of survival would affect the way you looked at this endpoint?

You know, we don't have a definite comparative efficacy standard but, nonetheless, I do think it is important we do consider these things, whether there is a large survival effect or not.

DR. TEMPLE: You have to be specific about the study. I mean, if you have a standard therapy out there that you knew something about and now along comes another drug and it actually shows improved survival, well, you know something about this drug. It is not worse than the other drug at least, and even if you are not bowled over by the effect it is sort of showing you that it does something other than shrink tumors. You might consider that as sort of proof of principle and a statement that, well, it is at least as good as what we have and actually it is probably better. Even if you think that one month is not of

particular value, it has told you something about the drug and what it can do. Whether that becomes standard therapy or not is a different question, but from our point of view maybe it has shown the kind of effectiveness you want if it is not over-toxic.

DR. PRZEPIORKA: Dr. Rodriguez?

DR. RODRIGUEZ: You are asking about developing principles and I think that coming up with specific numbers doesn't address a principle. I think a concept of principle would be, as Dr. Cheson has said, that there should be different guidelines for each malignancy. We are finding today that even within a defined category of malignancies we, in fact, have many biological variants of that same disease and we all have been bowled over at the recent meetings about how we now have to start thinking of proteomics and genomics in the definition of treatment for patients.

So, I think that this is, indeed, a moving concept and the principle should be that the endpoint should be appropriate for the disease and that it should be appropriate for the stage and/or status of the disease because patients who are in relapse are different from patients who are being

treated in the adjuvant setting, or for metastatic front-line treatment, and/or for post-transplant, or being considered for transplant, etc. I mean, I think we know as clinicians that we manage all of these patients very differently so we should not have "standard expectations" of any one of these categories of patients. They should be different.

DR. LEVINE: I would agree. I would add one more point to the principle. If, in fact, the survival benefit is a very small one, it would seem to me that I would want some confirmatory advantage as well as far as symptoms are concerned, or toxicity, or quality of life. So, one month in the hospital, you know, on IV morphine, or whatever, is not necessarily something that I would be aiming toward. I would want that in a small survival difference.

DR. TEMPLE: We don't really have authority to refuse a drug because its advantage over other therapy isn't big enough. We have said publicly that in oncology, unlike many situations where we would be obliged to approve something even if it was inferior, we would not feel obliged to approve an inferior cancer drug because there are serious consequences to that. But to insist that

it be better is really not within our statute. It doesn't have to be better.

It is important to make the distinction between showing that you are better as a way of showing that you work at all, which is what a superiority study does, and showing that you are better because you have to show you are better in order to be approved. You really don't have to show you are better to be approved. The statute and the legislative history is very clear that they were not trying to set a relative efficacy standard, much as one might want to know that a new drug was better. But we can't insist on that. What we do is we find superiority studies interpretable so that they show that the drug works. They also happen to show that it is better but that is in some sense incidental.

DR. PRZEPIORKA: Dr. Carpenter?

DR. CARPENTER: It seems to me that a couple of things may be helpful. One is that we have diseases, hematologic malignancies or breast cancer being examples, where there are a lot of therapies that are at least somewhat effective and that probably do impact survival. How one stacks up a new therapy at a given stage in that setting

and how one stacks up a new therapy in, say, disseminated melanoma where I think there is probably no generally accepted treatment that dependably improves survival are just going to be different scenarios and you almost have to have different rules there.

The other thing that has to be factored into this, but there is not a very quantifiable scientific way that such a committee always does, is to try to balance benefit and toxicity. Richard Gilber's analysis in breast cancer is one reasonably validated, not very scientific but it is an effort to quantify this kind of balance. I am not suggesting that we all adopt that but it is that kind of balance that I think is going to have to be left as a non-quantifiable but important aspect of this.

DR. PRZEPIORKA: Bruce?

DR. REDMAN: I think it is important--in reading the question, you are asking about comparing in a randomized trial against drugs that have proven survival benefit. I think that is a kicker because there are Phase III trials out there with a survival endpoint and the comparator is a drug that has never been proven to show survival.

It may be approved. Melanoma DTIC, and DTIC has never been shown to improve survival but it is used as a comparator. It may actually shorten survival; we don't know. So, if you are going to accept survival it has to be compared against a drug or a therapy that has been proven to affect the survival, or one that we think does.

DR. TEMPLE: Right. In a situation that you describe we would never accept non-inferiority as meaningful, obviously, but if it was superior, and ignoring your concern that the control might actually shorten survival--that is a big problem because you do have to assume it is at least neutral, in a study like that you would have to show an advantage over the available therapy and the available therapy would just be there as your placebo equivalent.

DR. REDMAN: Then the advantage of that has to be predetermined up front, what is acceptable. Then we are back to what Dr. Cheson was saying. You know, what is acceptable in stage IV untreated the same as the advantage in stage IV in someone who has received two prior treatments, specific in lung cancer, melanoma, kidney cancer.

DR. TEMPLE: I mean, historically we have

taken the position with the committee that if there is no available treatment that works for people we grant accelerated approval based on a showing of tumor response, time to progression, anyone of a number of non-clinical, borderline clinical endpoints. We would never worry if somebody managed to show improved survival and, as Grant said, even modestly improved survival. That has always been the basis for approval if you can show it. What you can show is really determined by the sample size you choose at the beginning as much as anything. I suppose if you made the study big enough you could show improved survival that a lot of people wouldn't think is very important. Historically, you would probably advise us to approve it anyway. That has been the pattern up till now.

DR. PRZEPIORKA: That is a very telling comment that you just made though since we are supposed to be approving drugs on the basis of clinical benefit, but I think I just heard you say, if I can paraphrase this correctly, that we always approve drugs on the basis of survival even if people don't think it is a very meaningful survival.

DR. TEMPLE: Yes, in practice studies are hardly ever large enough to show a completely trivial effect. So, we are in the 2-month, 2.5-month area and the recommendations we have gotten and our actions have usually said that is good enough in solid tumors; that is the best you can hope for so far.

DR. PRZEPIORKA: Dr. George?

DR. GEORGE: Could I address the second part--

DR. PRZEPIORKA: Yes, please, yes.

DR. GEORGE: --the confounding thing? This always puzzles me somewhat. If you have two therapies, let's say A and B, and then you have some other therapies that would be given after, say, recurrence or at some later point and often you don't have very good evidence that they have any effect, first of all. You might assume they do just to explain away the reason you didn't get any difference in survival. But whether or not they do, let's suppose that happens. You had a strategy of giving A and B followed by whatever is available at the time that they have recurrence, and let's suppose that that treatment does have some effect and sort of obliterates any potential survival

effect you would have gotten if you had done an unethical study, say, to force people to stay on treatment and not give them anything else no matter what happens--you couldn't do that, of course, ethically--so what is the overall conclusion you would come to? To me, it is that the treatment strategy you started off doing with A and B didn't work in terms of the outcome of overall survival in the context of that disease and in that setting with other potentially available therapies. So, in fact, if treatment A was the comparator and treatment B was the new treatment, in terms of overall survival you would say it doesn't have an effect. That is a simple answer.

Now, in terms of whether it is approvable, that means you had better have thought through other endpoints that you might be trying to use to get it approved. But in terms of overall survival it didn't work and it is not worth all the discussion about, well, maybe it was because we had all these other therapies or maybe it was this or that. The fact is it didn't work in this setting at this time.

DR. TEMPLE: The trouble is if the only endpoint that leads to approval was survival, then

this active drug has just failed.

DR. GEORGE: Exactly.

DR. TEMPLE: Even though if there weren't other therapies it would have been active in the usual sense. That is the problem.

DR. GEORGE: That just means you had better come up with the right endpoints and you had better not be using overall survival.

DR. TEMPLE: That is what we are here for.

DR. GEORGE: Well, I am just pointing out that people spend a lot of time discussing why it didn't work in terms of overall survival.

DR. TEMPLE: But that is because historically there has been a bias, not surprising and not unreasonable, in favor of a survival outcome because everybody knows that is tangible, that is a real benefit with some expressions of concern even about that. That is hard-wired. It is not subject to interpretation too much and everybody likes it. The trouble is the very things you are talking about can obliterate the ability of a drug that could be valuable to show its effect. That is what our trouble is, especially if the crossover is to the very drug that is being studied which happens for any marketed drug all the time.

DR. PRZEPIORKA: Dr. Cheson?

DR. CHESON: Harking back to something Dr. Rodriguez said, these diseases aren't failing these drugs. They are different diseases looking for the right therapy. We have certainly learned that in the hematologic malignancies where we started with, you know, leukemias and now we have separated them out into a myriad of different diseases. When we approve drugs, as we have seen recently, we are going to miss active drugs because the population in which they work is obscured by all the patients for whom the drug doesn't work, and there are some drugs that you all are approving that only work in small populations of a certain disease and, yet, they are getting generalized to the disease group at large and both of these are unfortunate circumstances for a variety of reasons. So, I think we need to recognize--and we certainly will be doing that more and more and we certainly do this in leukemias and lymphomas--that these are a bunch of very different diseases and we are going to have to be studying them like that. Instead of studying non-small cell lung cancer, we are going to have to find out, you know, what are the different subsets and how they respond differently

to drugs like Iressa etc., else we are just going to miss effective drugs and we are going to be spending a lot of money on ineffective therapies for patients in whom they don't work.

DR. PRZEPIORKA: Dr. Grillo?

DR. GRILLO-LOPEZ: I want to go back to what Dr. Rodriguez and Dr. Levine said earlier and add to what they said, that another consideration in choosing the appropriate endpoint and having an idea of what the expected magnitude of the effect should be is whether you are evaluating that new agent as monotherapy as opposed to that new agent within a combination therapy. If you are evaluating it as monotherapy and you are comparing it one-on-one, like the DTIC example that was provided by Dr. Redman, then I believe a survival endpoint becomes even less desirable because it is seldom that you see a single agent be curative in any malignancy. There are some exceptions but this is seldom.

The other extreme is when you are evaluating within a combination therapy. Now, we do have combination therapies that are curative in at least some percentage of patients with certain tumor types. However, how long did it take us as a

research community to find those optimal combinations? It takes years and years and years. Consider in your minds the ones that are available and you know how long it took to get there. It took many years after approval. Now, are you saying that you would deny the oncology community the opportunity to research this via an approved drug that can be worked into a combination, or that you would deny patients a drug that has shown efficacy in Phase II, that has reasonable activity, because you have not determined the optimal combination that would be curative and then you can use a survival endpoint? I would say no, you can't do that. Other endpoints are suitable to that outcome because it is very unlikely that during development, pre-approval, you are going to have the optimal combination identified.

DR. PRZEPIORKA: Dr. Williams?

DR. WILLIAMS: There is an underlying question that I don't think has really been heard. Let me just give you a situation. You have a marginal survival benefit out there. You are accepting TTP now; you believe in it as clinical benefit, let's say, but you are getting now this survival benefit over here so there are a couple of

different settings. One is something like fairly marginal, two-month median survival increase. You have a trial over here that is not even going to evaluate that; it is just going to use time to progression alone because of its clinical benefit too.

So, what is the tradeoff here? When do you have a survival effect here that is so significant that you can't do that trial; it is not ethical basically to use TTP to approve a drug? You wouldn't make the tradeoff for TTP because you have something else over here that is so good. One setting would be that you compare directly to this drug and you beat it in TTP. If you accept TTP, would that lead to approval?

Another would be that you evaluated TTP in another setting and you didn't beat it; you just showed that you had a TTP benefit. The question is when does the survival effect proven in one setting affect you so much that you can no longer accept this endpoint in another setting?

The way this happens is we have trials coming along. All of a sudden, one of these drugs is approved based on some survival benefit. It might be a little one; it might be a big one.

Then, at what point does that become so significant that it affects your ability to consider a different endpoint such as TTP?

So, that is the tension that I want to hear some discussion on. For instance, in the colon cancer setting, the lung cancer setting where you have one- or two-month survival benefit, does that then mean that you wouldn't even look at TTP as a separate benefit or that you would only look at it if you were beating that drug that had the little survival benefit? So, when I am talking about the size of survival benefit it is not necessarily would you approve it based on survival but how does that trade off and affect you looking at other endpoints?

DR. PRZEPIORKA: If I hear your question correctly, when would we actually insist on using survival as an endpoint and not use anything else?

DR. WILLIAMS: That is assuming that originally you had already accepted another kind of endpoint, such as TTP.

DR. TEMPLE: I assume this comes up because of the disconnected nature of the approvals. If there was something out there that had a survival benefit you would compare the new

drug with it because you couldn't really not.

DR. WILLIAMS: That is a question though. If you have a very small survival benefit you either have to say I am going to beat that drug, do a non-inferiority study which is impractical, or this is so small that it is not of any real meaning.

DR. TEMPLE: But it would be the standard and everybody would use it, but what you are saying is now you have just suddenly discovered something and you have all these people developing drugs without a comparison out there because they didn't know about it.

DR. PRZEPIORKA: Dr. Reaman?

DR. REAMAN: These trials are being designed and conducted to demonstrate a clinical benefit, not to dictate and define what the standard or a new standard is going to be. Correct?

DR. WILLIAMS: Yes, we don't do those kind of trials. We don't do trials to develop standards. So, yes, they are all being developed for clinical benefit but it is a different nature of clinical benefit here, the survival versus other drugs which might be TTP, let's say.

DR. REAMAN: But I think the question you raise really has to be considered within the context of the disease and the patient population in which the study is being conducted. I just don't think there are any absolutes that can be given, yes/no, will we always demand survival as the ultimate endpoint and can time to progression replace it.

DR. TEMPLE: Can I refine the question a little more? I guess if there were something that had a major effect in a particular setting, stage of disease--let's leave leukemias and cures, but had a major effect, most people would think the right way to develop a new drug is to compare it with that drug or add it to it or something like that. Right?

So, I think Grant is asking if you developed something that had an effect like that while other studies were going on that were looking at response rate, time to progression, would you be happy approving a drug not knowing how its survival effect compared to this thing that is now there? That is very important to people who are developing drugs without knowledge of what other people are doing at any given time. Does that capture your

question?

DR. PRZEPIORKA: Do you have a response?

DR. REAMAN: I would say yes. I mean, it may take a very long time to know about some of the impacts of drugs being approved and the impact that they could have on survival long-term, particularly using combinations.

DR. PRZEPIORKA: Dr. Grillo?

DR. GRILLO-LOPEZ: I have to say this is fun. I am practically jumping out of my seat here to address what Dr. Temple said. I did that. I developed Rituxan and we didn't find out until after the year 2000 that it was adding to the cure rate in intermediate grade lymphoma. We presented it to you for low grade lymphoma in a relapse or refractory setting, where survival was not an issue because it was not the appropriate endpoint, and you approved it. So, this is an example of an agent that had the potential of being curative within in a combination but got approved earlier on for relapse/refractory combination with a single-arm trial where survival was not the endpoint, and it was a regular approval.

DR. TEMPLE: Yes, we are well aware that the initial approvals of drugs do not define their

total use in the community. One of the reasons for accelerated approval was a barrage of arguments, often from the oncology community, that said, look, if you don't have the tools to do it, it is just impossible to develop drugs properly. Within limits at least, we bought that idea. That is why half of all drugs at least are now approved under accelerated approval based on response in refractory disease, the thesis being if refractory disease responds it is probably useful other places, and people are going to do studies, there will be cooperative studies and all that.

So, I think there isn't any particular debate about that question. There still is a lot of concern about what the standard should be given past guidance we have gotten for other kinds of approvals, not really most about accelerated approval which is sort of at least moderately settled if we know we could get the definitive studies done later. It is what should the standard be in first-line therapy given sample sizes, given crossover, and maybe that should be different from one tumor to the other. That is one of the things you are talking about.

DR. PRZEPIORKA: Dr. Redman?

DR. REDMAN: Regarding Dr. Williams' question, I guess a lot depends--you know, if you are talking about two randomized trials and if the comparator in the two trials is different, if the comparator arm is different and one shows a survival advantage while the other one was powered to show a time to progression advantage, I mean I guess you are never dissolve ODAC, you are going to have to ask somebody. I don't know the answer.

But if the comparator is the same and you said to them at the end of Phase II, listen, we will accept this as a valid endpoint as a clinical benefit, I think you have to.

DR. WILLIAMS: But it sounds like it is a value judgment and basically there is no over-arching rule that we are going to apply across the different diseases and it will be a case-by-case kind of discussion.

DR. PRZEPIORKA: I think we have beaten survival to death--

[Laughter]

Just to summarize, I think we started out with excellent philosophical points from Dr. Grillo, which is that survival is a goal but not necessarily an endpoint, and that survival can be

biased, as is pointed out in the questions, by subsequent therapy that is not standardized. However, under those circumstances we have to ignore the confounding factors if the original agreement was that we would look at survival; we should have different guidelines for each biological subset, meaning the disease, the status or any biological subset within a disease or disease status. At this point we can't demand survival under any specific certain circumstances. Everything has to be looked at individually.

Any other comments to add to that? Dr. Fleming?

DR. FLEMING: Well, it may be just a bit of a reinforcement but, to my way of thinking, choice of endpoints ought to be based on what it would be the patients really care about. In oncology, certainly, cancer has a huge effect on duration of survival and, certainly, from a patient's perspective to prolong survival would be of profound importance. That doesn't mean though that that is the only benefit that patients would look to. I would go back to Mr. Katz' comments, there may well be other measures but I would ask that we distinguish whether those other measures

unequivocally reflect tangible benefit to patients. Others that do, that we have heard a lot about, are disease-related symptoms or, as he was talking about, patient's functional status, being able to carry out normal activities.

Those would all be very tangible benefits. Those need to be put in contrast to the mechanisms by which we hope to achieve those benefits. In oncology classical measures would be tumor burden type measures such as response and time to progression. But I would only caution it may well be that we affect those measures which are the treatment mechanisms without, in fact, impacting the clinical endpoints of interest. I would argue then that our primary endpoints for registration should be these measures that unequivocally reflect tangible benefit or, as we will talk about a little bit later on, measures of biologic activity that have been validated.

I would like to reinforce one more thing that Dr. George pointed out, and that is the argument that has been given against survival is that it may be impacted by subsequent interventions. I would argue again from a patient's perspective that the goal here is to

formulate regimens which, when implemented in the best standard care approach in clinical practice, would prolong survival and improve quality of life. So, if I randomized to an experimental therapy against a control and secondarily supportive interventions allow for equal survival to be achieved, that is the truth. That is the truth. Even if the experimental therapy would give you an improvement in time to progression, if supportive care improves in the control arm such that there is no difference, that is the truth.

Now, it may be though that we have the wrong endpoint. In this case there may be clinical benefit in other measures. It may be that we are reducing the need for other toxic interventions, etc., in which case those factors need to be considered as well.

But the one thing that complicates this, and what Dr. Temple referred to before, is if best supportive care isn't what is being delivered to the control regimen but, rather, cross in to the experimental therapy so that you are looking at experimental now versus experimental later. That is answering the right question if you have established that experimental is efficacious and

you are just looking at what is the optimal timing for delivery.

But it is a circular issue if you are really trying to find out whether or not it is truly effective. I realize going down this path is going to be a very complicated pathway but I question the ethics and the scientific validity of crossing in to an experimental therapy that hasn't been established to be effective. Is it imperative to do so? No, it is not. An example would be the Evastin trials that have just been done in advanced colorectal cancer. Is it possible if you do that you will still be able to show benefit? The answer was yes, as was seen with Herceptin in advanced breast cancer.

But in general, as Dr. George had pointed out, crossing in to a best available standard of care is the scientific question of interest. That is not a bias. That is not diluting survival. That is the true effect on survival and if you are not going to impact survival in that way, then a different measure could be the relevant approach but it, again, should be a measure that unequivocally reflects tangible benefit.

DR. TEMPLE: I just want to make a

distinction between the best treatment of cancer patients and whether this drug is an effective drug because they are not the same thing. Tom, you are saying that if order doesn't matter, if you are studying drug A versus some treatment and now, when you progress everybody gets some other drug, if that drug turns out to be effectiveness enough, not necessarily more effective than the test drug but equally effective, say, it could obliterate or substantially reduce the apparent survival effect.

Now, that may be true information and useful information for the community of people treating cancer but it gives you the wrong answer on whether drug A works if survival is your endpoint. And, that is our worry. Also, if the drug is already available, if you are talking about a Phase IV study, you can rail about the undesirability and lack of ethics of crossing people over to the test drug but they are all going to be crossed over to the test drug anyway despite your view, which means that in many cases the confirmatory studies we want are perfectly predictably going to be much less powered than you wanted them to be in the first place. That is a consequence of insisting on survival.

So, I need to press this point because it comes up in conversations all the time and it is very important for us to distinguish between is this an effective drug and, therefore, should be marketed and what is the best way to treat people. It may be that, you know, using the other drug first is just as good, or the sequence matters, or any one of a bunch of conclusions. That is all fine. But what we want to figure out and we want to be able to tell people who come to us for advice how to figure out is what do you need to do to show that the drug works. And, I am very worried about survival where crossover is either predictable or unavoidable for the reason I gave before. I am sure somebody could model this. You probably need studies four times the current size, five times the current size. l

So, if survival is going to be the endpoint at least in certain settings, then everybody has to sit down and say, okay, we are not going to allow crossovers or we are going to try as hard as we can to prevent them, or we are going to do studies five times the size we are doing. You can't keep saying survival is the endpoint and not account for those things or then you get failure to

meet the desired endpoint and then you are scuffling for what you really meant in the first place.

I am hoping for real straighforwardness in this. If that is really, in practical terms, almost impossible to do, then we should hear that and not advise people to try to do it because they are not likely to be successful if the thing they cross over to is active, or somebody should model these things. It wouldn't be very hard. We could all do it. I couldn't but you could. We could model what the consequence of crossing over to an active drug is. You could calculate what the effect on power would be. But we really need to know the answer because otherwise we can't give anybody intelligent advice.

DR. PRZEPIORKA: Dr. George, last comment?

DR. GEORGE: Just to follow-up on that a little bit, you certainly could model it but it would be based on assumptions. And, one of the assumptions that seems to be behind this worry about the crossover is that when you cross over that agent that crossed over to, the same one, is going to have equal effect. In fact, that might entirely be wrong.

DR. TEMPLE: Fifty percent.

DR. GEORGE: Well, even if you assume some percentage, you just don't know. That is why you are worried about it I guess. But I think there are examples that show it is the timing of it that is critically important. So, later, at progression, it may not have the same effect or maybe a very small effect so you could still get a survival benefit. But I think your point is correct that you just have to think clearly about those endpoints, and if you think there is a possibility that that could occur survival may not be the best thing. You may get the right answer in terms of the strategy of using it but the wrong answer in terms of whether it is an effective agent.

DR. PRZEPIORKA: Let's move on to the questions regarding disease-free survival. The FDA has stated that disease-free survival can support regular drug approval in cancers where the majority of recurrences are symptomatic. Others propose that prolongation of disease-free survival should support regular approval in all clinical settings because a delay in cancer detection or a delay in the need for toxic cancer treatment is of clinical

benefit.

So, question number three is discuss whether disease-free survival is generally an adequate endpoint for approval of cancer drugs or whether additional evidence is needed, such as data demonstrating or suggesting that disease-free survival is a survival surrogate. So, I guess the question is, is disease-free survival an endpoint or is it only a surrogate. Dr. Brawley?

DR. BRAWLEY: I think they are two different things. I think disease-free survival without increase in survival could be a patient benefit. This is a purely hypothetical example where the patient's disease is suppressed for a prolonged period of time. The patient is without symptoms because of that suppression of disease. When that disease comes back and flares up perhaps even more aggressively, than if it had not been suppressed by the original drug--a purely hypothetical position, I think there is patient benefit there.

So, again, I am lapsing into what Dr. Cheson and Dr. Rodriguez have stressed before, that it is a disease specific entity and perhaps Dr. Redman is correct that we are going to prolong the

life of ODAC by making these arguments but I really do think you can use disease-free survival.

DR. PRZEPIORKA: Dr. Cheson?

DR. CHESON: I was just thinking but, no, I do agree with Dr. Brawley. I think disease-free survival is important, that the patient has no disease. The patient is generally seeing the doctor less commonly, has less complications, no treatment, less lab tests. So, even if there isn't a survival benefit there is generally a quality of life benefit and certainly the patients, as was mentioned before, would rather not have disease than to have disease around but it is just not progressing. But certainly from the quality of life aspect, visits and labs, and all that stuff, it is clearly a benefit. Now, whether that is important for regulatory approval of drugs is I guess something we are talking about.

DR. PRZEPIORKA: I would just like to add that I would also agree that disease-free survival is of actual importance, not a surrogate specifically in the leukemia patients. Patients we acute leukemia who relapse end up having to drop their job; put their lives on hold; get back to the hospital and be on therapy for another six months.

And, being able to delay that by one or two years makes a huge difference in their life, especially in young adults who are primary care givers in a family. So, I don't think disease-free survival as an actual endpoint should be limited to the adjuvant setting. There are some diseases now with very high response rates where disease-free survival could probably be a good endpoint. Dr. Taylor?

DR. TAYLOR: Well, I would agree that disease-free survival is a good endpoint but I think, again, you have to go back to it being very individual because some of the therapies we use to maintain a disease-free survival are very toxic, as with interferon with melanoma patients and it is something that you have to really weigh for each disease and each drug. I don't have any problem with disease-free survival but it may not be important if that entire time is spent doing high-dose chemotherapy and seeing the doctor anyway. Bruce already pointed out if you are going to have less doctor visits and less troublesome and better quality of life, that is an important aspect of it.

DR. PRZEPIORKA: Dr. George?

DR. GEORGE: I just wanted to be clear on this. This is a composite endpoint. It obviously can be closely related to survival just by definition almost. You know, if you die without a recurrence, I mean, that is an event in disease-free survival. So, it is going to be important to know in whatever setting we are talking about what is the likely percentage of patients that that might occur for. What are the sort of competing risks of death in the given disease setting you are talking about, and what is sort of known about the expected distribution about time from recurrence to death. Those are important considerations about whether this is going to be an important endpoint. I think in general it is a fairly good endpoint in a variety of settings because of those things but it just needs to be considered.

DR. PRZEPIORKA: Dr. Carpenter?

DR. CARPENTER: I think this is a critical area. I talked about balancing whatever these considerations are with symptoms. To be a little bit more specific, disease-free survival without major symptoms of disease or major symptoms of treatment is something that I think almost all of

us would say would be important. The bigger the impact of disease symptoms, the bigger the impact of symptoms from treatment, I think you would have to down-regulate that same benefit.

DR. TEMPLE: Wouldn't you presume that there are no symptoms from the disease if you are disease free? I mean, what would we be meaning if not that?

DR. CARPENTER: Well, let's give an example of allogenic bone marrow transplantation. You have no leukemia after your transplant but you have graft versus host disease which compromises your quality of life.

DR. TEMPLE: No, I understand about toxicity but not--

DR. CARPENTER: If it is disease-free, then you are free of disease and you have no symptoms from the disease. You are right.

DR. TEMPLE: Yes.

DR. CARPENTER: Absolutely.

DR. WILLIAMS: Dr. George, you mentioned duration from recurrence to death. I guess what you are saying is if there is a longer duration between recurrence and death it is a less important phenomenon. Perhaps for instance, you know, PSA

recurrence in prostate cancer might be many, many years. Is that what you meant?

DR. GEORGE: This needs to be considered. For example, if you have a very short time from recurrence to death you really are talking about sort of the same thing, especially if you have a lot of deaths that occur without recurrence. But, you know, you need to know that in a given setting because when you look at disease-free survival, for example in a setting where there is a long time between recurrence and death the curve is going to look real short and fast and then you have to kind of worry about that translation and relationship to survival. But that doesn't mean it is not a good thing. I think it is a very valid endpoint in many settings and is a good one.

DR. PRZEPIORKA: Mr. Katz?

MR. KATZ: Actually, what I wanted to cover was covered. I would just agree that definitely, you know, for a patient's standpoint it is a benefit to have increase in disease-free survival.

DR. PRZEPIORKA: Dr. Reaman?

DR. REAMAN: I would just argue that I don't think disease-free survival always connotes

the absence of symptoms for every disease. Certainly, individuals who have had surgical interventions for management of their initial disease may have long-lasting symptoms as a result of that. Patients with brain tumors may similarly have symptoms which aren't going to disappear. I also agree with Dr. Przepiorka that disease-free survival should be an endpoint and not necessarily be considered as a surrogate for survival.

DR. PRZEPIORKA: Dr. Levine?

DR. LEVINE: I was going to say the same about surrogate. This is not, to me, a surrogate; this is a valid endpoint. The only other point that I would like to mention is that if this is the only endpoint you will exclude some drugs perhaps unnecessarily. In other words, to get into that equation you have to be a responder in some sense and there may be other benefits of drugs that we are going to talk about later. But, to me, this is an extremely valid, real endpoint.

DR. TEMPLE: Well, you almost need either the adjuvant setting or something where there are a lot of complete responses or something not commonly seen in solid tumors certainly.

DR. PRZEPIORKA: Dr. Fleming?

DR. FLEMING: My own sense of whether I would consider a surrogate or not a surrogate would depend on the setting. We have heard a number of different potential benefits that could arise or could be accrued by having a delay in disease-free survival. One is if, in fact, this is a disease where at recurrence there is clear and frequent, if not standard, occurrence of symptoms, then clearly it is, in fact, a direct measure of clinical benefit.

One, of course, might argue that if that were the case then a direct symptom outcome measure ought to be able to also show that overall benefit. It has also been argued that there are potential psychological effects where, if we delay recurrence or detection of recurrent disease, there is that overall benefit to the patients. I would also accept that although that psychological benefit I would consider to be of much less profound importance than an actual delay in death.

As has been pointed out, what is the tradeoff in benefit to risk? If what we said is we are going to delay by six months or a year the knowledge of recurrent disease, how much toxicity would you accept for that benefit against saying I

am actually going to prevent the recurrence of disease; I am curing you of this cancer in 25 percent of the patients? I would consider that, as a patient, a far more profound piece of information, that I have a 25 percent increased chance of being cured than a delay in a year of the time in which I am going to have recurrence of disease.

So, it does become important to understand what it is that we can reliably conclude from a delay in disease-free survival. It is in part, in those cases where it is symptomatic disease, a direct clinical efficacy endpoint. In cases where it isn't it could also be a very relevant measure but now it is in the arena of a surrogate. We have to be able to know whether or not a delay in disease-free survival is reliably telling us we have a delay in death.

Maybe later in the discussion I will comment that there are specific standards that are emerging for what that evidence would have to be, but at this point I want to just distinguish that there are two different realms in which disease-free survival would be of interest. One is a direct clinical endpoint through the symptom

aspect and another is through its surrogacy for survival.

DR. PRZEPIORKA: Dr. Reaman?

DR. REAMAN: I guess I am still unclear about the symptom issue and why it would be a surrogate for survival. I am not aware of any disease that is easier to manage once it recurs. So, I don't understand why disease-free survival couldn't be an endpoint for determining clinical benefit. It is a clinical benefit if you prevent something from recurring.

DR. FLEMING: Yes, I think what I was saying is if, in fact, there was something tangible, such as symptom prevention or occurrence of symptoms or the psychological benefit, those are, in fact, direct clinical benefits. But that is separate from whether this is also predicting a prolongation of survival.

DR. REDMAN: But if it prevents the disease from coming back it could be predicting a prolongation of survival.

DR. FLEMING: Well, in fact, that is the hope and, yet, there needs to be some validation. Of all surrogates, this is one that tends to be much more plausibly valid, that if we can delay

recurrence of disease we are very likely to be prolonging survival.

DR. PRZEPIORKA: Mr. Katz?

MR. KATZ: I think given the fact that we are talking about diseases which can't be cured, I think we have to view this in terms of providing patients with options that they might not otherwise have that a rational person could perceive to be a benefit. Something like disease-free survival may be absolutely critical to someone based on where they are in their life. Someone may be in a position where being able to function without the disease for some number of years may be critical to putting their family in a financial position so they feel they have done the right thing. I mean, there is a lot of theory around this but I think it is all about patient options and that clearly provides patients with options that they don't have.

DR. PRZEPIORKA: Dr. Carpenter?

DR. CARPENTER: I was going to say something similar. Most of the situations we are dealing with here have to do with new agents for solid tumors and, in fact, curative medical treatment is generally unavailable for all these.

So, things based on a theoretical increase in cure are a little bit far out. Whereas, things that keep your disease from coming back for a tangible period of time or that keep your disease simply controlled for a tangible period of time seem to be a very direct benefit for that person.

DR. PRZEPIORKA: Dr. Brawley?

DR. BRAWLEY: No.

DR. PRZEPIORKA: There are two very interesting questions that are lumped into number four which come to the meat of what we do when things come here. Consider whether the adequacy of disease-free survival varies with the clinical setting in terms of an endpoint. B is treatment where the investigational drug shows prolongation of survival when randomized against an effective standard therapy where the standard therapy has already been shown to impart a survival benefit.

Would this august body be inclined to recommend approval based on disease-free survival for the investigational drug when compared against a drug that has already been shown to have a survival benefit? Dr. Carpenter?

DR. CARPENTER: Yes.

[Laughter]

DR. CHESON: This gets back to what Dr. Fleming was talking about before, that it is a bi-functional endpoint, the surrogate nature and the non-surrogate nature. Again, it is going to vary a bit with disease but I think in general--and I would think also when we were talking about time to progression before, it is not like you looked at survival and you didn't look at all the other endpoints along the way, like response rates and time to progression and disease-free survival. So, you will have some parameters to compare to this drug or this regimen that caused prolongation in survival and also had some point of disease-free survival and also had some time to progression and also had some response rate, looking at it backwards. So, you do have something to compare it against, which may give a little more support to using it as a surrogate endpoint in that particular condition.

DR. PRZEPIORKA: Dr. Carpenter?

DR. CARPENTER: And from a regulatory standpoint you just told us it doesn't have to be necessarily better to be approvable. It just has to be would we consider this evidence of effectiveness, and I think probably so.

DR. TEMPLE: Yes, I think B goes to, you know, you have one thing that shows that you know has an increase in actual survival. Now comes along something that is actually better on disease-free survival which you don't know the effect on actual total survival. How worried would you be not knowing that last?

DR. CARPENTER: Well, if you were to grant accelerated approval, I would think that would be the very right setting and you would hold that other in abeyance--

DR. TEMPLE: That is okay, other people would also want to know whether they could get regulatory approval on the basis of being superior to a drug that is already hot stuff in one measurement that isn't ultimate survival.

DR. PRZEPIORKA: Dr. Redman?

DR. REDMAN: I sort of agree with the statement that that would be fine but I would really like to see the data. What if the disease-free survival advantage was compared with the second-line regimen that prolonged the survival of the standard therapy that was given after those patients relapsed and they lived longer because they had the second therapy and now you have

brought it up front-line and there is no second-line?

DR. WILLIAMS: This is disease-free survival here.

DR. REDMAN: No, no, but something that has shown overall survival advantage. It may be that the overall survival advantage is then partly due to the regimen that you are now bringing up front.

DR. WILLIAMS: Well, I think what the question is meant to say is that you have a treatment that does improve disease-free survival. We know that; it is not secondary therapy. You have another treatment that comes along. It is either under-powered or the data aren't yet mature enough and it beats that treatment in disease-free survival but you don't yet know that it has the survival effect yet it is better in this surrogate or also maybe clinical benefit endpoint itself. Is that enough or are you going to be nervous about approving it until you see a lot more survival data?

DR. REDMAN: I guess I would have to know what the agents are, what the disease is. I mean, overall what you are saying is intuitively correct.

If it beats it in disease-free survival and, you know, the other one has gone out longer and shown an overall survival advantage, yes. But I couldn't in a blanket way say that.

DR. PRZEPIORKA: And I think a number of folks have already indicated that under the right circumstances disease-free survival is the endpoint. So, we would not be so worried about survival to demonstrate efficacy as opposed to let's look at the survival information when it is available for safety. Dr. Keegan?

DR. KEEGAN: Yes, I would like you to actually revisit the right circumstances because the right circumstances seem to be integrally involved with the toxicity of the agent. I think this is important if we need to meet with sponsors and tell them, well, it depends upon how toxic you are and your evaluation of the toxicity of this agent and the impact on the quality of life of the patient. Are you suggesting that for an agent which has more than minimal toxicity for adjuvant treatment or more than extremely short course that we need to be measuring some aspect of the quality of life and, if so, what aspects do you think are important? Because if, in fact, they lose on that

they have to have as a backup plan a trial powered to look at survival.

DR. PRZEPIORKA: Dr. Grillo?

DR. GRILLO-LOPEZ: Although there are exceptions, usually you are going to be evaluating an agent versus a combination therapy which may have some prolongation of survival and all of the issues of single agent versus combination come up again. It is unlikely that even though you are using the experimental agent within a combination that it is the optimal combination ever to be found with this agent. So, I would say in that situation disease-free survival is still a good endpoint.

If you are doing a single agent study, single agent versus single agent, standard single agent and experimental single agent, and you have a standard therapy that cures 100 percent of the patients and is totally free of adverse events, then disease-free survival is not the appropriate endpoint but I can't think of an example.

DR. KEEGAN: What about, for instance, areas where there is not a curative standard adjuvant therapy accepted so it would be single agent against observational control? I mean, obviously, it can't be less toxic than an

observational control so what components of toxicity should be evaluated? What are the important factors? One thought that was mentioned was that the individual is able to work and carry on all their activities of daily living. Is that the important component, you know, as opposed to just collection of adverse event information, which is hard to put into context of impact of a patient's physical functioning sometimes.

DR. PRZEPIORKA: Dr. Levine?

DR. LEVINE: A couple of thoughts. I differ a little from the group. In this example, B, my thought would be if we do have a curative regimen at some level, whatever it is depending on the disease, and now you have another drug which shows prolongation of the disease-free survival, in that setting I would say that is the surrogate marker. This, to me, is what accelerated approval should be all about. It is highly likely to convert into a survival benefit in the future. You don't want to withhold it from the people right now. In that example I would say it is a surrogate but I think it is still a good surrogate marker.

In answer to the question related to what would be important, I defer to Mr. Katz and others

but it seems to me that functionality is the critical issue. You know, if the patient is on this drug and the patient is able to work, or go to school, or care for family, that, to me, is critically important and far more objective--you know, the quality of life measures are very difficult to put meaning onto. Functionality is easier and more objective, it seems to me, and perhaps more valid.

DR. TEMPLE: This is the way you would measure how troublesome the toxicity is.

DR. LEVINE: Yes, can you function.

DR. TEMPLE: I have to say that we rarely get data of that kind.

DR. LEVINE: That is probably the most valid, I would think.

DR. CARPENTER: You should though.

DR. TEMPLE: Maybe. We do try. It is extremely hard to do in unblinded settings, which most of them are although not all adjuvant settings are unblinded. It is just very hard. I mean, in these quality of life things you usually don't know what to look for in advance. So, you are looking at multiple things and it is really hard. Many people have brought us patient-reported outcome

data and very few of them have been even close to persuasive.

I wanted to throw one thing out as part of the discussion. We are talking here about controlled trials where there is a control group. It is a fact though that for many years we have recognized the potential benefit of a very durable complete response, which is sort of related to disease-free survival, and we don't see that very often but where it does occur that has been a persuasive endpoint even on sort of historically controlled observations and I think that reflects the same thing you are saying here. All the treatments for testicular cancer that are approved were approved based on data like that.

DR. PRZEPIORKA: Mr. Katz?

MR. KATZ: Actually I have three points that have been stacking up here. One, relative to Dr. Keegan's question or comment, you know, I think that we have to distinguish between toxicities that are kind of quality of life issues and toxicities that are irreversible because, clearly, safety issues are a big deal.

You know, relative to Dr. Temple's comments, I think that that is one of the reasons

that we, patients, are really grateful to be at the table here because I think the size of the instruments that you guys come up with to measure quality of life is indicative of the fact of how hard it is to really explain. So, I think having real patient input on those things is really the only way to gauge that.

Also, I agree wholeheartedly with Dr. Levine. You know, when we are in the situation where we have low cure rates, low effectiveness of cure with these treatments I think we would all hope that people sitting around this table are basically asking themselves would a reasonable clinician gives to a patient and expect a better result even though we don't know for sure, and we don't want to hold back something that is potentially valuable. I think that is what I hear in this room and I am very encouraged by it.

DR. PRZEPIORKA: Dr. Carpenter?

DR. CARPENTER: I am just wondering about this issue that you asked about, functionality and how you measure impact. Functionality, even though hard to measure and maybe frequently we are unable to, I think most of us would accept is important. The other thing is some way to measure the impact

of the symptoms on the person's function. And, how many measures or how many other drugs in an adjuvant setting have to be used to take care of the toxicity or the side effects of the treatment, however you would want to quantitate that, it seems that one way to try to assess impact on quality of life and sometimes it is easier to count that or ask a few things. Pain medications are a long-standing thing but certainly not the only things used. Particularly in an adjuvant setting, you wouldn't expect to use many of them. But there are other things which may have to be used. Neuropathy would be a common thing that could have a big impact and is important in certain adjuvant settings--some way to try to measure that or sort that out because what you want is to control all the symptoms and not have the disease come back in this setting and some kind of way to quantitate how close you have come to do that. It seems to me a way to be able to compare and know what the impact of the new thing may be.

DR. PRZEPIORKA: Dr. Cheson?

DR. CHESON: Most of what I was going to ask has already been said. But it gets a little more complicated because some of these therapies

that prolong disease-free survival may be something you give immediately at the time you are initially treating the patient and some may be things you have to chronically administer and that has a different impact on patient quality of life, how you are going to follow toxicity, etc.

I certainly agree that we need in any circumstance to continue to monitor the AEs because there may be untoward events that are clearly unanticipated. Secondary malignancies are the ones that always come to my mind. It is nice that people are 100 percent functional but if five years down the line the risk of acute leukemia becomes eight or ten percent, then we have to reconsider what we are doing.

DR. PRZEPIORKA: Dr. Li?

DR. LI: I would like to hear Dr. Fleming's and Dr. George's comment on the single-point analysis discussed by Dr. Williams. The issue was raised for different assessment period imposed for the TTP or disease-free survival and that may cause bias and the need for a similar analysis at one-year survival or two-year survival as a single-point analysis for TTP or disease-free survival that may provide a kind of alternative.

So, I would like to hear some comment from the committee.

DR. GEORGE: It has a certain charm but is, like other things, I think a risky thing to do because you have to settle on what that point is. In terms of determining the progression you have to assess it at that time or enough to it, whatever that means, so it makes sense. If you miss it, that is worse than having a sequence of values of which you are missing one. So, it has some appeal in a setting where you know what that time would be and you are sure you are going to get all readings. Otherwise, I doubt that it would be of benefit. You are obviously losing some information and the question is whether that information is critical. I don't know. I would tend to say that is not the way to go. That is my feeling. You just need to develop procedures and carefully design studies so you kind of minimize the problems we talked about, that Grant talked about this morning, but not try to fix it with a single point.

DR. PRZEPIORKA: So, in summary, I think we are saying that--oh, Dr. Fleming?

DR. FLEMING: Had you already gotten to part C or are you still looking--

DR. PRZEPIORKA: No, C is open for discussion.

DR. FLEMING: Okay, if it is open discussion I might just add that C becomes much more problematic than B. I think we have discussed the complexities with B. In C, what we are saying is we haven't proven superiority; we have just ruled out that disease-free survival is meaningfully worse by some margin.

I think C is an extremely complex circumstance and I come back to this distinction again, is disease-free survival itself a clinical endpoint because it carries with it symptomatic improvement and it carries with it the psychological benefit? Or, is the major focus or a different focus of disease-free survival that it is, in fact, a surrogate at some level of validity for evidence for prolongation of survival?

In that first domain it is entirely possible to say that if, in fact, we are using this as a measure of symptom relief efficacy could follow if we establish that we are maintaining at least half of the symptom relief. On the other hand, if we are using it as a way of providing evidence that we are actually going to have a

survival improvement, which I still maintain, to my way of thinking, is a much more profound benefit if the intervention is actually providing a survival improvement. It is now very problematic as to whether or not not being a certain amount worse in disease-free survival allows me to conclude we maintained some of the survival benefit. So, I go back to some of the earlier comments and we will talk about this in more depth with time to progression later on this afternoon.

If we have established that an agent improves survival, let's say, and following Grant Williams' discussions from this morning we are saying we want to know that we are maintaining at least half the benefit we have to know not only that a benefit on the surrogate is telling us we have a benefit on the clinical endpoint, let's say survival. To do a non-inferiority argument we have to know how much improvement we can have or need to have in the surrogate to get a certain amount of improvement in survival. For example, it may be that, as with 5-FU, levamisole, 5-FU levorin in the adjuvant colon setting, we have a 40 percent reduction in the rate of disease-free survival and that translates into a 33 percent reduction in

death rate. If we want to maintain at least half that benefit in survival, how much reduction can we see in disease-free survival to maintain half? That is wishful thinking, to think we know the answer to that. So, essentially what we are doing is what I often refer to as my worst nightmare, a non-inferiority trial design in the context of using a surrogate endpoint.

[Laughter]

So if, in fact, here disease-free survival is of importance to us in a substantial manner because of its prediction of survival benefit, C becomes incredibly problematic. On the other hand, if all we care about in disease-free survival isn't because it tells us anything about survival but it is just that it tells us something about symptom relief, then it is possible to do this, although I would say it is pretty weak evidence that we know we are maintaining a small fraction of the symptom relief that standard of care would provide.

DR. PRZEPIORKA: So, in summary, I think what we are saying is that disease-free survival could be a primary endpoint rather than surrogate, most useful in diseases that have high response rates, testing drugs that have a very good

likelihood of giving a high response rate. It is important to keep people off therapy or on treatment with little more mostly reversible toxicities; that functionality is what is critical when looking at disease-free survival, and that we should also keep in mind the other endpoints that should be looked at just for confirmation of clinical benefit. In the situation for randomized trials where the comparator is already a highly effective therapy that has a curative fraction, there is some variation in thought regarding whether that disease-free survival should be an adequate endpoint or just a surrogate.

Let's move back to question number two--

DR. TEMPLE: Can I just comment on Tom's thing? I am sure it won't placate your nightmares--

[Laughter]

--but for the adjuvant setting, at least in breast cancer, we have asked for 75 percent retention of the effect on disease-free survival. Also, for what it is worth, even for tamoxifen I don't believe very many individual studies have actually shown improved survival. The meta-analysis does but that is not the same thing

if you are talking about an individual trial. So, that is not so easy.

DR. PRZEPIORKA: So, time to tumor progression, it has been proposed as an endpoint for regular approval, not a surrogate. Page two at the top lists the pros and cons that Dr. Williams has already gone through. What we need to do for the next 35 or 40 minutes or so is to discuss whether clinical settings exist where time to progression improvement should be considered an established surrogate for clinical benefit and should support regular drug approval. We need to identify the factors that determine when time to progression is an adequate endpoint for drug approval.

The factors that we are supposed to consider include reliability in measuring the endpoint, the relationship of disease progression to death, established benefit of available therapy, drug toxicity, and whether progressing patients are symptomatic. Dr. Williams has kindly provided us with a host of scenarios to stimulate our discussion.

If we could actually just pick up with Dr. Li's question from before about whether or not the

clinicians on this panel also have any comments about the single endpoint with regard to time to progression. Dr. Cheson is chomping at the bit.

DR. CHESON: I think using the single endpoint--again, I am thinking from my sphere of diseases, has the potential to be very dangerous. If you take some therapies where the initial toxicity, whether it be pharmacogenomic or for whatever reason, is exceptionally toxic and if you survive that you do well, then you are going to miss that initial real drop-off which might be a very undesirable effect. I drew a little curve here but, you know, the curve may go straight down and then sort of level off for the people who survive the therapy and you would miss that because of the same six-month point or whatever point you choose. Another therapy might get there but not have this initial somewhat disastrous effect on a large proportion of patients. So, I would be strongly opposed. I think you would lose too much very important information on patients proximal to that point in time.

DR. PRZEPIORKA: Yes, I would tend to agree in that the name of the endpoint is time to progression, not progression-free survival at some

point. So, if we really wanted to say that time to progression is what provides clinical benefit, we actually have to look over a course of time.

One issue raised earlier today is how do you measure this, knowing that patients come in for their staging at various time points and that can be somewhat difficult. My response to that was if the sponsor chooses to use time to progression as an endpoint, they need to do the work and they need to provide the data. If the data is missing, then they haven't done the study and they shouldn't get approval based on lack of data.

DR. TEMPLE: Could you talk about that a little more? One possible argument is that too infrequent measures decrease the precision of the measurement but, unless there is a bias tendency to get people in to look, it might not introduce a bias. So, how do you rate those two things? I mean, it might be true anyway even though you are only seeing them every three or four months. You might still be able to detect a difference as long as, say, the visits were similar in the two groups and there wasn't a bias. So, which is the worst problem or which problem are you focusing on?

DR. PRZEPIORKA: I think the problem that

I would focus on is missing patient data and missing the fact that if somebody doesn't show up for staging in a year you really can't make measurements based on every three-month interval. I mean, it is the difference between looking at a Kaplan-Meier and a life table analysis. In fact, some people put out Kaplan-Meier plots and you can tell how frequently they do their restaging because the Kaplan-Meier plots fall every three months. That is the kind of analysis that needs to be done as opposed to continuous analysis. The statisticians may end up having to come up with a new way to do comparisons using that sort of data because it is clearly not continuous.

DR. TEMPLE: So, they should make sure, if they are going to use this as an endpoint, that they are seeing people at some regular interval, every two months or every three months or whatever gives you the adequate precision.

DR. PRZEPIORKA: Hand-in-hand with that, you are looking at power calculations to determine how much of an interval in improvement you have to make, that interval has to be at least one interval between staging. You can't say you are going to stage people every three months and then you are

going to power to look for a one-month difference in time to progression. That would not make sense.

DR. WILLIAMS: I will follow-up on that because I have heard that and I honestly do not believe that is true. It depends on whether you are trying to precisely estimate the effect; maybe it is true then. But in terms of producing a highly statistically valid detection of effect, you can do it at one point just as well. So, the frequency really doesn't determine your ability to detect a small effect. It might determine your ability to precisely estimate the difference perhaps--maybe the statisticians can correct me on that point, but I have heard that discussed several times at ODAC and I don't believe it is true that you have to look at an interval that is smaller than the measured median difference that you are after.

DR. PRZEPIORKA: Dr. George?

DR. GEORGE: Just one thing about these kinds of measurements, of course, there is a whole big issue in statistics about how you handle this in data in longitudinal kinds of studies. This is a little different because here let's say you do a reading, then you have a long interval and you do a

reading again and there has not been progression, it is reasonable in this setting I think to assume that they never progressed. You are not monitoring a process that progressed and then un-progressed and you missed it. The problem comes in when you have those long intervals when you discover that they did progress and you don't know exactly when that occurred between this measurement here and here. So, you have to consider in this setting the disease I guess. We are back to that. What is the disease setting and what is your prior estimate of when these things would be occurring. So, you just don't want that to be too imprecise. You can quantitate that if you know something about the setting you are in.

DR. PRZEPIORKA: Dr. Redman?

DR. REDMAN: I agree with Dr. George. I got kind of thrown off by Dr. Przepiorka's one-year follow-up on a patient with advanced disease without progression. But I think, depending on the disease category, with the diseases I deal with you can define and I hate to say mandate but, you know, if you are going to say you are going to follow the patient every month by CT scans and every month you have to have the CT scans, and it has become less

of a problem in today's technology world. We just send them to a third party and they actually have copies.

I guess the question I have, and Dr. Fleming and I had a conversation, I am a little concerned about, you know, what happens in time to progression for the patients who die on therapy while they are responding. I got the sense from Dr. Fleming that those patients are censored and not evaluated and it has been diluted out. I am a little bit concerned about that because that somewhat speaks to the toxicity of therapy.

DR. TEMPLE: Certainly people look at toxic deaths as a separate item. How that gets factored into the analysis is something of a question.

I wanted to be sure about this, could I ask Tom and Steve, should we be advising people who are hoping to detect an advantage of, say, two months that if they don't see patients every two months they don't have a prayer; it is not valid? Or, could you, in fact, see them every three months and still detect a difference of a couple of months? That is the question Grant was raising. Is there a precise relationship or requirement?

This is very important for how we advise people. If they are looking for differences that are small, two or three months, they had better make sure they are seeing people at least as often as that or perhaps more often.

DR. FLEMING: It depends on the nature of the true distributions of time to progression. If we just said, for example, if we had exponential distributions for time to progression, i.e., time to let's say a certain amount of growth in tumor volume and there was a two-month difference in the median, you could look less frequently than two months and you could still see the difference. But, you know, sensitivity to that overall difference is going to be somewhat less. So, it is not a black and white, yes, you do; no, you don't but your sensitivity will be somewhat diminished if you are not following them with as great a frequency.

In fact, you said before how could you have a bigger survival effect than time to progression effect, this is one of the ways. This is one of the contributing ways. You are actually getting a noisy measure of what truly is happening by the intervention to tumor burden.

DR. PRZEPIORKA: Dr. George?

DR. GEORGE: I support Tom's opinion on that but I would also say that you need to consider the circumstance you are in. That is, there is no hard and fast rule that says if you are trying to pick up a certain difference you have to do the measurements like this. But you should be considering what you know about the rate, or what you suspect would be the rate of progression over time. I guess that is what Bruce was saying too. In other words, you would do it differently in different settings. So, I think you want to have reasonably careful measurements in that period where there is a high risk.

DR. PRZEPIORKA: Dr. Grillo-Lopez?

DR. GRILLO-LOPEZ: No.

DR. PRZEPIORKA: Mr. Katz?

MR. KATZ: I would suggest adding one factor to the list that we put here. We said whether progressing patients are symptomatic. I think whether stable patients are symptomatic is also germane here because you have tumor reduction but no symptom relief.

DR. PRZEPIORKA: Could you speak a little bit more about that with regards to who might

actually be a good candidate for a time to progression patient? If somebody is symptomatic already, is time to progression really an endpoint that you would consider clinically valid? That is, you are sick and as long as you don't get any sicker it is okay or, is this something for patients who have minimum disease and are not exactly ill?

MR. KATZ: Well, clearly if you start in a situation where you are highly symptomatic everything is valid. If you can get a treatment and it relieves the symptoms and it delays the time to those symptoms getting worse, then there is certainly an argument to say that that has a value to a patient. If a patient has profoundly serious symptoms that are horrible but you know that they can get worse but they are not getting worse because we have done this and it hasn't progressed, then I think that is also valuable. You know, things get more acceptable depending on what you are looking at coming next.

DR. PRZEPIORKA: Dr. Temple?

DR. TEMPLE: As Grant said, we have been encouraging people for years to look at time to symptomatic progression and I would say we have met

with total failure. Nobody does that for a lot of reasons. I don't know why. You probably know better than I do why. Symptomatic improvement in a group that is symptomatic has always been accepted as a valid endpoint. But as Grant also said, except for a couple of pain things with prostate, we have had very little success in attempts to do that and you have seen them--esophageal obstruction, you know, that works fine but most of the other things have been very resistant to success.

DR. PRZEPIORKA: Dr. Levine?

DR. LEVINE: I was just going to say that in considering time to tumor progression as the endpoint, not as a surrogate but as a real endpoint, it would seem to me that I would want it in the context of some sort of confirmatory clinical benefit other than that itself, i.e., symptoms are manageable; symptoms are better or have not re-occurred; toxicity of the drug is "acceptable"; quality of life. So, if it is just time to tumor progression alone without these other things, I don't know that that would be valid in a clinical sense.

DR. CHESON: Again, that depends on the

clinical sense because there are some settings where you start with nothing. When you "ain't" got nothing you have nothing to lose. If they start in an adjuvant setting or some setting where the patients just have disease, are asymptomatic, like early stage follicular lymphoma, and they don't have anything, then it doesn't work there.

DR. LEVINE: Right, you are right. So, in other words, it goes back again to disease specific situations.

DR. CHESON: Right.

DR. BRAWLEY: Can I ask for a point of information?

DR. PRZEPIORKA: Yes.

DR. BRAWLEY: Was gemcitabine approved for quality of life or for prolongation of disease-free survival?

DR. TEMPLE: Two reasons. Lilly invented a clinical benefit scale that had some elements of tumor progression and some elements of other stuff and they won on that. That is one thing.

But I think what actually persuaded people most was the one-year survival of 18 percent versus 2--not an official endpoint but it sort of looked pretty impressive. So, that is what it is for

better or worse.

DR. PRZEPIORKA: Dr. George?

DR. GEORGE: Can we talk a little more about the issue of the deaths that occur when you are looking at time to progression and death occurs before progression? I think if you are in a setting where there is some substantial percentage of patients for which that is true, that greatly decreases the value of the time to progression kind of analysis, in my view, because you don't know what that means. Further, even if you don't have deaths first it is pretty important to know something about that distribution from progression to death in different diseases, again to get back to the point I made earlier. If it is very short then, of course, it is sort of the same as survival really but if it is long, then you are in a setting where you probably need to consider this more as a surrogate or a potential surrogate. But I am worried about a situation in which you have some substantial proportion of deaths without progression and how you handle those then becomes critical. In the usual way you just kind of censor them but that is clearly subject to a lot of problems.

DR. PRZEPIORKA: Dr. Williams also talked about time to treatment failure as being an unacceptable endpoint and, yet, if we talk about time to treatment failure defined as disease progression or death would that satisfy your concern about how to incorporate death?

DR. GEORGE: Yes, but that is more like progression-free survival. I like that.

DR. WILLIAMS: I think we need to bring that up and the question is when we are looking at TTP as more like clinical benefit endpoint or surrogate, should we use progression-free survival, include the deaths and do a very careful evaluation and analysis to deal with the deaths or should we use TTP? It sounds like there is at least some consensus that progression-free survival is a good endpoint.

DR. PRZEPIORKA: Any disease categories where anyone here thinks that progression-free survival or time to progression simply would not fit and should never be used, or the converse where this is clearly the best endpoint because they will never get a remission and all you could hope for is progression-free survival?

DR. TEMPLE: Well, just to be clear, I

heard some uncertainty about that from Dr. Levine. I mean, if along with that you need to improve symptoms or something like that, then it is not just progression-free survival; it is symptomatic benefit too. So, I think we need to be clear on what people do think. But our initial question is, assuming you don't have all those clinical benefits, do you think progression-free survival or time to progression is a good stand-alone endpoint in this current, real world? If that is not clear, we are very interested in hearing whether it is or not.

DR. PRZEPIORKA: Dr. Redman?

DR. REDMAN: I think progression-free survival, at least in the tumor types I deal with, is fine. I don't think this is the implication, but if you have a drug that is coming in and you say, okay, we are going to pick progression-free survival and it cures 100 percent you are not going to miss it. I mean, it is going to be there. You are just saying what is the lowest, minimum activity or clinical benefit we are willing to accept.

DR. PRZEPIORKA: Dr. Fleming?

DR. FLEMING: Just to return to kind of a

general response to this question of where and when can TTP or progression-free survival be used for regular drug approval, I would return to the pros and the cons and, just in the interest of shortness of time looking at the cons, what we have to overcome are these uncertainties, uncertainties that arise because it is an indirect measure. The clinical meaning of TTP differences, of small differences is unclear. The reliability of unblinding interpretation results are issues. I would add to that another one that, in fact, did come up in the oral presentation, and that is just the noise and the variability factors add complications due to variability in imaging assessments or timing of assessments, as we were talking about some ten minutes ago, and missing data. There tends to be a bigger missing data problem with the TTP endpoint, less so with progression-free survival and, obviously, even less so with survival.

Because of this issue of clinical relevance and missingness induced by death, I find TTP especially problematic if I am using it as a registrational endpoint as opposed to a supportive measure of biologic activity. So, among the two,

if we were looking at it as a registrational endpoint, certainly I would prefer progression-free survival.

But I would like to just step back for a minute. Rather than say, yes, it is a good endpoint; no, it isn't a good endpoint, just talk a little bit about the principles that should guide the decision as to when it is a good endpoint and what kind of evidence we would like to have because there is now a lot of science behind what it takes to validate a surrogate.

So, in our November 12 meeting of the FDA ASCO working group, basically in that session we talked about a marker such as time to progression as being one of four levels. Level one would be the best. In level one forget about surrogacy, it is, itself, a clinical endpoint. We said examples of that would be when you have the event disease-free survival or progression-free survival it is inherently linked to symptomatic disease. So, symptomatic events, preventing or delaying symptomatic events are inherently of tangible benefit to patients. If that is the case, then we have an endpoint that is, in fact, in its own right a valid clinical endpoint and surrogacy issues

don't arise.

The second level would be an endpoint that reliably predicts clinical benefit. So, when I see an effect on time to progression I can know that I will see--let's say if it is a surrogate for survival--a certain level of effect on survival.

The third level is reasonably likely to predict clinical benefit where the agency then uses this as a measure for accelerated approval but with the understanding that the ultimate answer on clinical endpoints will still have to be obtained in a validation trial.

The fourth level I will call none of the above, none of the above often being a correlate. There are an awful lot of correlates out there that, in fact, aren't any of the top three levels.

What does it take to be in level two, versus three, versus four? Well, the first thing we will look for is if it is a correlate. Is time to progression a correlate of survival or whatever the clinical endpoint is on a patient specific basis? Almost certainly it is but, in essence, that doesn't tell us anything about whether specifically the benefit or the outcome on the clinical endpoint is mediated through that. For

example, you may have CEA correlated with survival but it is not through changing CEA if the disease process leads to an outcome in survival. So, changing CEA may not change survival. That could be a level four.

So, we have to go beyond that. The evidence that we typically look at to go beyond that is guided by the Prentice criteria. So, what we are typically looking for is not just having a correlate. That is a necessary condition. It is not a sufficient condition for validity of a surrogate. We want to find out whether or not the effects on that marker are, in essence, capturing the net effect on the intervention of the clinical endpoint. At a certain level of persuasiveness that would get us to level three and I think in many settings people would argue time to progression because it is, in fact, directly--getting at tumor burden is very likely to be at that level but obviously it needs to be addressed on a case-by-case basis.

The bigger challenge is to say when is it a valid surrogate such that I know if I achieve an effect on this measure I don't need accelerated approval; I have actually established clinical

benefit. That best evidence is obtained by meta-analyses of studies that have looked at an array of trials, an array of studies that establish treatment effect on the surrogate--in this case I will call it time to progression and treatment effect on the clinical endpoint I will call survival--specifically saying what is the functional relationship between a certain level of reduction in the failure rate on time to progression versus a level of reduction in the failure rate on survival.

Understanding that is really critical and, in fact, in many settings we don't have that kind of evidence and, as has been pointed out before, partly because we are looking at interventions that at this point don't establish much of an effect on the clinical endpoint. But the essence of validating a surrogate and saying we can use time to progression as a surrogate for, for example, survival would be having meta-analyses of studies that would show reduction in time to progression rates and reliably would tell us we would have reductions in whatever the clinical endpoint is, such as death rate--reduction in the rates. So, if we reduce the rate of time to progression we are

improving time to progression and we want to reduce the rate of death to improve the survival time.

DR. PRZEPIORKA: Dr. Williams?

DR. WILLIAMS: Dr. Fleming, I saw your categories at the workshop on colon cancer but when I was preparing my talk I was wondering what category we would put our practice of breast cancer hormones and response rates. I mean, perhaps category four, which is even worse than accelerated approval category or what I think it is, it is clinical inference about number one. I don't know if you have a category for that and I don't think you do.

DR. FLEMING: Well, my sense is that if you are talking about response rate in breast cancer--I think that is the example you were giving--

DR. WILLIAMS: Well, it was hormonal breast cancer where there is a long history with gemoxifen--

DR. FLEMING: Right.

DR. WILLIAMS: --and assume benefit but a long history of using tamoxifen and it was felt certainly by experts in the field that it was useful and this was used as a surrogate and maybe

the blood pressure and maybe some of these others. I don't see a category here that I could put them in. They are basically clinical judgment, clinical inferences about the benefit. So, what do you do with those?

DR. FLEMING: Certainly, my sense has been--and you can clarify what your sense is, but my sense has been for some of these interventions that provide a duality here, that are providing some direct evidence of benefit through, for example, delay in symptoms and a surrogacy aspect of them, saying that if you are in fact delaying progression that is some suggestion of a prolongation in survival. The duality of that in the context of a very safe intervention is giving you adequately persuasive evidence of benefit to risk. In the end that is what it comes down to. In the end is benefit to risk established to be favorable? The stronger the evidence of efficacy, then the more resilient you are on safety and, similarly, if you have an incredibly safe intervention you might accept or you might be more resilient in what you consider adequately strong efficacy. Certainly showing a survival benefit I would say in many ways is the most compelling thing

to do because it is the most compelling benefit and provides more resilience to issues of irregularities in trials and issues in safety that could arise.

In this case, what I understand you to be doing is really, in essence, saying we have partially a level one here because we have some very direct tangible benefits that are occurring and it is reinforced by an anticipation at some level, valid or invalid, that you are actually delaying death as well. With a very safe intervention that is favorable benefit to risk.

DR. WILLIAMS: I think that is really basically a lot of what we are doing here today with progression-free survival. Are there settings where we can accept, or the clinical experience with this endpoint, the broad experience it seems clear we don't have the strong quantitative validation we would like but, you know, what are those factors which might allow it to be used in some very specific settings at this time?

DR. FLEMING: Just one last response to this, you identified some of those in your appendix. So, specifically the ideal settings are C, E, P, J and N, C being itself patients are

symptomatic so you have at least in part a level one endpoint. By delaying time to progression you are directly getting evidence of an improvement in symptoms or delay in symptoms.

I might challenge whether there would have been another way to do that, specifically looking at a symptom endpoint as a way to establish that. I also might challenge that that is, in my own view, not as compelling as actually having evidence of a survival effect. But C does get, in my definition, potentially into level one. So, surrogacy issues are not as compelling.

If we don't have C, and many times we don't have specifically symptomatic disease at progression. In November 12 meeting that was certainly the agreement, that in first-line colorectal cancer at the time of progression we don't typically see symptoms. Then, these other aspects that come into play are do we have a large and precisely defined benefit? The larger the benefit on the measure, obviously the more plausible it is going to be that it actually translates into clinical benefit. Hence, P, a superiority trial, is far more persuasive a setting. A non-inferiority trial and surrogate, as

I have already said, is my worse nightmare.

Blinded trials are important and we probably can achieve that routinely so it does, in fact, diminish our confidence. We can in fact though, as you say in K, try to have some kind of an independent evaluation committee that is itself blinded.

N, drugs that have minimal toxicity, that is where I see in part the example you have given comes into play. The evidence on efficacy is somewhat less but if you have an intervention with an established record that is extremely safe you may, in fact, have a little more resilience on what the strength of evidence on efficacy would be.

DR. PRZEPIORKA: Dr. Grillo?

DR. GRILLO-LOPEZ: Having heard all of that with a bit of impatience--

[Laughter]

--I have to say that clinical medicine even today is still an art and clinical research resists our efforts to quantitate it; it is also an art. And, there is no such thing as a perfect endpoint. There is no such thing as a perfect endpoint and TTP has its problems but it has a lot of pros. You have to also make a distinction

between those problems that are inherent to TTP and those problems that have to do with how TTP is measured, presented, how the data is acquired in the clinic, issues like GCP, sloppy data or good quality data, and put those aside because your assumption has to be that the data is going to be of good quality. That should not be a deciding factor on whether or not TTP is a good endpoint. You have to assume it is going to be good quality.

DR. PRZEPIORKA: Dr. Cheson?

DR. CHESON: Just one more small comment. Listed under your pros there is a theme. TTP is a measure of tumor effect in all patients, rather than measure effect in a subset of patients. I would look at that as a con rather than a pro. We are talking about all the different subsets of patients that may respond totally differently and you have to have a very strong impact on the right group to overcome--going back to Iressa for example--to overcome the negative impact on another personal bias but that is how I would look at that.

DR. PRZEPIORKA: Dr. Temple?

DR. TEMPLE: I have a comment along the same lines. One of the difficulties, and you have described this repeatedly, is that we are trying to

look for an effect in an overall population when we are only probably influencing a small fraction. That is a real burden. In most other conditions you don't have to do that and you have some hope of treating everybody's headache even if that is not true. So, that is all going to get better when we get all pharmacogenomics--

[Laughter]

--I think Grant listed that as a pro for the following reason and I wonder what people think about it, that to actually shrink tumor volume by 50 percent you really have to be quite a good responder. There may be people who don't get quite that good a response but whose tumor growth is slowed, and you might think there are more of those than the former. That is why I think he thought that might be a more powerful measure.

But I also have a question. Remember, I don't treat patients with cancer so if you think this is really stupid just tell me. If there is no really good follow-on therapy, which is often the case, why do we monitor progression other than by symptomatic progression at all if there is nothing much we can do about it? If everybody progressed with symptoms then there wouldn't be any argument

about it. So, why do we do that? If that is really a stupid question, just tell me.

DR. PRZEPIORKA: Dr. Taylor?

DR. TAYLOR: No, it is not a stupid question. For many of us who have patients in whom there won't be treatment we don't do repeated x-rays and you do go by symptoms and you treat them by symptoms because that is the most practical thing to do. In essence, that is why ASCO recommendations are for follow-up after adjuvant breast cancer, to follow symptoms and to do mammograms and physical exams. So, that is not a stupid question.

The only time we are compelled I think to look for progression is when we are in an investigative setting in which we want to know what is going on with this particular drug.

DR. TEMPLE: For what it is worth though, we wouldn't mind seeing a study that was simplified and that only weighted for symptomatic progression. Whether it is ethical to do that is a different question. But if it was time to symptomatic progression there would be no debate about whether that was clinically meaningful at all.

DR. TAYLOR: Again, I would say that is

only specific diseases. There are some diseases where you do need to monitor.

DR. BRAWLEY: For example, in certain diseases--I live in the world of prostate cancer, the patients insist upon PSA to look for relapse. There are other diseases as well where the patients insist upon some type of radiologic imaging to look for relapse. Believe me, it is very difficult to explain to the patient that I don't really know if this is in your best interest.

DR. PRZEPIORKA: The other issue is always medical-legal. If you miss a diagnosis the patient always comes back and says, well, maybe I would have survived two years longer had you caught my tumor before it became symptomatic. So, that is another big issue. Dr. Rodriguez?

DR. RODRIGUEZ: The reality is, at least in the patient subset that I follow and I mostly treat patients with lymphomas, is that they can have other malignancies, not just lymphomas and that the second or third malignancies could be potentially curable if caught early. So, that is another overlying concern.

DR. PRZEPIORKA: Dr. Cheson?

DR. CHESON: However, there have been two

to three randomized trials--you say you don't know whether it is ethical or not to do them--in which patients with lymphoma both Hodgkin's lymphoma and non-Hodgkin's lymphoma, have been randomized to looking at patients presenting with symptoms, physical examination and simple things like that versus regular CT scans at certain intervals, and the overall outcome was identical. The patient was in general the best indicator of when the disease was coming back, although we all have patients where we do pick up things early and, in the grand scheme of things, survival was not adversely effect in any of those three studies.

DR. PRZEPIORKA: Dr. Williams?

DR. WILLIAMS: I wonder if all this discussion mostly refers to settings where the disease has gone away and you are not treating them. I am thinking that when you are giving cytotoxic therapy I think a lot of investigators feel like they need to know whether there is progression or not and generally they tend to stop the treatment, cytotoxic treatment--Dr. Temple brought up the question, if it is not a toxic treatment do you really need to know or you can just continue the drug anyway.

DR. TEMPLE: Of course, we don't really know if it is time to stop a therapy just because it has progressed. Maybe it is still providing benefit. We have had lots of conversations with companies about that with these newer non-cytotoxic therapies. But I guess if it is cytotoxic everybody wants to get rid of it.

DR. PRZEPIORKA: Other comments? Yes? Could you come up to the microphone? If you could just identify yourself for the record, please?

DR. SRIDHARA: Yes, I am Raji Sridhara, from FDA Biometrics. I am team leader. I have a question going back to the first one that George and Fleming commented on. You know, when you have crossover you are saying that, okay, it can't be helped; it happens and we leave it at that. I think we get to a point where actually the design is such that your primary endpoint is survival and then you don't know how much you will cross over and at the end you will have some crossover and you are left with all these secondary endpoints which were never powered properly, or we don't have specific secondary endpoints. Would you rather suggest then that we should have specific secondary endpoints which we can rely on just in case the

primary analysis is not feasible because of too many crossovers, loss to follow-up or any of those?

DR. GEORGE: You are bringing up a very good point. I think there was an issue some time ago, not in cancer, that came before the FDA in which the primary endpoint was not survival. The survival endpoint seemed to show a survival advantage and then what do you do? You know, it didn't show something in the primary endpoint which was not survival but did show a survival advantage in a surprising way; you didn't expect it. Could you get approval? That is not a question for me I guess.

DR. TEMPLE: Well, in other settings, other than cancer, the unexpected discovery of survival benefits turns out, not surprisingly, to carry a lot of weight. We agonize a lot but we tend to say, hm, that is good.

DR. GEORGE: I think so. I mean, I think that is the right kind of approach but you can get yourself into conundrums with saying this is the primary endpoint; survival is secondary. But to answer your question, if you really think all of the crossovers and subsequent treatments are going to be a serious issue in the trial you really do

have to rethink whether survival is the proper primary endpoint, and in those settings it may not be.

DR. SRIDHARA: Picking up on what you said about other settings where there was a survival advantage or where it was not termed as the primary endpoint, then should we be considering in all these settings co-primary endpoints survival and time to progression so that it will allow us to look at either one of them? Since generally until the trial is over we don't know really how much crossover is going to happen.

DR. GEORGE: What does is a co-primary endpoint mean? Does that mean you have to meet both of the objectives?

DR. SRIDHARA: One or the other, or however you want--it depends I guess on the disease setting and what we are doing.

DR. TEMPLE: Sorry, did you ask about co-primary?

DR. GEORGE: Yes, what does that mean?

DR. TEMPLE: Usually people divide the alpha appropriately, whatever appropriately turns out to be. There have been cases, but not mostly in oncology, where we expect a benefit on more than

one endpoint. But, as everybody knows, that becomes a formidable challenge and we get requests to reduce the alpha or make the alpha less demanding. But usually that means people have to make some accommodation to multiplicity--always tricky.

DR. PRZEPIORKA: Dr. Fleming?

DR. FLEMING: Just to return to this point, it seems to me that therapeutically what we are trying to do is improve the regimens and the therapeutic strategies. I think that was the term that Dr. George used earlier. We are looking at comparing a therapeutic strategy involving the experimental agent versus the standard of care strategy and trying to show that this experimental strategy is, in fact, better in a tangible way to patients. Obviously, that means that we should be delivering care in an optimal fashion and when the first intervention to which you are randomized leads to failure at some level you are going to follow-up with best supportive care, as you should.

In fact, we would hope that we can improve on strategies that will ultimately lead to an improvement in survival relative to what is available in the standard of care. So, clearly, in

many settings it would be an appropriate endpoint. But there are many other settings where it may not be anticipated that that would be the most sensitive measure to what beneficial influence we provide to patients. If, in fact, that is in part because of crossovers diluting the long-term survival effect, I would still argue that is the truth. That is what I am ultimately doing on survival. There may be need for other measures. I would argue that those other measures ideally should be direct clinical measures of benefit, measures reflecting improvement in functional status; measures that reflect overall improvement in symptoms. With bisphosphonates, for example, what we have gone to is skeletal related events as an alternative clinical efficacy measure. Beneficial effects may be reflected in survival but a more sensitive clinically tangible measure may be the measure in reduction in fractures and spinal cord compression and radiation and surgery to the bone, other rescue therapies. So, if I can improve that measure that is clinically tangible benefit. I would rather see that measure being the co-primary endpoint rather than a surrogate measure, unless that surrogate has been truly

validated.

I just want to come back to one of my colleague's earlier points that was raised in the criticism of time to progression. You are absolutely right, we want to do high quality studies. So, we are going to presume that people are going to the very best study they possibly can on whatever endpoint they are looking at. However, certain endpoints lend themselves to more readily being assessed in an unbiased, objective way. In an unblinded trial it is much more problematic when you have an endpoint that requires judgment, such as a symptom endpoint or a time to progression endpoint, as opposed to survival. And, missingness has over history been more of a problem when we are looking at these markers as opposed to survival as an endpoint. In particular, as we have said, with time to progression we are building in missingness because automatically time to progression, by censoring deaths, means you are missing what happens in time to progression subsequent to death in those patients who die. So, there are some inherent problems that exist with lack of blinding and with censoring deaths that even in the best quality study you are going to have some

difficulties with.

DR. PRZEPIORKA: If I could just summarize--

DR. GRILLO-LOPEZ: I disagree with that.

DR. PRZEPIORKA: Feel free.

DR. GRILLO-LOPEZ: I cannot agree that you can measure survival better than time to progression. I think that if you have an appropriately designed trial with the appropriate interval for CT scans you can measure time to progression better than you can measure survival because of all the biases in the survival measurement that I mentioned earlier. So, it all depends on how you design your protocol; how you schedule your evaluations and how good the quality of the data is. Again, there are so many biases inherent to the survival kind of endpoint that it is not an acceptable endpoint in most situations, in my mind at least.

The other thing that I would like to mention is that the issue of crossover goes away completely if you are not using survival as an endpoint. It is an important issue because if you have a drug, a new agent that has gone through Phase II trials you know of its clinical activity;

you know of its safety and you know what the patients know of its clinical activity and safety because they go ASH and they go to ASCO and they go to the websites and they know that there is an option which in some situations, in the refractory setting, may be the best option for them and they are not going to go into a Phase III trial and take a 50 percent chance of being randomized to a standard therapy that may not be as good in fact as the experimental therapy and never have the chance to get the experimental agent unless they know that there is some opportunity, not perhaps within the same protocol but some time later on, to get the experimental agent.

DR. FLEMING: But your response is presuming that access to that intervention on a delayed basis is going to provide the essence of what the benefit is when you deliver it up front--in some settings more plausible but in other settings much less plausible. And, your response hasn't addressed the issue of the inherent risk of bias that arises in what is typically done in oncology, which is unblinded trials, and it hasn't addressed the issue of the informative censoring that arises if you choose to censor deaths.

DR. GRILLO-LOPEZ: But that is not my assumption. I am saying that it is the patient's assumption. It is the patient's assumption that there is benefit and they want to get that experimental--

DR. FLEMING: That doesn't matter if it doesn't, in fact, carry a substantial part of the overall benefit up front. It doesn't matter if that is the patient's assumption.

DR. GRILLO-LOPEZ: You miss the point. What I am trying to convey is the difficulty of doing a Phase III randomized trial if the patient knows that he has only a 50 percent chance of getting an agent which the patient perceives as an active agent.

DR. FLEMING: The Evastin trial in colorectal cancer was just successfully completed in a manner that you are saying couldn't have been done.

DR. GRILLO-LOPEZ: It may be an exception.

DR. TEMPLE: Surely a company can control whether it makes an experimental drug available to everybody and allows crossover or not. It is their drug.

But I thought the earlier point you made,

and it is one of the reasons we are here, is crossover doesn't matter if you are measuring time to progression because crossover happens after that.

DR. FLEMING: If, in fact, time to progression is the answer to the question that we care about and can be addressed without the problems of these other biases that arise so it is not getting us out of the woods.

DR. TEMPLE: No, it just solves one problem.

DR. PRZEPIORKA: Dr. Brawley, last comment?

DR. BRAWLEY: Well, it was actually somewhat of a question. It is just sort of a gut check. I am just sort of remembering all those trials, many of them not in cancer treatment but in other areas where initial endpoints and initial surrogates seemed to be very positive and then, when we finally got to the randomized clinical trials we found out that the intervention actually was not as positive. I am thinking specifically right now of premarin in the Women's Health Initiative, although I have some rumblings of Iressa Phase III clinical trials in the back of my

mind, Iressa trials using Iressa and chemotherapy as well. We have to be very careful as we go down this path.

DR. PRZEPIORKA: A very good point. If I could summarize what I heard, there are actually a few parallels to our discussion on disease-free survival. Specifically for time to progression, we did not think that a single endpoint design would be attractive at all. There is concern about death on therapy and perhaps progression-free survival might be better than just time to progression.

We agree that there has to be rigorous assessment for scientific reasons, not for clinical reasons. So, repeated assessments may be done in studies where we would not usually do them in clinical medicine but we do want to get the scientifically valid results.

We would not use this therapy for patients who are very symptomatic because progression there would not be good for those patients as opposed to really trying to get a response. And, toxicity needs to be factored in as a risk-benefit for whether or not this is something useful.

So, it appears that progression-free survival would be for diseases with low CR rates in

therapies that would be unlikely to alter survival because of the underlying disease to be used as a primary endpoint, but in a comparative study when standard therapy is already shown to have a benefit it would probably only be as opposed to a real endpoint. Any other comments on that summary? Dr. Temple?

DR. TEMPLE: One of the points was that we don't expect these drugs to alter survival. I guess I am not sure that is the assumption. We think it may be difficult to demonstrate that because of crossover and because it is going to occur later, but I guess I think one of the assumptions is that if you have an effect on time to progression, or something like that, it probably does have a favorable effect on survival even if you are not able to measure it very well. Am I wrong in that?

DR. PRZEPIORKA: I don't think I would disagree with that but I think time to progression would be an excellent endpoint in a disease such as metastatic prostate cancer in the elderly where, no matter what you do, they are going to end up dying of non-cancer reasons. Whereas, if you can keep them symptom free it would be very valuable.

DR. TEMPLE: Actually, the last point is one we didn't talk much about, survival is tough if it is an old population that is dying of a lot of other things. We didn't really discuss that but in prostate that is probably a major factor.

DR. PRZEPIORKA: We will close this session with an announcement about lunch.

MS. CLIFFORD: The statement I made earlier, unfortunately, is not true about your badge. It will not grant you access into the building next door. I am sorry. At the front desk there is a list of six restaurants that are local, that are within walking distance that you are welcome to visit. Thank you.

DR. PRZEPIORKA: We will reconvene promptly at 1:00 p.m. Thank you.

[Whereupon, at 12:05 p.m., the proceedings were recessed for lunch, to reconvene at 1:00 p.m.]

A F T E R N O O N P R O C E E D I N G S

DR. PRZEPIORKA: In this afternoon session we will discuss non-small cell lung cancer endpoints and we do have a different group with us this afternoon so, for the record, I would like to go around the table one more time with introductions for everyone who is new this afternoon and everyone from this morning. If we can, let's start with introductions with Dr. Ettinger, if you could let us know who you are and where you are from, please.

DR. ETTINGER: David Ettinger, the Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins in nearby Baltimore.

DR. SAXMAN: Scott Saxman, in the Cancer Therapy Evaluation Program of the National Cancer Institute.

DR. BONOMI: Phil Bonomi, Rush Medical College, Chicago.

DR. JOHNSON: David Johnson, Vanderbilt University in Nashville, Tennessee.

DR. JOHNSON: Bruce Johnson, from the Dana Farber Cancer Institute.

DR. GRILLO-LOPEZ: Antonio Grillo-Lopez, acting industry representative.

DR. GEORGE: Steve George, Duke University.

DR. CHESON: Bruce Cheson, Georgetown University Lombardi Comprehensive Cancer Center.

DR. DOROSHOW: Jim Doroshow, City of Hope Comprehensive Cancer Center.

DR. RODRIGUEZ: Maria Rodriguez, M.D. Anderson Cancer Center.

DR. BRAWLEY: Otis Brawley, Emory University, Winship Cancer Institute.

MS. ROSS: Sheila Ross, Washington representative for Alliance for Lung Cancer, and I am a lung cancer statistic.

DR. FLEMING: Thomas Fleming, University of Washington.

DR. LEVINE: Alexandra Levine, University of Southern California Norris Cancer Center.

DR. REAMAN: Greg Reaman, George Washington University and the Children's Hospital in D.C.

DR. PRZEPIORKA: Donna Przepiorka, University of Tennessee Cancer Institute.

MS. CLIFFORD: Johanna Clifford, FDA.

MS. HAYLOCK: Pamela Haylock, oncology nurse from Texas.

DR. CARPENTER: John Carpenter, University of Alabama at Birmingham.

DR. REDMAN: Bruce Redman, University of Michigan Comprehensive Cancer Center.

DR. TAYLOR: Sarah Taylor, University of Kansas Medical Center.

DR. LI: Ning Li, FDA Biometrics.

DR. KEEGAN: Dr. Keegan, CDER Office of Drug Evaluation VI.

DR. WILLIAMS: Grant Williams, FDA Drugs.

DR. TEMPLE: Bob Temple, Director of ODE I.

DR. PRZEPIORKA: This afternoon's session is actually split into two. The first will be three talks regarding non-small cell lung cancer and clinical trials. We will have a brief break, followed by an open public hearing and then address the questions that have been posed to us by the FDA. We will start this afternoon's session with a talk by Dr. Cohen on non-small cell lung cancer, the regulatory background.

Non-Small Cell Lung Cancer Regulatory Background

DR. COHEN: I am going to review the approval in lung cancer that the agency has made through the years.

[Slide]

The data that I am going to present is the data that is in the individual labels for each drug. So, the data may be somewhat different from published data that you would find for each of these trials.

[Slide]

For non-small cell lung cancer there have been first-line approvals, second-line and third-line. There were five approvals for first-line. All of these approvals were regular approvals. For second-line there has been one approval, also a regular approval. For third-line non-small cell lung cancer there is one recent approval which was an accelerated approval. For small-cell lung cancer second-line there has been one regular approval and there has been one approval for palliation of non-small cell lung cancer.

[Slide]

This is a listing of the five approvals for first-line non-small cell lung cancer. There was one single agent, vinorelbine and four approvals for doublets containing cisplatin, and the doublet partners have been vinorelbine,

gemcitabine, paclitaxel and most recently docetaxel.

[Slide]

What I am going to do in the next group of slides is review each of these approvals. This is the vinorelbine approval. The approval was based primarily on an improvement in one-year survival and also, as supporting evidence, there was improvement in response rate. In this trial the comparator regimen was 5-FU leucovorin given in the Mayo Clinic type regimen.

There were 211 patients entered into the study. There was a 2:1 randomization in favor of vinorelbine. As you can see, the response rates were 12 percent versus 3 percent. Median survivals were 30 weeks versus 22 weeks and one-year survival was 24 percent versus 16 percent. The p value refers to the difference in the survival curves.

[Slide]

Vinorelbine/cisplatin was evaluated in two studies. In the first study vinorelbine/cisplatin was compared to cisplatin alone and 432 patients were entered. Response rates favored the combination therapy. Median survivals were 7.8 months versus 6.2 months. One-year survivals were

38 percent versus 22 percent, and the p value for the survival comparisons were 0.01.

The second study was a three-arm study that included vinorelbine, cisplatin compared to vinorelbine alone and the third arm was vindesine/cisplatin. You can see that the response rates in this study favored the vinorelbine/cisplatin combination. Median survivals were 9.2 months versus 7.2 months for vinorelbine alone versus 7.4 months for the vindesine/cisplatin combination. One year survivals were as listed. The p value for survival comparing vinorelbine/cisplatin to vinorelbine alone was 0.05 and the p value for the comparison of vinorelbine/cisplatin versus vindesine/cisplatin was 0.09.

[Slide]

Gemcitabine/cisplatin was also evaluated in two randomized trials. In the first trial the comparator regimen was cisplatin alone. There were 522 patients entered. Response rates were 26 percent versus 10 percent favoring the combination. Median survivals were 9 months versus 7.6 months and the p value for that comparison was 0.008.

In the second study, which was somewhat

smaller, the comparator regimen was etoposide/cisplatin. The response rates were 33 percent for the gemcitabine/cisplatin regimen versus 14 percent for the VP16/cisplatin. Median survivals were 8.7 months and 7.0 months. As you can see, that survival difference was not statistically significant.

[Slide]

Paclitaxel/cisplatin was evaluated in an ECOG trial that was a three-arm trial. The first arm included paclitaxel 135 mg/m2. There was a 24-hour infusion with cisplatin. The second arm was paclitaxel 250 mg/m2 with cisplatin. The comparator regimen was etoposide/cisplatin.

As you can see, both paclitaxel regimens had an increased response rate as compared to etoposide/cisplatin. Median survivals were 9.3 months for paclitaxel 135, 10 months for paclitaxel 250 with cisplatin and 7.4 months for the VP/cisplatin regimen. In terms of survival, which is listed on the bottom on the right, the survival comparison of paclitaxel 135 mg/m2 plus cisplatin compared to etoposide/cisplatin, the p value was 0.08 and for the paclitaxel 250 mg/m2 the p value was 0.12. However, if you look at response rates

which is a), and time to progression which is b) on the bottom, both of these were statistically significant in favor of the paclitaxel regimens, with paclitaxel 250 doing somewhat better than paclitaxel 135.

[Slide]

Docetaxel/cisplatin was evaluated against vinorelbine/cisplatin and also against docetaxel/carboplatin. A total of approximately 1200 patients were entered into this study. As you can see, the median survivals were relatively similar for all three regimens. This was a non-inferiority analysis and doing the non-inferiority analysis docetaxel/cisplatin retained greater than 50 percent of the therapeutic benefit of vinorelbine/cisplatin. On the other hand, docetaxel/carboplatin did not. So, the docetaxel/cisplatin regimen was approved.

[Slide]

Docetaxel was also evaluated as a second-line treatment regimen in two studies. In the first study docetaxel was compared to best supportive care and 104 patients were entered. The response rate to docetaxel in this patient population was 5.5 percent. Median survivals

favored docetaxel, 7.5 months versus 4.6 months, with a p value of 0.01.

The second study involved docetaxel compared to chemotherapy that was investigator's choice and 248 patients were entered. The response rates for docetaxel were again in the 5-6 percent range. The median survivals were comparable for docetaxel and investigator's choice chemotherapy. But one year survival for docetaxel was 30 percent versus 20 percent for investigator choice, and that p value was significant at less than 0.05.

[Slide]

Gefitinib or Iressa was recently evaluated as a third-line treatment regimen in patients who had failed a platinum and who had failed docetaxel. There were 143 patients who met these eligibility criteria. They were randomized to receive Iressa 250 or 500 mg/day. Overall, if one combines the two treatment groups and that was done because it was relatively comparable for each group, the overall response rate was 10.6 percent with a 95 confidence interval, as listed, and it was of interest that in exploratory analyses response rates were higher in females, in nonsmokers and in patients with adenocarcinoma.

[Slide]

The one approval in small cell lung cancer was Hycamtin or topotecan and that was compared to CAV, Cytoxan, adriamycine and vincristine. The eligible population for this trial were patients who had responded to first-line treatment and who had then progressed greater than or equal to 60 days after stopping treatment. There were 107 patients in the Hycamtin arm, 104 patients in the CAV arm. The difference in this study was only in response rate. The response rate was 24 percent for Hycamtin versus 18 percent for CAV and this difference in response rate was felt to be of sufficient importance to warrant approval.

[Slide]

The one palliative approval in non-small cell lung cancer involved photofrin photodynamic therapy, and that was compared to nd:YAG laser therapy. The patient population eligible for this study were individuals with symptomatic obstructive bronchial lesions. Symptom severity scales were used as the evaluation tool. Symptoms rated were dyspnea, cough and hemoptysis. Photofrin therapy was of comparable efficacy to nd:YAG laser therapy.

[Slide]

So to summarize the approval endpoints, in first-line, as I mentioned earlier, there were five studies. Three of the approvals were based on superior survival. One approval was based on non-inferior survival and one approval was based on superior time to progression and response rate with a trend toward improved survival.

In the second-line setting there was one study and approval was based on superior survival in that study. In the third-line setting, which was the one accelerated approval in non-small cell lung cancer, the accelerated approval was based on response rate. And, there was one approval based on symptom palliation.

[Slide]

In second-line small cell lung cancer there was one approval and that approval was based on response rate. That concludes my presentation.

DR. PRZEPIORKA: Thank you. We will hold questions until all three speakers have had the opportunity to presentation. Next, Dr. Paul Bunn will talk about the FDA ASCO non-small cell lung cancer workshop.

FDA/ASCO Non-Small Cell Lung Cancer

DR. BUNN: Members of ODAC, members of the FDA and guests, I would first like to say that I am honored to be here. It is a privilege to be here and I want to mention that I take this extremely seriously because what I do for a living is to take care of lung cancer patients and I think what you are deliberating is extremely important.

[Slide]

With respect to the history of why we are here, Rick Pazdur, in his infinite wisdom, I think agreed with a comment that Bruce Cheson made this morning and that is not all cancers are the same and in the future it is highly likely that we are going to have to look at these endpoints in individual cancers based on data from the individual cancers, not based on feelings but based on data from these individual cancers. Of course, this morning we heard a lot of theoretical discussion. Hopefully, this afternoon we are going to be talking about data-driven discussion.

So, to put the data into context, the FDA and the American Society for Clinical Oncology had a series of telephone conferences and a single open public hearing discussing endpoints for approval of

drugs for lung cancer. What you are hearing this afternoon is somewhat of a rehash of that. You will be asked some questions based on what you hear.

The way we have done this is that we have divided the discussion into two topics. The first topic is what has been called classical endpoints. The classical endpoints that we discussed were objective response, time to progression and survival. For whatever reason, we called another one non-classical endpoints. The distinction I think is incorrect but, anyway, that was largely patient-reported outcomes. After I get done talking about the classical endpoints of objective response, time to progression and survival, Richard Gralla is going to talk about patient-reported outcomes.

I have an apology to make. The slides that you have in front of you--my secretary and I were in a miscommunication mode and they have nothing to do with what I am going to say--

[Laughter]

--so don't bother looking at your handout. You will be very confused. You will actually have to look at the slides and I apologize for that.

Before I actually begin I want to make one correction to what Marty said and one other comment. Actually, the Albain study of vinorelbine/cisplatin versus cisplatin happened after the approval. Actually, the LeChevalier study for the combination was the primary study and the Crawford study for single agent was the primary study. The Albain study actually came later and confirmed what happened but was actually not known at the time of the ODAC presentation. I know because I am old and I was there.

I have great respect for the consultants here. I also have great respect for Dan Ihde. What I am going to say is something that I think in 1985 Dan Ihde and I agreed on and I wish he were here to agree with me now that what happened in 1985 was a big setback to lung cancer drug approvals.

[Slide]

I am going to begin by trying to keep this simple, stupid! Why are we here? Drug development takes enormous amount of fiscal resources and long periods of time. Currently we know more about novel targets than ever before. At the same time, there are fewer new drug applications. We could

ask why is that. It is undoubtedly for many reasons. It is possible that stringent FDA requirements for approval at the moment are a deterrent to new drug applications.

I think we could all agree that most knowledge about drug utilization and toxicity occurs after the initial approval. We might also agree that if we had safe and efficacious drugs, expedited drug development might benefit society. Therefore, I think it is appropriate that we are looking here at criteria for endpoints for NDAs, or new drug applications, for lung cancer.

As you heard this morning, FDA regulations require that drugs be safe and efficacious for a defined population by adequate and well-designed clinical trials. As you also heard this morning, simple statements are sometimes gray, not black and white. As you also heard this morning, FDA legislation does not require that a drug be shown to be superior to other drugs. It has to be safe and efficacious; it doesn't have to be better than approved drugs, with a single exception which I believe should be discussed openly and frankly in this afternoon's deliberations. Oncology drug divisions is determined that drugs given

accelerated approval should offer an advantage over existing agents.

DR. TEMPLE: It is in the reg.

DR. BUNN: It is in the reg? Okay. Well, we are going to discuss this during my presentation.

[Slide]

The question is, well, why would be here just for lung cancer? What are some of the differences between lung cancers and other diseases? One of the difference is that almost all the patients, three-quarters, present with advanced disease. That is, they are III or IV.

Most studies show that 90 percent of patients or more are symptomatic at the time of presentation. So, our discussion this morning about whether patients would be symptomatic or not, in lung cancer the basic idea is that they are symptomatic. When they get relapse they are symptomatic; when they present they are symptomatic. The majority of patients have co-morbid cardiopulmonary disease. Dr. George was talking about deaths from unrelated causes. This is a huge problem in lung cancer. If you look at trials of adjuvant radiation and adjuvant

alkylating agents the hazard rates are 1.2, so a 20 percent increase in the hazard rate of death is not due to the disease but it can accelerate the disease. Many of those deaths are not actual toxic deaths that you would define as a toxic death but these are sick people and when they get tough treatments sometimes they die.

In the current SEER data in the U.S. the median age is 70 years old. The majority of these patients are elderly. Recruitment to surgical trials is extremely difficult. In this disease at the moment, unfortunately, complete responses are rare. So, talking about disease-free survival is an oxymoron when you talking about stage IIIB and IV lung cancer. We don't have to have that discussion that we had this morning; it doesn't happen.

It used to be that objective responses or 20 percent were very rare. Fortunately, we have drugs that work now. We have drugs that make people live longer and objective responses oftentimes do occur in more than 20 percent of patients.

It used to be that second-line therapy did not influence survival but now, as you heard from

Dr. Cohen, it does. So, some of the issues we heard this morning about second-line therapy influencing survival will be an issue.

[Slide]

So, classical endpoints--objective response. Up until 1985 this was a major deal. In 1985 Dan Ihde, along with the FDA, looked at a bunch of data and there was not a wonderful correlation between response and survival. That probably would be true today for melanoma and other diseases where responses over 10 percent are rare. We are going to re-discuss that now in 2003 to actually look at what the relationship is between response rates and survival.

Time to progression has not often been used because it is very difficult to assess and, in the past, because second-line therapy didn't affect survival. The difference between progression and survival was very short but we will have a little bit of discussion about that. Survival I guess is not only FDA's favorite endpoint. As you heard this morning, most of us can agree that it is a real and important endpoint.

[Slide]

So, in the past objective response rates

were quite variable, not consistently assessed; did not always correlate with survival and most agents, such as the alkylating agents and the athrocyclines were toxic to smoking patients. Some of these agents produced response in up to 20 percent but rarely higher of untreated patients but there was no survival improvement. Thus, in 1985 the FDA decided that objective response rate was not definitely associated with patient benefit.

[Slide]

What happened since that time? I think that this is a very important study and one which really needs to be updated. In fact, after this morning's discussion I am thinking about having one of my fellows go back and actually do this. I partially did this but not in a real meta-analysis.

But there was a study that looked at the correlation between response and survival in 176 Phase II trials with 7000 patients between '76 and '95. Since that time, the drugs that Dr. Cohen mentioned have largely been approved and were not part of this. The average response rate in these trials was only 11 percent. I think since 1995 we are in a different place.

In these 176 trials they found 12 drugs,

or 11, that had a response rate of more than 20 percent. Those are cisplatin, vinorelbine, docetaxel and paclitaxel. As you heard, all those are approved. This also included small cell so irinotecan, etoposide, vindesine, epirubicin and ifosfamide and edatrexate showed up in that list.

They also did a correlation between response rate and survival time. You can see that the correlation coefficient and the p value. Then they did a logistic regression coefficient and you can see the p value between the relationship between response and survival was 0.0003.

[Slide]

So, what has happened since 1995 in terms of what is in the literature? These are the drugs that most of us would consider the most active cytotoxic drugs. We have the Phase II single agent studies of these drugs in untreated advanced non-small cell lung cancer. As you can see, these have response rates--these are limited institution studies now, not the big cooperative groups and I will get to those. They had response rates varying from 20 percent to 27 percent. They had median survival times ranging from 7.6 months to 9.7 months and one-year survival rates ranging from 22

percent to 41 percent. I think from historical controls, any of us would say, if you are an optimist, the median survival would be 5 months and the one-year survival rate would be 10 percent. Vinorelbine, as you heard a moment ago, is the only one of these drugs approved for non-small cell lung cancer.

[Slide]

What about multi-institution Phase III trials with these same therapies. You can see here that, again, there are large numbers of patients but there are some differences. The response rates before varied from 20 percent to 27 percent and now the response rates vary from 16 to 18. Why is that? The primary reason for that is that the cooperative groups require a post CT scan done four or more weeks later and most trials have them done eight weeks later. Many of the patients don't have the second scan and those are unconfirmed responses and the cooperative groups don't count those patients as having a response. So, it is generally true--and some of the ECOG or other people could comment on this--that in the multi-institutional cooperative group trials response rates are approximately five percent lower than in the

limited institutions primarily for that reason.

You can also see that the confidence intervals around these response rates are actually quite narrow. Largely, that is because people can actually use RECIST and actually have objective response rates that are fairly reproducible. Median survivals in these trials range from 6-7 months and one-year survival from 25-33.

[Slide]

I am going to come back to first-line therapy after a minute but something new happened, and that is patients are living longer. Now, just remember that the minority of patients have benefit. If you have a response rate of 20 percent means that most patients aren't having any benefit. Now, median survival is not likely to change a lot when 10 percent or 20 percent of the patients are benefited. Two-year survival goes from 1 percent to 20 percent in advanced lung cancer with treatment but median survival only goes up by a couple of months.

In the second-line setting the drugs that have been approved and the drugs that we think about the most are shown here. Response rates range from 9 percent to 16 percent in these trials

although the confidence intervals and the ranges are much broader in the second-line limited institution setting than they are in the first-line setting.

[Slide]

With respect to multi-institution Phase III single-agent therapy in non-small cell lung cancer, the data from the trials that we have had are listed here. Response rates vary from 8 percent up to 14 percent. Now, as you heard, docetaxel is approved and gefitinib is approved. Question number six in your handout could be viewed as a pre-setting for a pivotal trial looking at pemetrexed in the second-line setting and the response rate, median survival and one-hear survival from that trial are shown here.

[Slide]

So, a question that I hope you all will address, because I think it is extremely important--in 1985 it was basically determined that objective response was not either a likely patient benefit or a definite patient benefit, and in my opinion objective response that exceeds a certain threshold should be considered as likely evidence for patient benefit--likely, not proven. In Dr.

Fleming's terms this morning, that would be his group C. I think that objective response over 20 percent in untreated patients is a likely surrogate for patient benefit. It is possible that meta-analysis could change that into a definite evidence of patient benefit, as documented by symptom relief and/or survival.

Every drug that we know of with a response rate over 20 percent in limited institution trials and over 16 percent in multi-institutional trials has been shown in randomized trials to affect survival, and most of them have been shown to relieve [sic] patient benefit. I am not going to discuss patient benefit in terms of symptoms because Richard Gralla is going to talk about that.

So, if one could consider that objective response is a likely indicator of clinical benefit, the question is could accelerated approval be given based on objective response rates? Certainly, I think that they could. One could say that if the surrogate is definite it is full approval. If the surrogate is likely, it is an accelerated approval. Well, I believe it is likely. It could be definite but I think it is likely so it should be considered for accelerated approval.

Another thing is that RECIST criteria I believe are actually good and can be reviewed independently by the FDA and independent committees. So, I believe that the endpoint we are talking about here is a reproducible endpoint.

[Slide]

In the first-line setting one could argue that if an agent had an objective response rate of more than 20 percent in a limited institution study or 15 percent in a multi-institution trial that a drug might be given accelerated approval. One could argue in a second-line setting active agents have objective response rates of more than 10 percent in limited institution studies and more than 8 percent in multi-institutional studies.

Now, to demonstrate this type of response is actually not trivial. These data are I think almost right but not exactly right. I have a little bit better data from Dr. Piantidosi. If you want to show that a drug has a 25 percent response rate, plus/minus 5 percent, a 95 percent confidence interval of 5 percent, Dr. Piantidosi informs me that would be a 400-patient trial. If that goes to plus/minus 4 percent the number would be 625 patients.

[Slide]

This is not actually just an academic consideration here. Not all the drugs work that are developed. Current FDA policy promoting Phase III survival trials have led to the institution of multiple Phase III trials after the completion of a Phase I trial even when no single-agent activity was observed in the Phase I trial. No inactive drug has ever been shown to improve survival or improve patient symptoms when used alone or in combination with chemotherapy. However, going straight from Phase I to Phase III has led to multiple negative trials costing not thousands but millions of dollars and thousands, not hundreds, of patient live resources.

Examples of randomized trials of agents not showing any activity up until the time of a survival Phase III trial are shown here, tirapazamine, MMPIs and a Gentasense compound and a whole bunch ongoing.

[Slide]

This is what we have learned from these trials. These inactive agents when combined with active agents do nothing. This particular negative trial had 700 patients. No benefit to the patient.

Probably approximately 100 million dollars wasted. If objective response had been available to get accelerated approval, people would throw away the inactive drugs. Because they can't get accelerated approval for active drugs, they go straight from Phase I to Phase III, waste millions of dollars, thousands of patients lives. I would submit this is not a good state of affairs. Obviously, you may all disagree but it is not my favorite thing.

[Slide]

Single-agent activity of tirapazamine has never been established. Nonetheless, for the same reason multiple Phase III trials were done. Interestingly enough, one of these Phase III trials, shown here, showed an improvement in response rate of tirapazamine/cisplatin versus cisplatin. The response rate was higher, survival was higher but when this was done in another trial response rate was not improved nor was survival. This does show why we should also discuss in certain instances why you might want two trials instead of one. Perhaps we can discuss that.

[Slide]

Now, some drugs that get developed are not all that far from patent exploration. When

companies need a Phase III survival advantage trial to get a drug approved and it is going to take five years and they are four years away from their patent expiring, they may not want to develop the drug. So, a drug called oxaliplatin was done as a Phase II trial in lung cancer. Interestingly enough, it was done in performance status II patients which, as everyone knows, is a very bad group of patients. The response rate was 15 percent, median survival was 8 months and there was not a single grade III or IV hematologic toxicity.

If accelerated approval was available for this drug, on the basis of this probably one would want to do a big trial to try to get accelerated approval. The huge question is whether this drug will ever see the light of day for lung cancer patients because of the current interpretation of how to get a drug approved.

[Slide]

When we get into combinations response rate sometimes gets a little trickier. This is a trial that makes us all humble of course and it highlights the issue about response and median time to progression, and perhaps would be used to say that there should be surrogates for likely benefit,

not definite benefit.

This was a study from Germany that compared cisplatin to Taxol and cisplatin. The Taxol and cisplatin arm had a much higher and statistically significant higher response rate. It also had a statistically improved median time to progression. On the other hand, survival was actually a little worse, not statistically so but a little worse in the combined therapy arm.

I don't know what to make of this trial. It is certainly an outlier and it shows why outliers happen. One could argue that this is why objective response and time to progression should be surrogates as opposed to definite relationship to patient benefit.

[Slide]

Now, if accelerated approval was actually available and people took advantages, where would be today? Actually, docetaxel, paclitaxel, gemcitabine, irinotecan, pemetrexed and cisplatin would be approved for lung cancer and I don't think there is a single person in this room who thinks that would be bad. Drugs that would not be approved and have either been shown not to be useful under Phase III trials at the moment are

equally as many. And, why do we have to go through large, 1000-patient, randomized trials for inactive drugs?

There were drugs approved, vinorelbine and gefitinib, and gefitinib was actually approved by accelerated approval based on response. That precedent that you all set--I think what you did was right. I think what you did should be common, not uncommon. Not every active agent has a response rate over 20 percent. Carboplatin, I think most of us would agree, is a useful drug and makes people with lung cancer live longer but doesn't have a response rate over 20 percent.

[Slide]

So, just to reemphasize what you did, if gefitinib had not been studied in large numbers of patients and approved based on response rate, it would be gone because the company did what all the other companies have been doing, going straight from Phase I to Phase III, and they did that as well. They went straight into combined studies. As you all know, those trials were negative.

Besides the fact that most of us think that lonafarnib and gefitinib are drugs that should be approved for lung cancer, we have to learn how

to use them. Look at the time to progression in these trials. After the chemotherapy was stopped the groups that got gefitinib did better than the group that got placebo in both trials. I think everybody in this room thinks we need to understand why that is. We wouldn't be able to understand why that is if these drugs were not given accelerated approval--these would be gone.

[Slide]

Now I am going to talk a little bit about these EGFR inhibitors--

[Slide]

Before I do I want to say one thing, FDG PET hasn't been studied nearly as much as CT response. In every trial comparing CT response to PET response, PET response is correlated with survival better than CT response. There is not a single trial were PET response is not correlated with survival. I think, if nothing else, we should be encouraging our pharmaceutical colleagues to consider this for development as a potential surrogate endpoint that actually could be better than actually objective response by CT.

[Slide]

So, what about subsets? Lung cancer is

not one disease. We heard this morning that leukemias are not all the same. Bronchoalveolar carcinoma and large cell neuroendocrine carcinoma are not the same disease. Small cell carcinoma is not the same as non-small cell carcinoma. What are we going to do about subsets? If we require that for a subset approval a company has to do a Phase III survival trial, forget subsets. Forget it. If companies can get accelerated approval based on response rates in subsets, we might be able to make some progress.

[Slide]

Everyone sitting at this front of the room can identify as a classic patient with bronchoalveolar carcinoma, which is one subset of non-small cell carcinoma. Those of us who deal with this disease know this is not a very chemosensitive disease. We don't have a ton of data but what data we have suggests response rates are low in bronchoalveolar than in any other histology.

Anecdotally it was found that EGFR inhibitors often make responses in patients that have this chemorefratory disease. It is also anecdotally noted that these patients have high

expression of EGFR and HER-2, which was unexpected.

[Slide]

Now, we have a problem between the pathologist and the clinicians. Pathologists say that bronchoalveolar carcinoma has to be non-invasive. So, they are talking about infiltration among the alveoli septi where there is basically no invasion. They divide bronchoalveolar carcinoma into mucinous and non-mucinous forms. When we see these bilateral infiltrates what we usually have is invasive adenocarcinoma with bronchoalveolar features. So, that is something that we have to work out between the clinicians and the pathologists.

[Slide]

But as I mentioned, bronchoalveolar carcinomas have very high expression of EGFR and HER-2.

[Slide]

This is what we know about bronchoalveolar carcinoma clinically. Chemotherapy, as I mentioned, has response rates that generally are lower. So, Taxol which has a response rate of 25 percent in other Phase II trials had a response rate of 14 percent. There tends to be a little

more indolence so survival is a little bit better even despite the low response rates; median survival at one year 50 percent.

There have been two Phase II trials of erlotinib and gefitinib in bronchoalveolar carcinoma. Response rates were 24 percent and 19 percent. Median survival was 12.5 months versus not reached after 7 months. One-year survivals were 80 percent and 57 percent. Remember, these are pills compared to cytotoxic chemotherapy.

[Slide]

This is the Southwest Oncology Group, two consecutive trials, not randomized. Overall survival standard Taxol--this is the data we saw before. Response rate was 1 percent; median survival 12 months.

[Slide]

This is the data with gefitinib in the Southwest Oncology Group. The untreated patients had a median survival of 15 months and a one-year survival rate of whatever I said, 57 percent. Even the previously treated patients had a median survival of 10 months.

It is likely, when we get to randomized trials, that these single-agent pills will be

better than our standard two-drug chemotherapy. Remember, if accelerated approval had not been granted for these drugs--we only had those randomized Phase III trials--these drugs would not be seeing the light of day. And, in that large list of other drugs that went to Phase III trials, how many are actually active? We don't know because people were afraid to give approvals based on objective response.

[Slide]

Time to progression, there are a lot of problems that you heard about. One of the major of those is the frequency of assessment. We are looking at changes. Median time to progression in untreated patients is four months. A 25 percent reduction is going to be a difference of a month of less. We get CT scans every eight weeks. The frequency of assessment for time to progression is a huge issue here. Not only that, cycle length can actually affect time to progression. If the cycle length varies, therefore, the time you get the CT varies.

Another issue is sick and progressing patients may not be evaluated. Most of us who treat lung cancer patients, when they get sick and

get worse, that is the end of it. If they need a CT scan six weeks later and they have already progressed, and all that, a CT scan is not obtained. As you heard, oftentimes these patients die without any documentation of what actually happened.

[Slide]

This is an example of some of the problems with TTP that might argue it might be surrogate endpoint. This is the four-arm ECOG trial. The PIs of that trial are sitting to my right. It was comparing four different two-drug combinations. The response rates you see here. Time to progression varied from 3.3 [sic] months to 4.5 months. The 4.5 months with gemcitabine and cisplatin was actually statistically significant compared to the 3.5 [sic] months in the paclitaxel/cisplatin arm. But just remember this is a three-week cycle and CT scans are obtained every six weeks. This is a four-week cycle and CT scans are obtained every four weeks. As you can see, there is no difference in any of the survival outcomes. So, this might be a surrogate but it would be hard to say that this is a definite endpoint, definitely associated with survival and I

think in lung cancer time to progression has really a lot of issues.

[Slide]

You were talking about disease-free survival or time to progression in early stage patients. Certainly, if you progress you are symptomatic but the question is what is the timing of the assessments.

Another thing is that relapses are essentially always followed by a short survival. So, the advantage you have in some other diseases of doing this with much shorter intervals may not happen here.

Another problem is that, again, these patients are highly likely to die, not from toxic deaths but related to a toxic therapy. Those deaths are scored in very many different ways.

[Slide]

Just to show you that in the recent trials, this is a trial of a very toxic regimen, MIC. Three drugs, mitomycin, ifosfamide, cisplatin. Remember, ifosfamide-based treatments increase the hazard-related death. In this particular trial there was an improvement with the MIC chemotherapy. The hazard rate was 0.89. It

wasn't statistically significant. It certainly favored the chemotherapy. But look at what happened in survival. The people who got the chemotherapy were dying earlier. They did cross but the hazard rate for survival was 0.96 and, obviously, that wasn't statistically significant. So, if this had been a little bit better in progression-free survival there might have been an approval without an improvement in overall survival.

[Slide]

That actually happened. These are all trials, by the way, from ASCO this year or last year. This was an intergroup trial looking at chemo radiation versus chemo radiation followed by surgery. Time to progression favored the triple therapy. You can see this is the time to progression in the triple therapy and the p value was 0.02. It was better in terms of time to progression. What happened in terms of survival? The triple therapy arm had a lot of deaths early on. It was worse early on. Perhaps it was a little better later on, a p value of 0.51.

Now, some people have interpreted this to say that triple modality therapy is better. I have

a hard time with that. I think we still all agree that survival is a pretty hard and important endpoint. And, I think that in some of these trials we might have been misled by the time to progression analyses, not always, especially if the treatment is not so toxic.

[Slide]

This is a two-drug platinum based regimen, a more modern regimen looked at in the adjuvant setting. This is disease-free survival, statistically significantly in favor of the chemotherapy. Survival looked like this. Survival was statistically better as well. In this case time to progression or disease-free interval and survival were the same but it didn't take much extra time to find out that survival was also better as well.

[Slide]

So, I still think that survival does remain as a major indicator of clinical benefit and symptom relief may also be a major indicator of patient benefit. Richard Gralla is going to talk about that.

[Slide]

So, I believe that survival should remain

as a major endpoint for clinical benefit and for approval. Richard Gralla is going to talk about this, but I believe symptom relief can be considered as an indicator of clinical benefit and also granted full approval, but Dr. Gralla is going to talk about that. In my belief, objective response can be considered as a likely endpoint of clinical benefit and, therefore, an acceptable endpoint for accelerated approval.

With the current regulations, since new drugs are likely to offer an advantage in toxicity over existing drugs, requirement for a benefit over existing therapies is not a major obstacle if response was considered as a surrogate. But in the future this could limit drug development if this requirement of being better isn't gotten rid of. I hope that you, as ODAC, might advice the FDA whether they really ought to look at that accelerated approval improvement requirement for being better than existing therapies. Right now if you granted accelerated approval based on objective response, I think since we are going to have better toxicity with the new drugs it will be okay but in the future when we get a bunch of targeted therapies if you got two targeted therapies that

are active one is not going to be less toxic than the other, and why should one be approved and not another? I don't understand that. I think drugs should be approved because they are safe and efficacious, like the law says, not efficacious and better than something else. TTP--I am not sure if it is a marker for accelerated approval at the time or not. Thank you very much.

DR. PRZEPIORKA: Thank you, Dr. Bunn. The final speaker for this session will be Dr. Richard Gralla who will talk about quality of life and patient-reported outcomes as endpoints in clinical cancer trials. Due to technical difficulties, why don't we take our break a little early. Let's be back here at 2:10. Thank you.

[Brief recess]

DR. PRZEPIORKA: Would you take your seats, please? Dr. Gralla?

Quality of Life and Patient-Reported Outcomes as Endpoints in Clinical Cancer Trials

DR. GRALLA: Thank you very much. We had an unplanned pause but it looks like we all benefited from it.

[Slide]

It is always a pleasure to share the

podium with Dr. Bunn and to be here at the FDA to discuss these interesting areas. I am going to add to the non-small cell lung cancer a little bit on mesothelioma, given that it fits all of Dr. Bunn's criteria in terms of being a difficult disease with very similar parameters.

I also want to thank the many members of the group that contributed to the presentation. Obviously, we are not all going to agree. Where you agree with me, those are my ideas. If we disagree, those are the other folks on the committee.

[Laughter]

[Slide]

This new term, patient-reported outcomes, PROs, sort of defines clinical benefit or a term that probably could have stayed as palliation for this purpose and quality of life.

For quality of life we need a multidimensional concept that includes areas less likely to be affected by chemotherapy, the spiritual, perhaps less the psychological and social but certainly the physical and functional.

For clinical benefit, with talked about the original definition. It includes areas more

likely to be affected by the treatment choice. Why isn't it just symptom benefit? Well, performance status is not a symptom is probably the reason. So, it includes functional and physical aspects as well but areas likely to be affected.

So, this is sort of the overall working of PROs--symptom palliation, quality of life of life as well, but quality of life used in a denotative way, not as a connotation of oh, it must affect his quality of life.

[Slide]

This is probably my slide that I should have entitled much like Dr. Bunn's, sort of the why are we here? Is there really a need to look at PROs? I think the answer is absolutely yes. Every physician knows that hardly a day goes by that a patient doesn't say to us, you know, doctor, I am interested in my quality of life as well, and why isn't that involved in drug approval? It should be and I think we have heard the desire for it to be.

Lung cancer mesothelioma are a highly symptom diseases. Survival response reveal only a portion of the experience that our patients and families have. Our treatments vary in their side effects and risk profiles, some of them really

being quite toxic but this applies to surgery radiation and chemotherapy. So, we have to be able to balance that experience in some way. The response rate simply won't do that. Actually, if we are honest with ourselves, meaningful survival differences are most uncommon. Every trial is designed to look at the survival differences but they are extraordinary when the occur.

[Slide]

The question came up before do we really know what symptoms to look at? You are darned right we do in lung cancer. We absolutely do, mesothelioma as well. Look at the frequency on presentation or during the time for non-small cell lung cancer and small cell lung cancer for these common symptoms that our patients present with and tell us about.

In the development of the better instruments, which I will talk about, the input of patients is absolutely crucial or we could not have been able to assemble such instruments. These were not developed by people in "ivory towers."

[Slide]

Our patients are highly symptomatic at baseline. This is a large, 30-center trial. We

looked at using a validated quality of life instrument in the beginning. As you can see, 80 percent of patients present with three or more of these symptoms, 92 percent with two or more. So, another way perhaps of doing it, to get away from some of the multiplicity issues, is to look at how patients rate their overall symptom distress, what the symptoms really mean to them. It gets back to some of the functional issues as well. Unfortunately, people at presentation first-line are extremely symptomatic.

[Slide]

Looking at survival, and this is just a compilation of large randomized trials over the past decade. The red bar represents supportive care. We no longer have the issue does chemotherapy improve survival over supportive care. Seventeen out of 17 trials with this design--way too many--showed improvement over supportive care. The majority of those trials independently showed an improvement in survival. Way too many trials were done there.

The next bar, next to the red, is just platinum alone and Dr. Cohen told us about platinum alone. But if we look at the last three bars,

carboplatin combinations, older cisplatin combinations and newer cisplatin combinations, yes, the newer drugs have a little bit of a benefit for us; they are easier for us to use in many ways and we prefer them. But in terms of survival benefit, it is very, very difficult to have a meaningful survival benefit although, God knows, we don't want to talk about what a meaningful survival benefit might mean. We have already sort of addressed that one. But it is pretty hard to have survival benefit that gets our attention.

[Slide]

Dr. Janet Dancy really put together a lot of this and I think she is just right. Here PROs can create an accurate picture of the disease. Without this we are missing what are patients tell us about in every single patient encounter. We must have this to really understand about the disease.

The second paragraph--unfortunately, many studies have shown us that we are not so good as nurses as doctors in predicting how our patients feel about these things. It is too bad but, unfortunately, has been reproduced even in the JNCI and in the Miles trial was shown once again.

Interestingly, why we need this is that response rates under-estimate the benefit. It appears we don't need a major response to be able to have enough change to be able to have benefit.

Finally, how do we have this balance between symptom improvement, toxicity, the difficulties of treatment and the benefits? There are many examples where more toxic regimens are associated with greater patient benefits, including their symptom relief, etc. So, to be able to put this together is not easy--actually, it is easy, we have to ask the patients and they can tell us.

[Slide]

So, the four questions I have always had with these areas are can we define quality of life? We surely can define pain, dyspnea and cough.

Can we measure quality of life? That is what a lot of the conversation was about. Can we quantify the more subjective aspects? We quantify subjective aspects all the time in many different areas in behavioral science.

Can we agree on how to analyze the data? I am not sure we are quite there yet but I think we are getting closer. We have a lot of good people around the table who can help us with that.

Can we present the data in a way that is clear and useful, not looking at 99 different endpoints, etc.? That is nuts!

[Slide]

Define it. If we ask each one of us in the room to define quality of life in one, two or three sentences we will probably end up with some disagreement. If we sat here for a while we would probably come pretty close and be able to carve out one paragraph. One thing we can agree on is this is probably made of these dimensions, the physical such as symptoms and side effects; the functional which we talked about earlier, psychological, social and spiritual. Spiritual doesn't have to mean religious; it can be meaning of life. So, these are the denotation areas of quality of life. Now, the other PROs, the patient-reported outcomes, deal more with the physical and functional.

[Slide]

This is the model part of the content or actually the construct validity for quality of life. Dr. Patricia Hollen publishes for the LCSS instrument. Well, if we look at the physical dimension and the functional, those are what are, for the most part, discovered or looked at in the

other PRO dimensions, the symptoms, the performance status. Yes, we can look at functional dimensions. The FACT-L actually does a very nice job of looking at the differences in function and how function is meaningful, and we don't have to look at these as a lot of different endpoints. So, we can focus on the physical and functional, which account for about 75 percent of the variance in many of the studies, and globally capture quality of life in the others.

[Slide]

Instrument development has changed, or instrument use has changed in quality of life. We have instruments that are good for all populations that are kind of interesting to look at, but I think it is clear that there would be a need for instruments that are more cancer specific than, say, osteoarthritis. The pace of these diseases can be quite different.

We talked a little bit about lymphoma. The B symptoms of Hodgkin's disease are a great deal different than the symptoms of lung cancer. Issues such as fertility are issues that we think about all the time in younger patients with lymphoma but it is not really such an issue in lung

cancer. So, we need disease specific instruments. We might even need treatment specific. We talked earlier today about adjuvant trials. In adjuvant trials in lung cancer in patients with stage I and II, we want to look a year later to see if our interventions in an adjuvant trial in somebody who has undergone a right pneumonectomy whether we have good quality of life a year later. That may be a different instrument that refocuses on the functional endpoints than we would use in a clinical trial in stage IV where that patient has an expected 7-, 8-, 9-month live altogether and we have such instruments as well.

[Slide]

Here are the three instruments with acceptable psychometrics. We will look at the psychometrics in a second, the LCSS, EORTC QLQ30 and the FACT-L. The latter two, the EORTC and the FACT-L are similar. They are 30-40 items total, a general module 7-13 for the lung cancer. The LCSS was developed specifically for clinical trials and clinical management. It is shorter; 8 items in mesothelioma, 9 in lung cancer and 6 observer items but the observer scale is optional. They take between 3 to 10, 12, 15 minutes. These are not the

99-item instruments that are out there, and more. We do not need those.

[Slide]

What kind of validation have they been through? They have been through very serious validation methods. These validation methods were not set up for cancer; they were set up for behavioral science and they are very strict and are much more difficult than, say, RECIST or most of the other things that we have been talking about. We can see that these instruments to be useful must be valid, reliable and feasible, able to be used in a real clinical practice in real time studies.

Here are some of the psychometrics that are there. As far as the content validity, the content of what we looked at if we didn't have patient agreement, patient input, it wouldn't be worthwhile. Fortunately, that is true in all these instruments.

[Slide]

If we look at internal consistency, if we look at the reliability, stability--do you get the same results if you give it again to the same patient? Do you get it if you give it in different groups of patients who have the same

characteristics? The answer is yes. Dr. Nunnally wrote the textbook in this area, not as far as oncology is concerned, and the instruments that I showed you, those three instruments stand up very, very well.

[Slide]

If we look at two of the lung cancer instruments, for instance, that are used the most in U.S. trials which is why I looked at them, if we look at their reliability coefficients, the Cronbach's alpha for their core measures, they come out very, very well, and much better than needed for a new measure. For the lung cancer module they come out really quite well also. In fact, we have a new publication from Dr. Chris Earl and Jane Weeks that looked at quality of life and PRO instruments in oncology and the lung cancer instruments, specifically the LCSS, are among the very best in all of oncology. So, as far as lung cancer is concerned, we are blessed by having some really pretty good instruments and most of these instruments now are being put into electronic format so that they can be very, very easily done with very little extra time for patients or data managers.

[Slide]

If we look at other types of validity construct criterion related, they are really there. They compare well to gold standards and other aspects. So, there is no doubt that the validity process that has been used for these types of measures in a variety of different conditions are met by these validated instruments, not necessarily by other instruments.

[Slide]

We talk about this clinical meaningful difference. I am just floored why it is that this should be answered for these PRO endpoints and quality of life but not for survival. I really am amazed that we can even talk about non-inferiority if we can't set what the border is for survival that would be important. I think that this really becomes rather difficult. We know it doesn't meet non-inferiority but what was the border? Why was that boundary selected? The same thing is true here.

I like what Dr. Williams said, we look at whether there is a statistically significant difference, whether we can be confident that there is a difference. Let's apply whatever we are

applying to these PRO or quality of life endpoints too. Either we have a difference or we don't. It is for somebody to look at and say that three percent difference doesn't mean much to me. We heard the five-week difference didn't mean much. But, of course, Dr. Cohen presented a lot of five-week differences here that we have approved drugs on, and there is value to normative data being collected as well.

[Slide]

Phase II trials, single-arm, non-randomized trials, these trials suffer from the same problems that survival studies do. We talked about the gefitinib trial before. We were all glad to see that patients had a rapidly occurring change. Of course, that was really looked at from the subscale FACT-L, not necessarily the whole FACT-L and, yes, there was symptom improvement and these are all very nice things to see. But the problem with these is, just as with survival analysis, that with the lack of a control group we don't have a context.

[Slide]

What makes it particularly difficult in symptom control is that we are giving standard

palliation. It is not a blinding issue. Of course, we are giving pain medicines to people who have pain; cough medicines to people who have cough and oxygen to people who are dyspneic. We wouldn't want to do a trial that was any other way. These are confounding problems but they are what we deal with in clinical medicine every day. So, without having something for context I have no idea whether or not that is a great response rate we see or not. So, in Phase II these are helpful in hypothesis generating but difficult for us to say that they lead to true improvement.

This can lead to an overestimate of benefit. On the other hand, if we just looked at the response rates, since less than a major response gives benefit, that has been an underestimate of benefit. So, there are problems with Phase II. It is probably really good to analyze these data in Phase II studies so it can be more useful in trying to guess what difference we need to look at in Phase III.

[Slide]

What about Phase III trials? What kind of problems do we run into there in comparison trials? Well, these are the complaints that we hear the

most, cumbersome instruments. Yes, but actually the three instruments I showed you are not so cumbersome, the 3-, 5-, 15-minute analysis isn't so bad. It takes a whole lot less time than the MRI that we get all the time or the PET scan or the CT scan. People say how can you ask a sick person to complete this questionnaire that might take them five minutes, you mean as opposed to getting into an MRI machine? It is really very easy. It is tough to get the sick patient who may have progressed over to the PET scanner but it is not so hard to do these instrument and many of these can be done by phone.

Patient deterioration is a big problem and this can lead to the sloppy data that we heard about before or asymmetrical follow-up--nice term; I like that term. If we don't follow-up equally in two groups in a Phase III, that is not good. So, we need to be looking at patients even after they progress. Lack of investigator commitment. How do we prevent that? We emphasize it from the very beginning.

[Slide]

This looks at those same 673 patients that I showed you before with those symptoms. We wanted

to see after three cycles how many were staying on study, 64 percent. The main reason for coming off and not having assessment was disease progression. This is completely controllable simply by following with something as simple as an instrument that costs pennies, not thousands of dollars, to be able to follow this.

Another advantage of following the PROs is we talked about the problem of contamination with crossover. This isn't crossover. We don't have to worry about that. It is eliminated from looking at this. So, we should be able to improve this follow-up by at least 20 percent to be able to get 80-90 percent adherence rather than the 64 which is certainly not good.

[Slide]

Who drops out? Who is in the attrition group? Well, we looked at age which is not a prognostic factor in lung cancer and there was no difference between the on-study group and the attrition group by age. Indeed, if the symptom burden was worse or if the quality of life was lower, those patients were disproportionately seen in the attrition group. Think what that does. That takes an arm that is inferior in terms of

response or survival and it drops out the more symptom or lower quality of life patients, artificially making the inferior arm look better. So, that is a real problem. Is it surmountable? Easily and it has been surmounted.

[Slide]

This is from mesothelioma study. I will talk a little bit more about it. Nick Vogelzang published this study in the JCO this summer. It is pemetrexed-CIS versus CIS in advanced mesothelioma. What did then do? They conducted a brief training session so that everybody involved understood why quality of life and PROs were being done. They included baseline quality of life data as part of the randomization which emphasized the importance that we really want this as much as we want the CT scans. They continued to have emphasis while monitoring the trial and, as a result, more than 90 percent of the planned assessments--this was done weekly which I think is excessive and there are reasons to believe it is excessive, but more than 90 percent of the planned assessments were done. So, this is probably the industrial standard.

[Slide]

We talk about survival, quality of life

and response as being separate. We need to analyze them separately, that is correct but, of course, they are more related than different. They are related because they are largely determined by the malignancy. If we cannot control the cancer we will not be able to improve survival very likely or quality of life. Of course, if the treatment is harsh then this could have a negative impact on survival or quality of life or both.

But when we look at the approved regimens that Dr. Cohen showed us, they are all pretty similar in terms of their toxicities. There are not big differences. So, we shouldn't expect with modern care that that is the problem. So, they are inter-related but they are not identical, these endpoints, and quality of life is a very important one. But I don't think we should ever look at quality of life without looking at survival or looking at survival without looking at quality of life, but either one of these could be a primary endpoint.

I like what Dr. Bunn had to say about response and accelerated approval but when we talk about large trials response is probably not of great value if it doesn't contribute to quality of

life or if it doesn't contribute to survival, and probably any good treatment will contribute to both because it is mediated through the malignancy.

[Slide]

This looks at the survival based on quality of life at baseline. If we look at that group that scored their quality of life in the lower half of the group, they had a much inferior ultimate survival when compared with the group that scored their quality of life in the top half of the group. That is not too surprising but this was a more important prognostic factor in multivariate analysis than any other, including stage III versus IV, including gender, including performance status. So, ignoring quality of life is missing the boat on a lot of these areas. Yes, it is more difficult to measure quality of life than to use the instrument that we use for survival, that instrument being a calendar, but I should think we are little bit more sophisticated than just having the ability to use a calendar.

[Slide]

For Phase III we have problems in analysis. The standards for statistical approaches remain controversial. I do agree that the less

modeling we can use, the more data that we can include, the better off we are. There are problems with simply averaging scores. Survival differences complicate quality of life analysis because the attrition is not random. But these are correctable.

As Dr. Fleming has emphasized, results from all patients on trial need to be analyzed. Instead for looking for a way to adapt for that, we need to follow all the patients. They did that in the mesothelioma trial and we can do that too.

[Slide]

Well, does it really add to response or to survival, the common endpoints? Let's just look at these data. This is almost a 500-patient study. If we look at this in terms of the PRO outcome of pain, which is something Dr. Carpenter brought up as something important, it is not too surprising to us that patients rated their pain control as better if they had either a CR or PR, but we know there are not real CRs--a major response versus stable disease versus progression disease.

But what we didn't expect to see is if you just look within response, because we think of response as a blunt instrument and you either have

a response or you don't, if we looked at how patients rated their pain there was a major difference between the pain control for those who got the combination regimen, in this case pemetrexed-CIS versus the single agent. You can see the yellow bar versus the blue bar. These patients were all followed to the same degree. They all responded but there was a change in pain. In fact, in all 8 LCSS parameters the same pattern existed within responders and patients on the combination rated their patient-reported outcome, including quality of life, as being better. So, it is possible that this is a more sensitive measure than the blunt instrument of response.

[Slide]

What about survival? Well, Dr. Vogelzang reported in the JCO that there was a survival difference between the combination regimen and cisplatin alone. If you look at 12 weeks there was no sign of this. At 18 weeks there was only a slight suggestion that there might be a survival difference.

But let's look at quality of life and symptom distress--this covers all the PRO aspects. If we look at quality of life we can see that there

was already some difference at week 12 and a larger difference at week 18. When patients rated distress from their symptoms the same pattern was seen. At 12 weeks this was not significant. At 18 weeks this was highly significant, even if one addresses the issues of multiplicity, showing that it was easier to show quality of life differences and symptom distress as the patients reported which was significant earlier on than was survival. In fact, this is predictive validity, predicting what will happen to survival which is considered to be a very strong validity point.

[Slide]

My conclusions would be, and our group said, yes, this is ready for "prime time." There are validated instruments but when we do these studies we must select carefully. We need to use a validated instrument but, remember, some of these instruments measure different aspects, such as a clinical trial versus an adjuvant trial, a little bit different and we need to be sure that we have the right languages and cultural aspects which many of these instruments address.

As with other study endpoints, before the trial begins we need to delineate what are the

primary endpoints. We need to address areas of multiplicity and of analysis. Too often I see protocols that say, well, here is the instrument we are going to use and we are going to analyze it and then later comes the analysis. No, that has to be thought out ahead of time. If so, we will have something that we can present to our colleagues at FDA that I think they can probably get their arms around.

We need to follow all patients whether they are progressing or not. That is one of our biggest areas of problems so we need to follow all patients throughout a predetermined interval. So, if we have an interval to follow the patient, how long should that interval be? Appropriate to be able to see response and appropriate to be able to see the toxicities. If we can see that, we can see that area.

There are other uses for quality of life. In terminal care we can look at it in those areas but that is a different issue. But in the beginning in a clinical trial, follow for a specified time but follow all patients. When patients die, is that a problem? It is not a problem. Quality of life is a function of life.

If some patients have died, that is what occurs; we don't follow those. But we don't look for the patient who is no longer contributing, the patient lost to follow-up. That is as bad as with toxicity and response.

[Slide]

We need to use an appropriate control group. Sometimes this is difficult. And, all these comments refer to quality of life measures when we are looking at drugs that are likely to have their benefit by means of anti-cancer activity. We are not talking about pain medicines here. We are talking about anti-cancer drugs and looking at approval for those. Their appropriate control group is important.

We need to emphasize compliance throughout the study and as long as the investigators and the patients understand this, then I think we are likely to have people included. When it is feasible to blind the patients and the doctors, especially the staff, that is great but it is not always possible to do that and I am not sure that is the biggest objection.

[Slide]

So, can we define quality of life

adequately? Can we measure quality of life? I think we have some decent instruments. They are not perfect but they are decent. When they are put in electronic media they take almost no time from the staff, almost no time from the patients. Can we agree on how to analyze quality of life results? We are getting closer. There are thoughtful ways that we can talk about. Can we present quality of life findings clearly? Sure, we can. We don't have to present every last aspect, especially when we have determined at the beginning of a trial which are the primary endpoints that we wish to look at. Thanks.

Clarification Questions to the Presenters

DR. PRZEPIORKA: Thank you. Before we have our introduction to the questions I would like to actually ask the three speakers to take the podium together and have the committee have the opportunity to ask them questions. While the synapses are all firing up here, I will take the prerogative to ask the first question.

Dr. Gralla, you went through what validation means or quality of life which, in the lab, would qualify as qualification rather than validation which would be predictive of an outcome.

You did mention "the gold standard" but did not identify it. What do you use as the gold standard? For example, if we had a surrogate as a response rate we would hope that would predict for survival. What do the quality of life instruments measure for?

DR. GRALLA: For instance, predictive validity from an instrument, and this could be true for time to progression or whatever and for quality of life, predicts for another validated endpoint. But when you do against gold standards, if we looked at instruments such as the American Thoracic Society dyspnea scale, if we looked at the Melzack-McGill pain scale, etc. we now have huge numbers of questions to ask. So, what we look for are correlations between using these already validated instruments. So, for pain the Melzack-McGill scale is one that one could select, there is a whole variety of different scales that are out there for different aspects that are used for use as gold standards.

This is why if you read the papers, and each one of these three instruments have published psychometrics, they tell you exactly which scales they used, the PONS, etc. to look at various

aspects. It takes years to validate these scales which is why we don't want to see somebody just ad hoc make up a scale to be used in the next myeloma, lymphoma, lung cancer trial. So, there are specific scales that are found in each of the publications.

DR. PRZEPIORKA: Dr. Levine?

DR. LEVINE: I have kind of a crazy question but people are all different. I saw this on one of your slides but, you know, one person may call something pain and that is not pain at all to somebody else.

DR. GRALLA: Right.

DR. LEVINE: So, is it valid to just look at what I say is my quality or maybe what you should be looking at is change, you know the delta, in each given patient. How do you analyze that?

DR. GRALLA: You brought up a very good point. For many of these instruments, that is what the Cronbach's alpha, the internal consistency, can look at. When you look at certain items that don't make sense--for instance, the fatigue question, 15 years ago when we looked at that we said we don't think people understand what fatigue is. So, we will look at tiredness; we will look at weakness.

Well, they all meant different things to different people. It turned out that the right term to use, years later with much more testing, was fatigue--

[Laughter]

--and only by testing could you find that out. So, you must find that out. In emesis scales, which is different, nausea means something rather different. Don't ask my mother-in-law what nausea means to her. It is entirely different from what it means to others. And, that is a real problem. But for each of these instruments those points are there.

Now, do you ask about change over time? You must have a time period. For instance, if you ask a patient how did you feel nine weeks ago it is really difficult for us to say. So, for many of these instruments the time frame is in the past day or in the past week.

DR. LEVINE: I didn't mean that. I meant let's say the instrument is done at baseline and then every week. I guess it is an analysis question, couldn't you just look at changes between week 1, week 2 and week 3 and that they have answered in a timely way?

DR. GRALLA: Indeed, that is the way that

many analyses are looked at.

DR. PRZEPIORKA: Dr. Bonomi?

DR. BONOMI: Along the same lines to Dr. Levine's question, maybe we could define a quality of life response just relating to the physical elements, not the whole quality of life instrument, and the point that you made, a baseline and, say, four weeks and eight weeks. What is the statistically significant change? I know in gefitinib they talked about a difference of two points. I don't know the statistics of it but it sounds like an awfully small change to be considered significant. It seems like we need to look at that. Could we define some type of quality of life response that could be then applied across studies?

DR. GRALLA: Phil, I think that Dave Cella meant 2 points out of his 7 questions, and of 29 total points yielding a 7 percent difference. We can either accept that or not as such. It is kind of the same discussion that we have had before. Think of the risk-benefit aspect there. If you were looking at imatinib versus marrow transplant in CML, clearly you would have to have a better benefit in the marrow transplant to be able to be

worthy to most people than, say, just giving Tylenol or just giving imatinib. So, the risk-benefit probably comes in there and it is just the discussion that we talked about before, in rapidly progressive disease, highly symptomatic.

One of the problems is when the baseline is 70 percent where 100 is perfect and 0 is terrible and you improve by just 6 or 7 percent, that doesn't sound like very much but actually it is 25 percent of the amount that you could improve. So, it is the relative difference versus the absolute. These are very, very difficult things to answer. In a progressive disease like lung cancer is it the number of patients who report an improved quality of life, a stable quality of life, or is it when treatment A preserves more quality of life over that entire group versus treatment B even though there is a deterioration in both groups? I favor the latter rather than looking at the quality of life response.

DR. PRZEPIORKA: Ms. Ross?

MS. ROSS: Thank you. I guess this would be to Dr. Bunn and Dr. Cohen. Dr. Bunn made the statement that only in oncology drugs is accelerated approval dependent on showing an

advantage over existing drugs. Was that your statement, Dr. Bunn?

DR. BUNN: Right.

MS. ROSS: I heard someone say that is not true.

DR. TEMPLE: The accelerated approval rule refers to showing an advantage over available therapy. That is why you would accept a lesser standard of approval.

MS. ROSS: Is that only on oncology?

DR. TEMPLE: Oh, no, it is for everything, for any accelerated approval.

MS. ROSS: Has that ever been changed? Is it a rule?

DR. TEMPLE: It is a rule; it is a regulation.

MS. ROSS: It is a regulation?

DR. TEMPLE: Yes.

MS. ROSS: Or is it law?

DR. TEMPLE: It is actually now in law as well. It is part of the fast-track provision of FDAMA as well as the rule.

MS. ROSS: Thank you.

DR. BUNN: As I mentioned, right now that is probably not a huge problem for oncology because

many of the new drugs have less toxicity so they do have an advantage over existing drugs in terms of toxicity. I brought that up in terms of thinking about the future. You know, laws and rules are made to be changed so perhaps in the future one would consider whether that provision for accelerated approval is a bit too strict. Certainly for regular approval that provision doesn't exist, only for accelerated approval. Is that right, Bob?

DR. TEMPLE: Yes. There is one thing that is important. The Commissioner has announced this. We were trying to decide among ourselves whether this has made it into a rule but you can or will be able to have a second accelerated approval, say, for another drug that is not cytotoxic as long as it still has an advantage over anything that has full approval. I don't think that completely--

DR. BUNN: It is halfway there.

DR. TEMPLE: I don't think it goes completely to where you want to go but that is important.

DR. WILLIAMS: Dr. Bunn, as I read it, there is no reason to have accelerated approval. You know, according to your proposal you could use

a different endpoint then that would be tantamount to full approval and there wouldn't be any particularly setting where you needed it. It would be in every setting. You would get approval in every setting for the surrogate endpoint. Right? That is what you are proposing? There is no particular setting--

DR. BUNN: No, no, if you had a response rate in an untreated population of 25 percent and you had the same toxicity profile, then you wouldn't be able to get accelerated approval. If you had a response rate of 25 percent and you had less toxicity, then you could get accelerated approval.

DR. TEMPLE: If it was still accelerated approval now and it was based on response rate alone and there was no other drug and the second, third, fourth still had an advantage over available therapy, they could still be approved. I think you really want to say if it is a useful drug none of that should matter and you would like to make that a standard for all cases, but we haven't done that--

DR. BUNN: Right.

DR. TEMPLE: --but accelerated approval is

not terminated by the approval of one drug under the accelerated approval rule.

DR. PRZEPIORKA: Dr. Cheson?

DR. CHESON: Paul, response rate in lung cancers to you is an important endpoint. Does it matter how long the responses last?

DR. BUNN: Of course, it does but--

DR. CHESON: Is there a minimum duration of time which you would accept for that?

DR. BUNN: We don't know that. That hasn't actually been looked at and it is something that probably could and should be looked at. But, surprisingly, there is very little variation in duration of response. They are very similar. I don't know why it is. You know, why is 20 percent, more or less, sort of the magic threshold for what will lead to an improved survival. It is hard to say. Almost all those drugs have a median duration of response in terms of three months. If you had one that had a median duration of response of a year it might make a bigger impact on survival. If you had one that only had a median duration response of a month would it still affect survival? I don't know and that is because we don't have any examples. So, it is something that we should

certainly look at but there is not a lot of data and there is not much we can say about it at the moment. Do any of the experts over here disagree with that? I mean, I think at the moment it would be hard to put median duration of response into the equation.

DR. PRZEPIORKA: Dr. Fleming?

DR. FLEMING: Actually, I have questions for both Richard and Paul but to avoid confusion let me just start with Richard.

DR. GRALLA: I was afraid of that, Tom!

DR. FLEMING: Actually, I was pleased to see that you addressed a number of the issues with PROs that we struggle with, issues of how imperative it is to ensure you are following everybody so you are getting an unbiased assessment. I still struggle a little bit with how to handle the deaths in that regard.

With the validity issue, you talked a lot about that. Blinding still troubles me as to how we could address that. I think blinding is really key to the objectivity of measuring these.

A question that I would like to ask or a comment maybe in response to one of your questions, you had pointed out this committee, in a sense,

dodged the question of how much of a survival effect you need to see for it to be relevant and you were saying why should we be asking the same thing for PROs. At least for some of us the reason that there is a difference there comes down to a multiplicity issue with PROs. There usually is a wide array, as you have mentioned, with these various scales, 6-plus 9 or 15 measures, 30-40 measures etc. It really is important to formalize this into something that is a primary endpoint. Sometimes that may be based on a composite. What you get then is you compromise interpretability for enhanced sensitivity and here is the issue, you might now have exquisite sensitivity to small differences in these composite measures and then it is, in fact, much more likely that you could achieve statistical significance there and wonder if it is clinically significant. It is much less to occur on survival, for all the reasons we have heard--it is difficult to get an even adequately powered survival study. So, I would say there is a reason. I don't know if you wanted to comment specifically on the issue of multiplicity on this.

DR. GRALLA: I agree with you entirely, Tom, it is a real issue and that is why you need to

define it in the beginning. First, it is simply something as simple as looking at quality of life which can be looked at globally, or looking at symptom distress or looking at pain, whatever you feel would be most important in this population. You don't need to look at all of them. The problem we have had the most is with people looking afterwards and then choosing, oh, here is the one that came out, or overwhelming us standard data in a 99 instrument and 44 looked at this and 33 didn't. That is over. That time is over. Those aren't the issues.

When we use these instruments we can look at families and maybe we do give away some sensitivity but, in fact, in looking at some of the data that I was pleased to see with some of the trials that I mentioned, we in fact don't have a multiplicity issue. When we look at two or three of these areas, even if we adjust for the fact that we are looking at three endpoints, it is still significant.

I know that that gets back to your other point of looking at small differences in survival. Again, we are talking lung cancer. Marty showed us approvals with five-week, three-week survival. So,

I don't think that we should be rushing to worry about those small differences. I can't understand why a patient would say to me, well, let's see, doctor, there was only a 7, 9, 10, 12, you-name-it, percent difference, why wouldn't I want the one that had that 12 percent difference? And we look at what patients want and whether we are fulfilling those needs.

The blinding, it is great to do when you can and often it can be done and should be done but, you know, when you think about it, you have a large trial and you are looking at pain control and you give the patient the pain visual analog scale. The patient I think is pretty honest about telling you what it is and as an investigator in a 400-patient trial I have no clue as to how that affects. In other words, I am not putting my input in, the patient is. I am not sure the patient understands which one is better in that regard.

Where it is also important though is the context. Did it require more pain medicine to be able to get that pain control result? So, we do need to look at that. Anyway, that is sort of how I would address some of those key issues that you bring up, Tom. They are important but they need to

be thought of, just as survival, ahead of time; just as whether we are going to look at disease-free survival, TTP, TTF and survival. I think they are similar issues.

DR. FLEMING: I think it is when we use the composite scales that are harder to interpret and then we can see very small differences. Yes, I would say a small difference is better than no difference if I can get it for free but then it is benefit to risk.

Let me get to a question that is probably more for Paul although it relates a little bit to what you were talking about as well, Richard. Paul, one of the take-home messages I get from what you are saying is you are identifying concerns with launching large-scale Phase III trials because we have to show survival effects when there really isn't adequate evidence at hand at baseline to say the plausibility of achieving that positive effect on survival is adequately high. Gee, if we had responses and we were looking at 15, 20 percent responses, then your sense from the data you are looking at is that it is much more likely that we will see a survival effect. I guess one take-home message I get from what you are saying is then we

ought to have fewer study settings jumping from Phase I to Phase III. Let's do that Phase II trial with 100 people and see if we get a 15 or 20 percent response rate.

The issue that is troublesome here, and is a little bit related to what Bruce's comment was before, as I look at response it seems to me that response is a component of what we would think of as an integral causal pathway through which the oncology disease process is influencing outcome like survival. My worry is that when we look at percent of patients that achieve a certain level of tumor shrinkage would dichotomize the world and that dichotomization may be missing part of what the intervention and disease process is really doing here. It is not just a matter of did you achieve a response. What was the magnitude of that response? What was the durability of that response? It is easy to envision that an intervention could readily be achieving intended benefit on clinical endpoints like survival and an oversimplification of what is really happening to the disease process, to the tumor burden may not be adequately captured by percent responding.

One of the things that troubles me too,

and you and I had a brief chance to talk about this, when you look at that meta-analysis of the 176 Phase II trials, those studies are looking at the relationship between whether somebody responds and what the overall survival is.

So, Richard and Paul, you are vigorous and I am frail at time zero. In fact, Richard, you have a better quality of life than I do and, Paul, you have a better response than I achieve and both of you survive longer. What do we see from those data? That there is an obvious correlation between quality of life and survival and response and survival. Now, Richard, I don't care that that is the case in what you are advocating because quality of life is a value to me whether or not it is a surrogate for survival. But with response, Paul, I do care because I do want to know that this is, in fact, giving me evidence that mediated through that response I am causally inducing what I really care about.

Here is the rub, we could have a million patients in the data set that you have been providing to us. What it does is it tells us about a correlation that exists but it could be that the causal mechanism for that correlation is not

induced responses leading to prolonged survival. What I need for that, and this is critical information, is properly controlled trials that can compare what is the treatment induced influence on response versus the treatment induced influence on survival. That relationship across a meta-analysis is telling me whether or not I am causally influencing survival mediated through response.

DR. BUNN: I don't really disagree with what you say. One of the issues gets down I suppose to semantics but, you know, it has to do with cytotoxic versus cytostatic. If a lot of the drugs that we have actually worked by being cytostatic this would be a huge problem. Maybe bevicuzimab will be the first but maybe some day we will get confounded by cytostatic. But most of the drugs that improve survival and, in fact, in my belief all of them at the moment, have actually worked because they are killing cancer cells. Even tamoxifen causes objective responses in patients and certainly Iressa causes objective responses.

So, I think when the mechanism is to kill cancer cells, that objective response actually makes sense. Sometimes, you know, examples are useful. I think it is not out of school to be

actually thinking about what is coming along. You heard about a trial that looked at a non-inferiority survival advantage in second-line non-small cell as the major endpoint. In every efficacy parameter, including symptoms, both pemetrexed and docetaxel were identical. It is the biggest trial ever done in second-line non-small cell. But the non-inferiority p value was 0.051. I don't know what the committee will do but I do know that the response rate to pemetrexed was 9.1 and to docetaxel it was 8.8 and the symptoms were just as often relieved.

So, if the committee can't deal with a single trial with a p value of 0.05 in terms of non-inferiority, accelerated approval could be given on the basis of response for, you know, a drug that I think needs to see the light of day in this disease and killing some of these drugs may be the end of the light of day. Erlotinib is going to come before this committee in a trial where the hazard rate for the study was a hazard rate of over 30 percent reduction for a single pill in second- or third-line non-small cell that is a big change and that may not make it against best supportive care in terms of survival but I will eat my hat if

in terms of response it is not highly statistically significant and if it isn't eight percent or higher.

DR. FLEMING: But your example is a bit changing the topic here because you gave an example where you were talking about evidence on response and time to progression and survival, and you are really asking the question, in a non-inferiority setting, what is an adequate amount of evidence on the aggregate of those measures, which is different from the thrust of your presentation which was let's reexamine whether or not there is adequate evidence that if you can induce an impressive response rate at a certain level that is now adequately reliable evidence for benefit.

DR. BUNN: Right, if erlotinib has nine percent and best supportive care has two percent I would say accelerated approval should be given.

DR. PRZEPIORKA: Dr Cheson?

DR. CHESON: Paul, coming back to part of your elegant presentation, there are some drugs which you had on your list that never should have gone on to Phase III because they are inactive as single agents. I take issue with that because there are some drugs, particularly one of them that

you had on your list, which are probably not active as single drugs but work better by enhancing the activity of other agents. What I am thinking of is Gentasense, for example. So, I would be reluctant to throw out some drugs like that have a unique mechanism of action. Some of the growth factor receptors may be the same sort of thing. The typical cytotoxics, okay, but when you get to the new targeted therapies I think a lot of them may work better and should be studied going right from Phase I to Phase III if there is in vitro rationale for such combinations.

DR. BUNN: I am sorry I don't have my slide to put up but the bottom sentence on that was unless there is very good compelling preclinical evidence for why that would happen. So, that is not uncommon to the situation up until now but I certainly don't disagree with your sentiments but I think there should be compelling preclinical reasons for that. Again, you know, bevicuzimab may be the first one to actually prove me wrong but I will be happy to be wrong.

DR. PRZEPIORKA: Dr. George?

DR. GEORGE: Richard, I have a couple of things. One is that you make very compelling

arguments of why we should be able to these kind of studies in quality of life. One of the frustrating things to me, sitting on this committee, is we don't see these things. We don't see good, well done studies in this area and I was wondering if you have any notions, accepting what you have said, that we are not seeing them because they certainly could add a lot to a lot of these kinds of applications.

DR. GRALLA: Steve, I agree with you 100 percent. The problem is in the past we really haven't seen so many good ones. In fact, over the last five years what we have seen is sort of leapfrogging. Each trial gets a little bit better than the last at doing these. We see more trials that start to use validated instruments. We have even heard of some ad hoc instruments. I think now with the electronic way of keeping the data we are there on some of these. So, I think that we are now poised for you to be seeing more of these.

The second line in small cell approximated some of these, approximated one of the validated instruments. It wasn't really an elegant presentation for looking at the topotecan second-line but it was getting there. So, I think

why we are here is to encourage that and to try to set some points along the road to help those who are doing these studies to be able to present trials in that way to this group so that you are more able to evaluate these results.

We have had some presentations at ASCO this past year that looked in that way, and maybe the year before. So, I think that is what we are going to be seeing in the future.

DR. GEORGE: This just seems to be an area where theory and practice seem to be far apart.

DR. GRALLA: You have a very good point but I think we are getting much, much closer now and I think you will see them soon.

DR. GEORGE: One quick question, just a small point, on this blinded evaluation, blinded to the interventions, there are other types of blinding that can be equally important in this area. I guess we saw some of that before. For example, just knowing sort of the clinical development of things could presumably influence quality of life. That is, you have to know when you are asking these questions if the patient was just told that they had, say, a response--

DR. GRALLA: Right.

DR. GEORGE: --Mrs. Jones, your tumor is shrinking. Now, would you please as this question, how do you feel?

DR. GRALLA: Right. That is why all of these instruments believe your point and have taken it for granted. It is not just a response; how about your white count? Your white count is 1.9. We are not going to treat you today. Oh, my God, I am going to die. So, for almost all of these instruments is when you repeat the measure. You do it before the patient sees the doctor and before the patient gets any clinical results. You are 100 percent correct. That must be done or you could have wonderful impact on the study through more subtle means. So, those areas have been addressed.

DR. BUNN: I would like to make just one comment. I think, you know, we are getting better. The FDA actually has said for a long time that symptom benefit could be for a primary approval but sometimes the studies have been so bad that that hasn't happened. I will just give you that same example again where there are going to be three endpoints. There is going to be survival, and in my opinion the study is a bit under-powered because it is looking for a big survival advantage, but

there is symptom benefit. This is erlotinib versus best supportive care. I believe full approval should be granted if there is a tend in survival and there is symptom benefit that is statistically significant if you believe it was done well. If you don't believe it was done well and there is a statistically significant difference in response and the response is eight percent of higher, then I believe accelerated approval should be given based on response. So, I mean, you have three endpoints and you need to decide what to do.

DR. PRZEPIORKA: We are approaching the scheduled time for the open public hearing but I don't want to squash questions. I see a few more hands back there. Dr. Bonomi?

DR. BONOMI: I have a question for Tom. I think there is no question that response is at least a treatment-related diagnostic factor but, you know, the cause and effect thing--we have been talking about it for 25 years and we used to plot out the curves, the PRs and the stable disease and we can't do that because maybe the people who were better, who were going to live longer also exhibit a biologic response. But with all the data we have and all the cooperative group studies, is there

some type of statistical modeling that could be done to try to elucidate this? You know, my gut feeling is response does translate into some benefit for the patient but how can we go at this?

DR. FLEMING: Absolutely, there is and you are exactly right to say that it has been 25 years since we have recognized this issue that, you know, responders live longer than non-responders but that is not evidence that I have a treatment-induced effect on survival mediated through response because, as you say, people who are intrinsically better may be the people who would have survived longer and would be more likely to respond and treatment has just labeled those people who were better.

It is, however, the first step. If I have a marker that I am going to use as a potential replacement endpoint the first thing I need to know is, is it correlated. So, it is not a useless step. By the way, if it is correlated then, in that sense, it can be useful in other ways. PSA can be correlated with prognosis and it could be a very good measure to counsel patients or to detect disease but that doesn't mean that it is a good measure to indicate treatment effect. What we have

to know for that is that the disease influence on the clinical endpoint is predominantly captured by this marker, that this marker is in that pathway mediated through which these benefits occur. And, we have to have some sense that it is unlikely, and this is tough, that there aren't unintended mechanisms that can influence outcome not captured by the marker.

Those are clinical insights that are important to supplement the data. The data, as you point out, can also though be very helpful and it needs to be analyzed in a much more sensitive way. It is only the first step to see that people who respond live longer than non-responders, have a better quality of life, blah, blah, blah. What I really want to know is if you have 20, 30 or 50 or 100 studies that have been done, and these need to be randomized, controlled trials, and those studies have measured treatment-induced effect on the marker--let's say it is response, let's say it is time to progression, and treatment-induced effect on the clinical endpoint, what we need to understand is what is the functional relationship between the level of treatment-induced effect on that marker, such as response, and the level of

treatment-induced effect on the clinical endpoint, which is other than what that meta-analysis of 176 studies did. It is a different issue. An example of this is the analysis that was presented on November 12, looking at whether disease-free survival--this as Dan Sergeant's analysis--could be a surrogate endpoint for survival in the colon adjuvant setting. They at least did a meta-analysis on all potentiated 5-FU colon adjuvant trials and showed a fairly strong relationship between the magnitude of treatment effect on, in that case, disease-free survival and the magnitude of treatment effect on survival.

So, the kind of thing that would be very informative here, in this setting if we were talking about time to progression for example, is this meta-analysis looking at an array of studies to see whether or not when you achieve a given level of reduction in failure rate and time to progression, does that translate reliably to a given level of reduction in survival.

My biggest concern is to be able to rule out cases where when I achieve a certain response rate or when I achieve a certain reduction in time to progression, does that ever translate into no

benefit? How big do those effects have to be such that we don't get no benefit on survival? Those are answerable questions. We can go to the data and start doing those meta-analyses. They will give us very important insights. Those, however, have to be supplemented. Just to quickly repeat what I said before, we really do need to have a clear sense of mechanism. So, if we are talking about biomarkers, is the biomarker the result of the tumor burden and it is not mediated through the change in the biomarker that the patient has worse survival? I suspect that is the case. So, that wouldn't be a classic example of what we would go for. But basic measures of tumor burden would be the likely candidates that we would be looking for, and if we have interventions that are thought to be fairly safe so that it is unlikely that there would be major unintended negative effects, then we are in the ball park of the kind of evidence that we would be needing to see and the kinds of settings we would need to be in.

DR. PRZEPIORKA: Any burning questions before we move on? Dr. Temple?

DR. TEMPLE: Actually I have a burning question for Dr. Gralla. Most of the time when you

study symptoms you make sure the people entering the trial have one. You wouldn't study headaches in people who didn't have the headache but you thought might get one some day. A lot of the quality of life efforts we have seen do not make sure that the people who are entering the trial are impaired in those dimensions and, even more, even if they have one of the things on your list of physical symptoms they don't have all of them. So, anybody trying to show improvement is starting out with a huge disadvantage because there is no prominence to the symptom.

So, my question is this, we have urged people to think about this, for each patient identify a target symptom, namely, one that they actually have and try to focus on that, even if it was actually different for each patient in the trial. I wonder if you have any thoughts about that. I mean, if I were doing it that would seem the way to find an effect if there is one because you are at least identifying people who have the problem, whereas in so many of the trials we have seen the people don't even have that problem. It is hard to win.

DR. GRALLA: Yes, I understand your point

and I think that is another reason why we have to be careful about setting an absolute number on improvement. Three percent of patients are asymptomatic, three percent. When people ask me how do you treat the asymptomatic patient, I don't worry about it, I just wish more would walk in the door. So, everyone has symptoms.

The question of looking at symptom burden, how do your symptoms affect you is not a bad one to look at in that way because, therefore, it doesn't matter whether it is pain, cough or dyspnea.

DR. TEMPLE: But you want to be sure they are having an effect. It wouldn't be a good question to ask if they said, no, it doesn't bother me, I get through it.

DR. GRALLA: No, no, everyone rates that question from zero to 100. You can rate it zero, you can rate it 100. So, you can see the whole group. If you have 200 patients in an arm, you make up the number and you can see what the scores are. If you start out at baseline with one group being much more symptomatic than the other, then you have big problems but that is not what usually happens. And, what you can see here are differences, real differences when you see drugs

that work. So, what you can see is patients rate the effects of their symptoms as being improved more on treatment A versus treatment B. It is not a huge effect but it is there.

If you want to, you can start with those patients. People have correlated different scores on a visual analog scale with mild, severe and marked. So, if you want to say I only want to look at those patients who rate their pain above 25 at baseline and what happened to that group, you can do that from this same set. But now what we are doing is getting to Dr. Fleming told us. Maybe you don't want to go there; now you are looking at a subset analysis.

DR. TEMPLE: Yes, but I could also stratify and I could make that my primary hypothesis.

DR. GRALLA: You could; you could.

DR. TEMPLE: You could say to yourself if they don't have a whole lot of impairment in this dimension I am not likely to say much benefit. So, I want to make my primary hypothesis people who are very impaired in this dimension.

DR. GRALLA: Yes, I like to think of the opposite criticism. So, you only looked at those

patients who rate their pain. So, is your drug no good for people who don't have pain?

DR. TEMPLE: It doesn't improve their pain.

DR. GRALLA: But what I showed you before, looking at the difference between pemetrexed and CIS, even within responders was eight out of eight parameters favored the combination, a significant difference in itself. This is what the patients say and, to me, that is very compelling. I don't know how the FDA would see that but to me that was very compelling. But no one of those was hugely different but in each one of those areas people looked at it being different. Your suspicion would have been that many of them would have been the same.

DR. TEMPLE: I am only asking because we see so many "unsuccesses" and one of the possible explanations for that is that there isn't much room for improvement. You know, if you have ten items in a score and only one of them is capable of being improved, that is pretty tough. If all ten are, well, you are much more likely to show something.

DR. GRALLA: But the differences in the areas that are looked at here--for example since we

were talking about mesothelioma, there are only five. In the validation studies for the instrument there were only five that were important. When you think of pain and dyspnea and cough and anorexia and this sort of thing--I can't remember the other one, you know it is not too surprising when you get a tumor response. The problem is lung cancer comes up with dyspnea where you have COPD as a concomitant illness. If we have a drug that fixes the COPD we are really in good shape. There you have the confounding variable problem.

DR. PRZEPIORKA: Thank you. Thank you to all the speakers. I would like to now open the open public hearing and call to the podium Mr. Mark Scott. While he is coming up to the podium I have been asked to read a statement about financial disclosure.

Both the FDA and the public believe in a transparent process for information gathering and decision-making. To ensure such transparency at the open public hearing session of the advisory committee meeting, the FDA believes that it is important to understand the context of an individual's presentation. For this reason, the FDA encourages you, the open public hearing

speaker, at the beginning of your written or oral statement to advise the committee of any financial relationship that you have with any company of any group that is likely to be impacted by the topic of the meeting. For example, the financial infection may include a company's or a group's payment for your travel, lodging or other expenses in connection with your attendance at this meeting. Likewise, FDA encourages you at the beginning of your statement to advise the committee if you do not have a financial relationship. If you choose not to address this issue of financial relationship at the end of your statement, it will not preclude you from speaking. You may go ahead.

Open Public Hearing

MR. SCOTT: My name is Mark Scott. I am the executive director for development in the U.S. and I work for AstraZeneca Pharmaceuticals so that would be the financial interest, and they did pay my way here today.

[Laughter]

Madam Chairman, members of the committee, ladies and gentlemen, thank you for the opportunity to speak. I am representing actually AstraZeneca Oncology for this presentation today and I believe

in your package you received a seven-page document outlining a number of points we intended to make as part of this committee meeting.

I believe that most of the points have already been discussed today so I want to go into them with the detail I had originally intended. Some of the points were made this morning and some of the points are directly relevant to the discussion you will have after this with respect to the questions that are being addressed.

The first point is that we wanted to endorse the committee discussion on symptomatic improvement as used as the basis for full approval for oncologic agents, and especially for non-small cell lung cancer as it is a disease of symptoms. With well validated scales that are available, including the lung cancer symptom scale, a demonstration of relief of these symptoms as determined by well conducted and controlled patient-reported outcome studies could be acceptable as a sole basis for full approval of new agents.

The next area was in trials in subsets of patients, specifically performance status II. This wasn't necessarily directly germane to the

discussion but, given that you are talking about lung cancer, we thought it to be important. Inclusion and exclusion criteria for many clinical trials in non-small cell lung cancer exclude performance status II patients because of their short life expectancy and because many are considered unsuitable for cytotoxic chemotherapy.

Novel agents with better tolerability may offer a chance to bring clinical benefit to this ill-served patient population. The FDA has recently granted fast-track status for a compound to be investigated in a trial in performance status II patients and we are asking the committee do they agree that a PS-II population in advanced non-small cell lung cancer is an identifiable population worthy of clinical study, and for whom an indication could be written? If the answer was no, how would they propose to define the population of patients often considered too unfit to tolerate chemotherapy and, therefore, being excluded from many current clinical trials?

Another area that we wanted some debate about which got covered this morning is that we are very encouraged that there was a recommendation by the committee that progression-free survival could

serve as the sole basis for approval in certain situations.

The last area we wanted to discuss was the efficacy standard, and I will not go into it in great detail but it has to do with non-inferiority trials, which I will talk about at the end. We would briefly like to reinforce the implications for oncologic drug development as raised by Dr. Williams this morning. It is actually through an article by Rothman et al. that was published in the January, 2003 edition of Statistics in Medicine on non-inferiority trials. The methods described in this article are increasingly used by regulators in the United States and Europe to evaluate the design analysis of trials of new agents. The consequences for trial size are enormous as a result of this paper.

In this context, there has been something of a paradigm shift though in the approach to cancer treatment over the recent years. Academia and industry alike are now fully engaged in the discover, research and development of novel, well tolerated, biologically targeted anti-cancer agents. It is hoped that these new treatments will offer significant advantages to patients in terms

of improved tolerability, but they may not always demonstrate increased efficacy. This naturally leads to the use of active control in non-inferiority trials to compare the new agent standard to standard agents, with the conventional aim being to show no clinically relevant loss of efficacy.

But the key problem for researchers, physicians and patients alike is that with Rothman's approach there is a dramatic increase in the size of the trial required to determine non-inferiority. We don't believe that the answer is to avoid non-inferiority trials. We believe that there are situations that are clinically relevant where a non-inferiority trial would be the trial of choice to define efficacy.

We don't believe that the scientific statistical debate about how to best draw inferences from active control, non-inferiority trials should be considered complete. Rothman's approach serves to highlight that considerable statistical, methodological and philosophical issues remain, and failure to consider these issues constructively will, at the very least, lead to ever-increasing drug development costs, time, and

delay the availability of new therapeutic options to patients with life-threatening diseases. At worst, the barriers posed will discourage drug development where it otherwise might have been feasible and so prevent potentially useful new medicines from becoming available to patients.

We sincerely hope the scientific community, together with regulatory bodies worldwide will give this important area further careful thought, and we, at AstraZeneca, recommend that the advisory committee here, as well as academic interest and industry interest have a panel like this meeting to address this issue. Thank you.

Questions for Discussion

DR. PRZEPIORKA: Any questions for Mr. Scott?

[No response]

Thank you. Our hosts have provided some guidance, if you will, on the importance of the questions and, given the hour, we will be taking these out of order.

The first question to be discussed will be question seven, under the surgical adjuvant setting. The FDA has stated that disease-free

survival can support regular drug approval in cancers where the majority of recurrences are symptomatic. Others propose that prolongation of disease-free survival should support regular approval in all clinical settings because a delay in cancer detection or a delay in the need for toxic cancer treatment is of clinical benefit.

In non-small cell lung cancer, should a disease-free survival improvement from adjuvant chemotherapy support regular drug approval? If so, clarify why you consider disease-free survival an established surrogate for clinical benefit in this setting.

Part b) is if not, could a disease-free survival improvement support accelerated approval? Would a survival advantage ultimately be required for conversion to regular approval?

So, the question before us is should disease-free survival in the adjuvant setting be a primary endpoint or a surrogate for survival. Dr. Johnson?

DR. B. JOHNSON: I think this is a more a philosophical than a real question in that adjuvant therapy hasn't yet been proven to play a role in lung cancer, and I can't imagine--I don't know of

any company that has a plan to look at this. So, it is not something that is going to come up for three to five years. So, I think yes is probably the answer but I don't think it is terribly important to define the answer at this time.

DR. PRZEPIORKA: Just to question you, you indicated that there has been no drug that has been shown to have an advantage in that setting. Was that based on survival as opposed to disease-free survival, and would you be willing to suggest that disease-free survival would be an appropriate endpoint rather than survival?

DR. B. JOHNSON: There are two studies that have been presented in abstract form that Paul talked about, and it looks like there will likely be an advantage for at least one of those two studies when it gets published and the disease-free survival fits with the actual survival. The point I was trying to make is I can't imagine that somebody is going to submit for approval a new drug unless you are going to be approving it for a new indication.

DR. PRZEPIORKA: Dr. Johnson?

DR. D. JOHNSON: Dr. Bruce Johnson and I decided ahead of time to avoid the confusion that

the good-looking Johnson--

[Laughter]

DR. PRZEPIORKA: You are also in alphabetical order!

DR. D. JOHNSON: I would say yes, disease-free survival can be used as a primary endpoint and I would say that I would interpret the two studies that have been presented slightly differently. One will be published in The New England Journal soon, which was presented at a plenary session at ASCO this year. It is really the only study that is sufficiently large to address this question. It was an international study, done largely out of France. The disease-free survival essentially mirrors the overall survival. This is essentially identical to what we see in breast cancer adjuvant trials.

The second trial, which shows the same pattern, is a trial out of Japan which used a drug that is not available in the U.S., UFT. It too showed a disease-free survival that was reflected in the overall survival.

So, I personally think that this is a worthwhile endpoint. If it is going to be used in future trials, I think DFS can be used as it is in

breast cancer adjuvant trials.

DR. PRZEPIORKA: Other comments from our experts?

DR. BONOMI: I agree and I think you are going to see that there is going to be a lot more activity in this area with these trials, especially with the ALT trial turning out to be positive. I know the cooperative groups are gearing up to do new studies.

DR. ETTINGER: There are two studies. One is the Canadian study that has been completed with vinorelbine/CIS that we await with bated breath in early disease, stage I actually, and there is the CALGB study that is very similar with a different set of drugs, hopefully, going in the same direction otherwise we will have a real problem on our hands. Right now we have the ALPI study, although there was a trend that was negative, and we have the ALT study that obviously is positive.

So, I agree that disease-free survival in that study as well as the UFT study in Japan show that the disease-free survival and survival are in the same direction and should be able to use either one of them or both.

DR. PRZEPIORKA: Other questions?

Comments? Ms. Ross?

MS. ROSS: Just a quick comment because my duty here is to represent patients, and the status quo is not acceptable. We can't remain with a 14 percent survival rate with lung cancer. We have to open this up. Yes, I would agree with that position. Please open it up.

DR. PRZEPIORKA: Do you have other points you want us to discuss with that question? No? Okay.

DR. WILLIAMS: There is one other issue though. I would like you to vote on it.

DR. PRZEPIORKA: To vote on it?

[Multi-member discussion]

DR. WILLIAMS: What we are asking for, call it what you want, is would you grant full approval for this? That is the question before you--or regular approval.

DR. PRZEPIORKA: If we get a positive vote on a) we won't need to vote on b) then. Going around the table then, the question before us is in the surgical adjuvant setting would one accept disease-free survival improvement to support regular full approval for a drug. Dr. Ettinger?

DR. ETTINGER: Yes.

DR. PRZEPIORKA: Dr. Saxon?
DR. SAXON: No.
DR. BONOMI: No.
DR. D. JOHNSON: Yes.
DR. B. JOHNSON: Yes.
DR. GRILLO-LOPEZ: Although I don't have a vote, if I had one I would like you to know that I would vote yes.
[Laughter]
DR. GEORGE: Yes.
DR. CHESON: Yes.
DR. DOROSHOW: Yes.
DR. RODRIGUEZ: Yes.
DR. BRAWLEY: Yes.
MS. ROSS: Yes.
DR. FLEMING: Conditionally yes. Sorry, I have to give a condition because it wasn't totally clear to me. If we can say consistently that at recurrence there are symptoms, then that makes it what I would call a level one outcome. Short of that, if we can put forward data that would indicate that there is a clear consistency between effects on disease-free survival and effects on survival that would also be the basis.
DR. LEVINE: Yes.

DR. REAMAN: Yes.

DR. PRZEPIORKA: Yes.

MS. HAYLOCK: Yes.

DR. CARPENTER: Yes.

DR. REDMAN: Yes.

DR. TAYLOR: Yes.

DR. PRZEPIORKA: It is overwhelmingly yes so we will forego b).

Back to the first page of the afternoon session, first-line non-small cell lung cancer treatment setting, approval based on demonstrating superior time to progression. So, considering the pros and cons that we all discussed this morning in the time to progression session, for approval of drugs for first-line treatment of advanced lung cancer, could time to progression benefit of a new drug compared to a standard first-line regimen justify regular full approval? Assume that the standard control arm has a known small, two-month, benefit. Comments?

DR. CHESON: So, we are really keeping this at time to progression and not progression-free survival?

DR. WILLIAMS: Why don't you change it to progression-free survival?

DR. PRZEPIORKA: Progression-free survival.

DR. WILLIAMS: Thank you, you have made it easier.

DR. PRZEPIORKA: Dr. Johnson?

DR. D. JOHNSON: Actually, my comments were relative to time to progression, but actually I just want to make one other point that may be self-evident to everybody at the table but it may be more germane to Dr. Bunn's comments vis-a-vis response. One of the problems I think in lung cancer studies is the tremendous heterogeneity of the population that we study. I think one of the problems that FDA faces and this advisory committee faces when it comes to lung cancer is the fact that there has been a stage creep that affects us. Stage IV disease is very much more homogeneous and a lot of the data that I think that Dr. Bunn presented really applies principally to stage IV disease. When you start including unresectable stage III disease, first of all, you have to define unresectable and then you have to define which stage III disease one is dealing with. At least in cooperative group trials, a review of the database shows as much as a three-month difference in median

survival in various so-called unresectable stage III patients relative to stage IV. That is actually the difference that many trials are designed to see. None, as Dr. Bruce Johnson has shown, actually has quite achieved that level in advanced disease. Typically, the best one sees is about a two-month improvement in the so-called statistically positive trials in stage IV.

So, I just want to make this point. It also has to do with response rates because response rates are consistently higher in patients with unresectable but locally advanced disease as compared to patients that have metastatic, extrathoracic metastases. So, there is a huge issue here that I didn't really hear addressed but I am assuming, maybe incorrectly, that this particular committee is familiar with and knows about.

DR. PRZEPIORKA: Would you feel more comfortable asking this question in a metastatic setting versus the non-metastatic setting separately?

DR. D. JOHNSON: I think it would be helpful to our colleagues at FDA but maybe they can answer that question for themselves.

DR. PRZEPIORKA: Would you like to hear that?

DR. WILLIAMS: Certainly, if it makes a difference, we would.

DR. PRZEPIORKA: Other comments before we move to vote? Dr. Fleming?

DR. FLEMING: I would be interested to know if there is more evidence to put on the table than what I have heard thus far. The distinction here between what I have been calling a level two as a marker versus level three is profound. Level three means it is reasonably likely to predict clinical benefit. Level two is, is it reliable? It is reliable evidence; it is established. Across clinical areas the number of established surrogates is really small. They are very rare. It takes striking evidence to be able to reliably say that the effect on this marker will tell us the effect on the clinical endpoint.

When this FDA/ASCO group met, after several meetings the summary of the conclusions, which are presented in this document, basically were it has not been established that the benefit on TTP reliably predicts benefit on survival--reliably predicts. Listening to Paul's

presentation, the vast majority of it was advocating for greater attention to response. His comments indicated, if anything, some real skepticism, pointing out a number of inconsistencies in time to progression prediction of survival. So, I would consider that a fairly negative summary that, in fact, endorsed what the FDA/ASCO summary indicated after its sessions. But maybe there are more comprehensive analyses other people have done that can give a more positive view than this.

Essentially I am trying to summarize what I heard at FDA/ASCO and what I heard from Paul. It sounds as though for time to progression these data are well short of what we would typically think of as necessary to say reliable.

DR. WILLIAMS: Tom, I think some of those things we were talking about this morning really need to be discussed a little bit here. Does it matter that there is a short difference between time to progression and survival, and which way does it matter? Does it make it more acceptable or less acceptable? Do you think there are symptoms when people progress and, therefore, is that the reason you would accept it? You know, what would

be the pros and cons of accepting it here? So, I think a bit of discussion on that point would be helpful.

DR. B. JOHNSON: One of the potential means for this is that this will pick up an important endpoint that survival misses. The length of time between time to progression and death in advanced disease is very short. So, the help of that would be very small as a surrogate to outcome.

The second potential problem is that now with therapies in the second- and third-line you would have problems in interpreting data that the randomized did not take care of. To me, that is a hypothetical problem; not a problem that has been proven to be shown. So, I don't see that adding a time to progression or progression-free survival would be particularly helpful in interpreting the trials.

DR. D. JOHNSON: I don't know if this helps, Tom, but one thing that we have done over the last several years is to do a detailed analysis of the ECOG database for advanced disease, with all of the recognized limitations of such an analysis. But what I can say is that at least in stage IV

disease--which is fairly reliably diagnosable, perhaps even more so today but certainly in the '80s and '90s with CT scans one could pretty reliably diagnose stage IV disease--one thing we observed is at the time of progression, as documented by the individual taking care of the patient, typically by a physical finding or a new radiographic finding, before widespread availability of second-line treatment or the widespread acceptability of that, the median survival of patients from that point forward was approximately 14 weeks or so. That was borne out in the docetaxel study that Dr. Cohen alluded to where the median survival of patients after first-line therapy was four months. What docetaxel did was extend that by approximately two and a half months, more or less, in one study not in the second study.

We did an analysis which we then presented this year at ASCO, looking at the ECOG trials subsequent to the approval of docetaxel. That is, presumably the widespread availability of second-line therapy. What we found was that the median survival of patients from progression was extended by approximately six weeks beyond what it

had been according to the data prior to that. Again, this more or less validates in my mind the data that we saw in that relatively small trial of docetaxel.

Another thing we did during that same analysis which was of interest to me, and I presented this at the forum, were two separate analyses. Again, we are talking almost exclusively about stage IV disease. These data were developed in patients, 85-90 percent of whom had documented stage IV disease. Patients that had disease control--forget about whether their tumor got smaller or not but they didn't progress, did as well regardless of whether their disease got smaller by X amount, 30 percent, 40 percent or whatever. Those patients had virtually identical survivals.

The other thing we looked at was percent of progression at various time points. We chose time points when physicians would have evaluated patients according to the protocol. So, that would be every three weeks or every four weeks, whatever. It didn't really matter whether one chose three weeks, six weeks, nine weeks or whatever. If one selected a time point and then calculated the

percent of progressors, non-progressors, in only those studies where there was a statistically significant survival benefit was there a difference in percent of non-progressors in favor of the arm that did better, if you follow what I am saying.

So, it is a little bit different than progression-free survival, but it is a fixed time point where one can say X amount of patients are progressing at this point in time, fewer in this group and this group does better. And, that was surrogate, if you will, of survival. So, we looked at those. I think that was something you were talking about earlier, could one use some marker of that nature to do that.

DR. FLEMING: The evidence that we really need here would be a wide array of studies, conducted in a given setting where we are advocating the use of a given marker as the reliable evidence of benefit that would show treatment-induced effects on that marker at a certain level which are always going to tell us that we have treatment-induced effects on survival and, more generally, that the relationship between those two is very strong. Some of the examples that Paul gave were ones that gave very

inconsistent results in progression from survival. He also mentioned the ECOG 1594, saying that the GC arm was a month and a half longer in time to progression, suggesting a difference but the survival effects were the same.

DR. D. JOHNSON: Actually, those survival results are not the same. They are not statistically significantly different but actually the better survival is in that arm. But that is a whole other argument. I would disagree with Paul's analysis of that particular data.

But let me say this, that what we did was develop those markers in one set of data, 5592 which was the predecessor trial and was a three-arm trial, and we tested the model in the 1594 data. We also went back and tested it in another data set, 1583, which was a study that Dr. Bonomi chaired back in 1983. He is not that old; he just looks that old--

[Laughter]

--and again validated those endpoints in the same direction. There was a survival advantage in his study with carboplatin as a single agent and, yet, it had the lowest objective response rate. But the percent of patients who progressed

at various time points was lower in that particular arm. There was "crossover" but only a small percentage of patients actually crossed over. But it was that percent of noon-progressors that actually best correlated with outcome in that particular study.

DR. FLEMING: But you are saying the aggregate data showed a lower time to progression in the arm--

DR. DL. JOHNSON: No, what I am saying is the objective response rate in 1583 for carboplatin as a single agent was nine percent. That was the lowest overall response rate. The highest response rate was 27 percent, as I recall, a three-fold difference in response rate, and yet the 27 percent group had the lowest, statistically less survival compared to carboplatin. But then when we applied our rule of non-progression, and you could pick the point you want, after two cycles, after three cycles or whatever, not looking at objective response rate but non-progression it comes out in favor of the carboplatin arm, just as we had predicted from the 5592 data and 1494 data and then applied to the 1583 data. So, there were three separate databases.

DR. FLEMING: It is this kind of data that certainly gives one concern about the reliability of the response predictor where you are telling us it goes in the wrong direction. More broadly, for time to progression or any other measure of tumor burden what one needs is much more evidence than what I am hearing, and it may exist but just needs to be looked at in a meta-analysis framework to understand whether treatment-induced effects on whatever measure you are advocating--time to progression right now-- is reliably telling us treatment-induced effects on clinical endpoints such as survival.

DR. PRZEPIORKA: Dr. Bonomi?

DR. BONOMI: I want to make one comment. The MBP regimen is a peculiar regimen. I don't know if Dick Gralla is still here. We used a very low dose of cisplatin, 40 mg/m2, and some people would say, and I think Dick would be one of them, that dose might be below or right at the minimum effective dose. The point I want to make is there is discordance between response and survival in the study but that particular regimen isn't a good one to base it on because in three consecutive studies it gave the highest response rate, statistically

significant in I think two out of the three, and a trend for a shorter survival. In fact, when it was lumped together it actually gave a significantly lower one-year survival rate, MBP did. So, higher response rate, lower survival. We thought that regimen either was doing something detrimental in people or possibly the platinum dose was too low. Mitomycin might have been detrimental. We thought it was a combination of toxicity and the actual anti-tumor effects. That is a peculiar regimen. I wouldn't want to base any correlation response and survival on that particular one.

DR. FLEMING: But that really gets at the essence of what leads these predictors to not be reliable. It is not that they are irrelevant; they are relevant but are they adequately relevant? Are they adequately capturing the complexities of how the disease process influences the outcome, and are they adequately capturing some of the unintended effects? This is the heart of why these are often misleading.

DR. GRALLA: If I could make a comment?

DR. WILLIAMS: You need a mike.

DR. PRZEPIORKA: Will you take the podium?

DR. GRALLA: There are other aspects, suck

as Lucio Guino's study where, with different doses of cisplatin, he finds that the same drugs put together differently equal, for example, gemcitabine/cisplatin which is approved.

I think that we can find exceptions, but what I think Paul was trying to do was to put them all together. He was looking at single agents. When you put single agents together at the doses at which they are used, you do find exceptions but what you find is a fairly strong correlation between response and survival. You know, we can put together regimens in ways that don't have duration of response, that are too low to do that. So, I think Paul was looking at single agents, not combinations that are more subject to that because when you put that together differently you can get a different result.

DR. FLEMING: But, Richard, a lot of that single agent was Phase II data and that is not the kind of data that you need to have to validate a surrogate because that is just getting at correlation of response and the outcomes. That is just a foot in the door step.

DR. GRALLA: It may be. I mean, you are right, many of those were Phase II studies. I

think if you looked at the randomized studies looking at single agents though you would come up with a clearer correlation between survival and response but we only have about 15 or 20 of those in the last few years.

I must say, in my heart of hearts I believe really ultimately response does agree with survival. The question is are the data robust enough to agree with that at this time, and that I am not sure of and why wouldn't we want to look at the data to see that rather than just have an opinion?

DR. PRZEPIORKA: We will get back to the question of progression-free survival. Dr. Johnson, before I could answer this question the question I really have for you or anyone else in the expert row there is would you limit enrollment in such a study on the basis of performance status? If, in fact, we want to use progression-free survival as the ultimate reason for approval and we think progression-free survival is actually a measure of clinical benefit, is it going to be likely in somebody who has ECOG performance status II or are we looking for people who are pretty healthy looking people?

DR. D. JOHNSON: Well, I think most of the data that have been developed in the last decade has really been restricted to patients with performance status 0 or I. We could debate about II should be allowed or not but, frankly, the numbers here are not generally a problem. So, I personally think restricting to 0 or I is still the way to go. There is a higher level of toxicity associated with performance level II. Actually, response rates tend to be fairly similar across the performance status and we have shown that several times in the ECOG database but the toxicity levels are much different. So, I personally think it should be preferentially in patients with performance status 0 and I. I wouldn't mandate that it be limited that way but I would certainly urge that that be done in that fashion.

DR. PRZEPIORKA: Dr. Ettinger?

DR. ETTINGER: Since progression-free survival in my opinion is a fuzzy endpoint, it seems to me the quality of life issue becomes paramount. Therefore, I would say you want patients that are symptomatic if you are going to use that as an endpoint because then there is clinical benefit, and I think that is critical and

I think that is what the patient wants. If the survival didn't come out to be statistically significant, at least there was a clinical benefit and that is enough to approve a drug, especially if the progression-free survival was in the right direction that was statistically significant.

DR. TEMPLE: Just to make the point, we have long said that improvement in symptoms is a basis for full approval. That is why we haven't been asking you about that. So, that is already true and we haven't had any reason to debate it. The question here is suppose you don't have that. So, if you have that along with whatever it is, you are fine; that is not an issue.

DR. PRZEPIORKA: Dr. Williams?

DR. WILLIAMS: First, I believe Dr. Johnson is saying that you believe there probably is a correlation, at least that it could be that progression-free survival could be a substitute or a surrogate for survival. Perhaps we don't have all the data yet to validate it as such. So, I would like to pursue a little bit further also whether or not in these patients you believe that progression is an indicator of symptoms and that would be the other basis where you might consider

this endpoint--a little discussion on that matter.

DR. D. JOHNSON: Well, I got off in a little o- bit of a tangent. The point I was trying to make when I was talking with Dr. Fleming is the fact that I do believe progression-free survival is a valid endpoint, and I do think that upon progression, even in this era when we have second-line therapy, the overall survival after that is not that good. I mean, it is really pretty modest and those patients are for the most part symptomatic. Most of the recurrences take place because the patient walks back in your office not on a scheduled visit but because they have new lung pain, or they had a seizure, or they are short of breath, or they are coughing up blood, or they are coughing their lungs out. So, this is not a subtle thing in most instances. We don't find it on screening PET scans. It is the type of thing that patients are really quite symptomatic.

So, I do think prolonging their progression-free is almost tantamount to their symptom improvement, not symptom free because they rarely completely resolve their symptoms.

I might add that the first drug that showed benefit in non-small cell lung cancer that

we know about was published in 1948 in Cancer by David Karnofsky and it was nitrogen mustard. Nitrogen mustard actually--the reason that he recommended its usage was not because it induced tumor regression but because it improved symptoms in 70 percent of patients. I am mindful of the fact that the FDA did approve gefitinib because of its objective response in symptom improvement, and the rapidity with which that occurred I think was on average eight days. If you go back and read Dr. Karnofsky's paper you will note that nitrogen mustard which, by the way, most of us don't use to treat lung cancer these days, improved symptoms in approximately six to seven days. Procarbazine has been shown to do the same thing too in non-small cell lung cancer. So, this is not a new concept. This has been going on for 55 years.

DR. PRZEPIORKA: Other discussion that you need before the vote?

[No response]

As recommended by Dr. Johnson, we will split this out looking at locally advanced versus metastatic disease, and we will start with the metastatic patients. So, would you consider progression-free survival as an appropriate

endpoint for full approval for a patient with metastatic non-small cell lung cancer? We will start with Dr. Taylor and work our way around.

DR. TAYLOR: no.

DR. REDMAN: Yes.

DR. CARPENTER: yes.

MS. HAYLOCK: Yes.

DR. PRZEPIORKA: Yes.

DR. REAMAN: Yes.

DR. LEVINE: Yes.

DR. FLEMING: No, and just to amplify a bit, there is a correlation here but I still think that the essence of the nature of what we need still maybe hasn't gotten clarified adequately. There is a correlation between those people who have a longer time to progression and those people who have a longer time of survival. The evidence, at least as was brought forward before the ASCO/FDA group and the evidence that Paul Bunn brought forward today certainly brings out that there are serious concerns about whether we can rely on time to progression effects to predict survival effects. Symptomatic effects have been mentioned. I wonder if the best way to measure symptom improvement is through time to progression or whether it would be

through some of Richard's approaches that he has indicated using PROs.

But, in essence, the number of truly validated surrogates are rare in clinical practice. I think the data that we would need potentially could be out there but they haven't been brought forth to be analyzed.

DR. PRZEPIORKA: Ms. Ross?

MS. ROSS: Yes.

DR. RODRIGUEZ: Yes.

DR. DOROSHOW: No.

DR. CHESON: No.

DR. GEORGE: Yes.

DR. B. JOHNSON: No.

DR. D. JOHNSON: Yes.

DR. BONOMI: Suggestive but no.

DR. SAXMAN: No.

DR. ETTINGER: No.

DR. PRZEPIORKA: So, it is 8 no and 11 yes.

DR. WILLIAMS: Can we do a subgroup analysis? Any particular group occur to you?

[Laughter]

DR. PRZEPIORKA: Let's do the second part and see if that changes.

DR. WILLIAMS: Okay, go ahead.

DR. PRZEPIORKA: So, those with inoperable, locally advanced disease, would you use progression-free survival as your primary endpoint for approval? We will start with Dr. Ettinger.

DR. ETTINGER: No.

DR. SAXMAN: No.

DR. BONOMI: No.

DR. D. JOHNSON: No.

DR. B. JOHNSON: No.

DR. GEORGE: No.

DR. CHESON: No.

DR. DOROSHOW: No.

DR. RODRIGUEZ: No.

MS. ROSS: Yes.

DR. FLEMING: No.

DR. LEVINE: No.

DR. REAMAN: No.

DR. PRZEPIORKA: No.

MS. HAYLOCK: Yes.

DR. CARPENTER: No.

DR. REDMAN: Yes.

DR. TAYLOR: No.

DR. PRZEPIORKA: Overwhelming no. So, clearly that reflected the discussion earlier

regarding a slightly better prognosis group that you want to get good, hard endpoints in.

DR. WILLIAMS: So, in patients that might be more symptomatic or more likely to be symptomatic upon progression the "non-lungers" said yes and the "lungers," except for one, said no. That is what I heard.

DR. PRZEPIORKA: Do you want us to continue on question two regarding the metastatic patients?

DR. WILLIAMS: No, why don't we move on?

DR. PRZEPIORKA: Well, we can move on because we have said no. If it doesn't support full approval, would it support accelerated approval? We will again start with Dr. Ettinger.

DR. ETTINGER: No.

DR. SAXMAN: I think that would depend on the magnitude so I guess the answer is yes.

DR. WILLIAMS: Let me just give a little guidance here now. The accelerated approval regulations say that you must show an advantage over available therapy. Let's say this is a first-line therapy with a survival advantage and you are showing a TTP advantage over it so what you need to ask is, is this endpoint reasonably likely

to predict clinical benefit. You don't have to show that there is clinical benefit. So, that is the call for accelerated approval, to feel that this is reasonably likely to predict clinical benefit. So, you can also discuss the magnitude but I just wanted to make sure that that was clear.

DR. SAXMAN: That is TTP.

DR. WILLIAMS: Or progression-free survival, or we will substitute that for each of these.

DR. SAXMAN: What about accelerated approval?

DR. WILLIAMS: Accelerated approval. In other words, you are getting the best thing out there with respect to time to progression or progression-free survival.

DR. SAXMAN: With the idea that full approval was intended upon subsequent survival advantage.

DR. WILLIAMS: Right.

DR. BONOMI: I will say yes on that one.

DR. D. JOHNSON: Yes.

DR. B. JOHNSON: Yes.

DR. GEORGE: Yes, assuming all those methodologic issues are addressed that we

discussed.

DR. CHESON: Yes.

DR. DOROSHOW: Yes.

MS. ROSS: Yes.

DR. FLEMING: Abstain.

DR. REAMAN: Yes.

DR. PRZEPIORKA: Yes.

MS. HAYLOCK: Yes.

DR. CARPENTER: Yes.

DR. REDMAN: Yes.

DR. TAYLOR: Yes.

DR. PRZEPIORKA: That is overwhelmingly yes. Then we have to answer the more important question which is what would be the interval that you would want to see to say that your progression-free survival was of clinical benefit. It is open for discussion. Dr. Johnson?

DR. B. JOHNSON: About three months beyond control.

DR. WILLIAMS: We are talking about accelerated approval now, right? So, we are talking about what would be a surrogate reasonably likely to predict clinical benefit.

DR. PRZEPIORKA: Dr. Carpenter?

DR. CARPENTER: All the differences in

therapy we have heard about were all either in the two-month or the three-month range of any therapy over another, if I understand the experts. It would seem unrealistic to expect anything larger than that of a new therapy, or not very likely. So, David mentioned the biggest difference in survival and the disease-free survival threshold level usually pretty closely parallels that. I think that the data needed for accelerated approval would have to be pretty compelling and there would need to be a large, well-controlled study that showed a difference that is larger than we typically see for survival with best supportive care with a doublet. I think it would need to be at least three months.

DR. PRZEPIORKA: Dr. Johnson?

DR. D. JOHNSON: Just to give some context and, again, I think you have to think about this in stages and stage IV I would argue is the most homogeneous group in a group about whom we have the most data accurately in terms of these numbers. So, median survival in stage IV disease is about seven and a half, maybe eight months with PS-0 in one patient. If you throw II's in that drops down. The median time to progression in SWOG and ECOG

trials is pretty reliably--the time to progression, not progression-free survival--is about three and a half months. You saw that in the 1594 data. That is unbelievably reproducible. I use that all the time. You can just about double the time to progression in most of the cooperative group trials and you can come up with the survival, median survival. That is what it is going to be.

Now, progression-free survival is a little bit harder to come up with because those data haven't been as well characterized, at least within the cooperative group data. But I would agree with Bruce. I think if one is looking for accelerated approval one needs to see something that is more than just a few weeks difference in progression-free survival, and I think three months may be unattainable. I don't know but you are talking about accelerated approval here and I would agree with that number.

There is one method-logic question that has been posed which I think may be germane even in the accelerated approval setting and that is should the trial be blinded and, if it is not, or even if it is, should progression be verified by a blinded central reading of scans. One shakes their head

yes, one, no.

DR. BONOMI: I don't think so. David has pointed out it is pretty obvious when these people are progressing and I think probably you don't need to go to that degree of rigor. Maybe David might dissent.

DR. D. JOHNSON: No; I don't dissent. I just want to point out that, in the studies, at least the ones I have been involved in, where there has been a review committee that reads the X-rays, there is as much disagreement amongst the review committee as there is amongst the original investigators. So I am not sure who is truly accurate in reading these.

Actually, it is my personal view that the way to get better rigor is not to have someone else read the films but to have someone consistently read the films at one's institution. That way, I think one gets more accurate. But that is a debate for another day, I think.

DR. BONOMI: One other thing. I think more and more places now have digital radiographs with a cursor and you can measure it. There was just a paper in JCO that is what Dave said; it should be one person reading these things

consistently. You can keep it, put it in a power point presentation. If somebody wants to look later and see what you did, they can see exactly what they did. The reading stays right on there in millimeters. It is much more reliable than it used to be but it should be one person.

DR. B. JOHNSON: One point of clarification. When you talk about blinded, is it blinded to the treatment or is it blinded for determining the time of progression?

DR. PRZEPIORKA: Either.

DR. B. JOHNSON: One of the things, and I think we have heard this consistently, it is nice to blind you to the treatment but, if you are getting some kind of I.V. infusion, I don't think it is going to be ethically or practically possible to blind you to the treatment.

So I think it depends on the circumstances. If it is a pill, certainly. If it is a 14-day infusion, no.

DR. PRZEPIORKA: Dr. Temple.

DR. TEMPLE: I am having a disconnect. The question here is about time to progression irrespective of whether the person is symptomatic. What you are all saying is they are always

symptomatic, or almost symptomatic, and that is what makes you know they have progressed. But we never see that. We are never given data that show symptomatic progression. If it is that easy, why isn't everybody collecting it because then there would be regular approval. It wouldn't be accelerated. There wouldn't even be a discussion.

DR. D. JOHNSON: I am reminded of the time that I sat in this committee informally as a member and this is like deja vu because I remember your comments many times, Bob--

DR. TEMPLE: Sorry.

DR. D. JOHNSON: No, no. I am glad to find you are consistent. In my after-ODAC life, I have been involved in advising folks and I have made that point many times that it is something. I think Richard has made the point many, many times as well. We, basically, agree with you. We do think that that is a reason for approval of drugs and we would like to see more of it ourselves.

So I can't answer why people don't do it. But I am also reminded of one of my favorite quotes. I actually put it--after I heard you make this quote, I actually had my wife embroider it and it is on my wall. It is listed there, "Bob Temple,

FDA, Survival Trumps Everything." That was a quote from you and I have never forgotten that. So we always remind people when they--

DR. TEMPLE: Just one other observation; we have also asked people, even if you are not absolutely sure that, at the time of radiologic progression, there are symptoms. It has always been our assumption that, in something like lung cancer, symptomatic progression must be fairly near at hand, even if they have crossed over or stopped the drug.

We have invited people to look for symptomatic progression at any time, even if they are off therapy or moved out and, again, gotten very little interest in doing that.

DR. B. JOHNSON: Let me make a comment about this. It has to do with the clinical practice of it. One of the things that happens is, when we go in to see somebody and they tell us they have shortness of breath, you examine them and they have decreased breath sounds half of the way up, you send them for a chest X-ray and you get the chest X-ray and it shows a new pleural effusion and enlarging nodules. The thing I always tell the patient--well, usually I tell them when they are

responding, responding, getting better, it is easier to make jokes when they responding.

But we say, well, one of the things that's nice about being an oncologist is it is not that complicated because 95 percent of the time the radiographs agree with the symptoms. Now, we have grown up with radiographs as our objective criteria for assessing disease progression. So that gets categorized not as a symptomatic progression but it gets categorized as a radiographic progression because that is what has been reviewed in every cooperative-group study.

Now, one of the things that Richard has talked to us about is that the symptom scales have evolved so that they may be more objective than assessing radiographic response which will be a step forward in being able to recognize and use the data. That hasn't been something that hasn't been easily available to us outside of a clinical-study setting.

DR. GRALLA: One of the problems has been feasibility. The point is if you see the X-ray that Bruce is pointing to, you say, well, why do I need to validate this on a scale. I have this. Unfortunately, we have often gone from the chest

X-ray to the CAT-scan so it is $1,000 procedure that you wait for a little while on.

It has been necessary to convert these scales to easy ways. They are not like on a palm-pilot, some of them. They are just being rolled out in trials. This should make it easy. But how do we now adopt that into clinical practice because we are not used to doing that and, God knows, getting us to change is the hard part.

So you have got this case-report form that is 40 pages long and the rest of this and now you want to add something else to it. That is why I think you haven't seen it but I think it is up to us now, from the cooperative group and from other areas, to get this so you so you can see it in a way where most of the patients have it.

DR. PRZEPIORKA: Dr. Temple, just to bring his point back to you and your definition of symptomatic progression. Would you be looking for something on a scale that is objective and you can measure or, as he points out, the patient says, I'm short of breath? Is that enough to say this is a symptomatic progression?

DR. TEMPLE: That is a fair question. If we are all blinded, it would be a much easier

question because then you could accept a lot of things. But there are people here much better able to think about that than me, but somebody showed the five or six things that are most of what bother patients.

If there were some systematic question that even asked them on a ten-point scale, how is your fatigue, your this, your this, your this, your this, and that was done regularly. When it looked worse, you then sent them out for an X-ray. That would greatly help the persuasiveness of that finding of progression as a meaningful thing.

The other thing, of course, is if, in several studies, it always came out that way, you would have at least some case for saying that progression pretty much always means symptomatic progression. Then we wouldn't have to do all that anymore.

DR. GRALLA: I think Dr. Taylor pointed out in second line, where we saw these response rates of 6 to 10 percent, do we need to send all these patients for X-rays for this? When the patient tells you that they can't breathe, and you have got a valid way of measuring it, that they have more pain, that they are using more pain

medicine and they are dropping weight like a stone, I am just not sure that we need the chest X-ray, the MRI, the PET scan.

DR. TEMPLE: We totally agree because symptomatic progression is a no-brainer approval, if you believe it--if you believe it. That's important

DR. GRALLA: These instruments do that now. The problem is getting them incorporated into trials in a feasible way. It is the feasibility that is the problem.

DR. B. JOHNSON: There is one other problem that comes up with this. Richard may want to address this. We have gone through the design of a trial now where the symptoms as being assessed on one of the formal scales and the design want to withhold that information from the physician because they think it will bias the physician's decision-making.

We are wrestling with the ethical dilemma about do you withhold patient information from the treating physician with the potential of biasing the outcome. I would like to hear Richard's comments on this.

DR. GRALLA: It is a great point, Bruce.

We are doing a 200-patient trial in Ontario right now trying to look at that, trying to look at how these data affect--did these data affect the physician decision-making. So I hope we have some information there. I think it is going to be difficult to say because the patient comes in and has pain. As David said, it is not at the regular visit that the patient comes in with this. The patient comes in telling you this. It wasn't on the screening PET scan.

But we have a 200-patient study looking at this where the physicians are given this prospectively and they are given the data each time. We will see what they tell us. It will also be interesting to see the average number of cycles that they use.

DR. WILLIAMS: I guess the biggest problem in my mind is what about blinding. Can we believe it? How do we know we can believe it. These validations of this and that, they don't seem to be taking into account the placebo effect or the effect of knowing your treatment.

So how do we address that? If can't blind trials, then can we use these endpoints? We have basically moved down to No. 7 and 8 with this

discussion, I think. Can we? I wonder what Dr. Gralla would have to say about that.

DR. GRALLA: So, by "these endpoints," you mean these subjective endpoints, the pain, et cetera?

DR. WILLIAMS: Right.

DR. GRALLA: Let's look. We have talked about 1594, this four-arm lung-cancer trial. Was the patient supposed to feel that they should mark it better because they were getting the docetaxel or the paclitaxel? Most of these trials are in that way.

Now, if the patient is getting the gemcitabine or the or the paclitaxel, my guess is that we could tell which one the patient was getting if we were blinded. So I think that actually maintaining the blind is unlikely and that these are, to me, almost moot points because we are usually looking at Treatment A versus Treatment B. The patient is usually told if we are using the best standard versus a new agent, well, you are getting the very best that we know of.

I don't think that patients answer that their cough or pain is different six, eight, twelve weeks into a study because of this. Now, I think

it is important, such as in the gefitinib study, et cetera, that the patient then being given a pill is given a placebo on the other arm when they maybe are getting nothing in second line. I think that that really is important.

But, in most of these first-line Stage IV patients--and that is the other reason that the normative data will be important, also, to be sure that this is a group.

DR. KEEGAN: Dr. Gralla, I guess having lived through enough of the hype of certain drugs--Herceptin was one, Iressa and Gleevec were others--in a lot of trials, some patients actually are concerned about which arm they are randomized to and do have a strong feeling. Perhaps patients might not be as concerned about being on a certain arm and declaring symptoms as patients who are on the "unfavorable" arm, or what they perceive to be unfavorable, and want to hurry up and declare their symptoms so they can be crossed over. Is that a concern in an unblinded trial, because I think that has been a concern we have had.

DR. GRALLA: I certainly think whenever possible to blind, why not. There is absolutely no reason not to. The are many studies where we

didn't see that being done. However, I must say that, in most of the trials that we have done in the '90s, this really hasn't been where people have been so excited and where they have dropped out in that way.

If you look at the *pemetrexed study that I showed, basically, you can see a lot of patients showing improvement on the cisplatin study, et cetera. There is a strong correlation with response there, on the cisplatin arm, et cetera.

I agree that it is an issue and whenever possible to blind, it is reasonable to do. But maybe the burden of proof is on us to show that your concern actually occurs because it is like the placebo effect, when they looked at it carefully, it was pretty hard to show it was really there.

DR. WILLIAMS: That is kind of our tradition to have the sponsor show that something exists. That is hard to get around.

DR. PRZEPIORKA: Dr. Bonomi.

DR. BONOMI: Just very brief. One other objective thing that could be done in every Stage IV lung-cancer trial is just measure the serial weights. Obviously, people with edema would throw that off. But, otherwise, if I had one thing I

could look at in a patient, just show me their serial weights and pretty much that is going to tell you what is happening to them.

DR. PRZEPIORKA: Is performance status still a valid--

DR. BONOMI: Oh, absolutely but it is--you know, the weights are so--it is a quantitative--one, two is not--Karnofsky is a little bit more detailed.

DR. GRALLA: These are all valuable. But they are not surrogates for quality of life. So they are all valuable. They are components of quality of life. But they are not, by themselves, that. So performance status is really a function scale. It is of real value, what is your ability to do things.

Actually, we like now the patient-generated activity scale where they fill that out. That can be useful. These are all valid points that are very helpful in clinical management. It is pretty hard to see a patient who is losing weight like crazy and think that you are doing something good for that patient.

DR. PRZEPIORKA: Dr. Saxon.

DR. SAXON: Getting back to the original

question which was to choose a magnitude of progression-free survival that one would think would be clinically relevant, it seems to me that the problem with that, and maybe I don't understand this correctly--but the problem with that is that it dissociates that endpoint from the toxicity issue.

Whereas, I think a three-month progression-free survival advantage in a minimally toxic drug may be quite interesting and important, a three-month progression-free-survival advantage with a very highly toxic drug probably wouldn't be. So my own opinion is you can't choose an absolute magnitude that is of clinical relevance, that you have to take into account the toxicity of the agent. So it is going to be a judgment call each time this comes up.

So I guess, in that regard, I disagree with Dr. Johnson, B. Johnson. I don't think it is going to be possible, quite frankly, to choose an absolute magnitude. That consideration is too important, I think.

DR. PRZEPIORKA: Dr. Fleming.

DR. FLEMING: I had voted against use of time to progression as a full reliable endpoint

because of the uncertainties we have talked about. I abstained on the issue of its use as an accelerated approval because I am a bit on the fence. I think we are getting at some very good discussion that I think are the relevant factors that would pull me off the fence one way or the other.

If we are conducting these studies with a high level of rigor that minimizes bias due to unblinding which does concern me, and minimizes missingness, those are issues that certainly are important. I am very favorably persuaded by my colleagues' comments that, if we were relying on time to progression as an accelerated-approval endpoint, it would have to be based on a very substantial evidence of benefit.

I think Scott makes the good point; ultimately, it is benefit to risk. So what that level of benefit is going to have to be will be dependent on what the overall safety profile is. That is certainly relevant although it is helpful to get Dr. Johnson's sense, three months. My own sense here is it should be something very substantial taking into account, of course, the toxicity profile.

We didn't talk about statistical strength of evidence, but it should be strong statistical strength of evidence. Traditionally, we call it strength of evidence of two trials, 0.25 squared, something on that order, something on that order. It should be strong evidence than I might have asked for for survival because, in fact, it is not as reliable a measure.

The study presumably will give us some information on PROs or survival. Certainly seeing some suggestive evidence that those results look to be trending in the right direction, obviously, would be also very importantly reinforcing.

The final point that I would make is a very important issue; is accelerated approval tantamount to full approval and, if it is, then I would argue we should be using criteria close to that for a full approval. But, if accelerated approval really is to get early access while we complete the validation trial in a timely way and, if we have procedures in place that would give us a process to withdraw the accelerated approval if the validation study shows lack of benefit, then I am much more willing to say yes, this lower level of evidence that we would have is, in fact, a basis to

providing an accelerated approval.

So I guess I am saying under all of the conditions that we have talked about, I would also support the accelerated approval. But those conditions mean that we need to have considerable strength of evidence on time to progression. It would be useful to have supportive evidence on survival and it would be important to know that, if the validation study, when completed, showed lack of benefit, that this wasn't going to lead to indefinite access. If it were, then we should be looking at full approval criteria.

DR. PRZEPIORKA: Dr. Johnson.

DR. B. JOHNSON: I wanted to get back to Dr. Williams' point about being concerned about using the PROs and the blinding issue. One of the things that we don't have a lot of examples of in lung cancer is a big dissociation between patient-related symptoms or patient outcomes and what is happening with the underlying disease.

The duration of time is relatively short that we typically see so, until we come up with some examples where there is a moderate dissociation between the patient's perception of outcome and what we typically measure in the

disease, I think it should be okay. It is not something I would lay awake at night worrying about.

DR. PRZEPIORKA: Dr. Temple?

DR. TEMPLE: I guess something I want to flag for a later discussion is the difference between what we usually measure, which is medians or the shape of the curve, and the possibility that there are widely different results from one piece of the patient population to the other; that is, a small responder set.

I don't want to try to resolve that now, but has is always sort of bothered me because I have always been struck by the end of the tail that goes out real far. That seems, in some ways, more important than the median. None of our analyses really reflect that. But I don't want to talk about it now. I just want to flag it for later. Much later.

DR. PRZEPIORKA: In that case, we will move on to the Question No. 6 which we are now getting into dreaded territory. First-line non-small-cell lung-cancer treatment setting approval based on the noninferiority analysis of time to progression or progression-free survival

and/or response rate.

So, specifically addressing the following situation; a less toxic experimental drug demonstrate noninferiority of both response rate and progression-free survival compared to the standard toxic regimen. The standard toxic regimen has previously demonstrated an estimated two-month survival benefit one trial comparing it to best supportive care.

In the current trial data, 95 percent confidence intervals cannot establish whether the experimental therapy retains the survival benefit of the standard regimen. Could approval be based on noninferiority analyses of response rate and/or progression-free survival in situations where the noninferiority analysis of survival cannot be performed.

Examples would be when there are insufficient patient numbers to allow the survival noninferiority analysis or when there is confounding of the survival analysis by crossover.

Discussion? Dr. Fleming?

DR. FLEMING: 5 and 6 are related. They are both noninferiority questions. 5 was on survival, 6 was on surrogate for survival. I am

just wondering, since 5 lays out the fundamental issues that have to be considered for a valid noninferiority trial which also have to be considered in Question 6, is it okay to consider those two questions together, or can we start with 5?

DR. WILLIAMS: I would prefer not to get into the details. Let's suppose that we have everything we need for a noninferiority trial, for time to progression and response rate. I don't want to get into whether we do and how you would do that, but let's suppose we do.

Not a likely situation, but let's suppose. Given that, and given that we can't deal with survival compared to this marginal survival benefit of this other agent, but it is less toxic--I mean, this is a real situation that we definitely will face with several drugs in the near future. The question is can you do noninferiority comparison with response and time to progression.

Certainly, you can do it with response rate. And they are less toxic. So that is the question. I don't want to get into the details of what are the various numbers of trials we have in order to demonstrate the time-to-progression effect

and the response-rate effect. Let's just assume that we have a margin that we can establish and we can establish that we have the same noninferiority rate and time to progression.

I would like to take that as a given, in this question.

DR. TEMPLE: It didn't say noninferiority on the surrogates.

DR. WILLIAMS: Right. Response rate and time to progression.

DR. TEMPLE: But not for the survival, but tolerability advantages.

DR. WILLIAMS: Yes. This is an extremely real example. All of the doublets have very poorly documented survival effects. It is very difficult to do an noninferiority survival analysis. So you have either got to beat them or the other alternative would be to say, I have the same response rate, time to progression with some sort of rigor and that I am less toxic.

So it is sort of a value judgment. You have already said--part of committee said they wouldn't take progression-free survival as a benefit anyway. On that basis, maybe it seems obvious. But the situation may be that you cannot

deal with survival here unless you beat the drug. So I would just like you to kind of struggle with what we are struggling with.

DR. PRZEPIORKA: So, if I can reinterpret the question, if you have a drug that is really not toxic and it gives you the same response rate and time to progression as your current standard which is, come in, get your white count wiped out and have lots of nausea, vomiting and throwing up and, on the basis of numbers, response rate and time to progression are exactly the same for the toxic and nontoxic drugs and there is no way you could look at survival--

DR. WILLIAMS: We have to go a little better than just on the numbers. We would have to satisfy Dr. Fleming they are noninferior.

DR. PRZEPIORKA: But there is no way you could look at survival in those patients because there is just not enough. Would you be willing to recommend approval?

DR. D. JOHNSON: Regular approval.

DR. PRZEPIORKA: Regular approval.

DR. WILLIAMS: Or even accelerated approval. That would be a possibility.

DR. SAXON: But that is not exactly what

this says. What this says is that you cannot establish whether the experimental therapy retains the survival benefit. So the confidence intervals here are overlapping null.

DR. WILLIAMS: Well, no. When we are talking about with respect to survival, you are correct. But we cannot establish it either because we don't have enough data or because the effect is so poorly established historically that it could never be practically done.

DR. TEMPLE: Realistically, if you have a two-month survival, the lower bound for confidence interval is added somewhere less than that, and you want to preserve 50 percent of it, you would have to rule out a loss of half a month or something. The size of study that could do that is not really thinkable.

DR. B. JOHNSON: Can you give us an example of the sizes. The unspoken thing here is that it would take a huge trial to do that with a two-month difference.

DR. WILLIAMS: I would say 2,000 or 3,000. I don't know what the statisticians would say.

DR. B. JOHNSON: Can you give us an idea about the size we are talking about?

DR. FLEMING: It is easier if you go with me for a moment. It is easier to start with the perspective of survival and then move into the perspective of time to progression. But the size, just to jump ahead, of the trial is going to be dependent on what alternative you are presuming.

The way this would frequently be done, if it were survival, for example--let's suppose we have a three-month advantage in survival and it is estimated with considerable precision, plus-or-minus a month. So it is three months, plus or minus a month.

Now, by the way, that clearly is going to be based on a metaanalysis because three months plus-or-minus three months is what you get when you have a p-value that is two-sided 05. So you are talking about very strong evidence to be three months plus-or-minus a month.

Then the typical approach is to say, all right, that means it is at least two months. I will preserve half the benefit so I will have a one-month margin.

DR. TEMPLE: In that case, you could do it.

DR. FLEMING: In that case, it is like the

iridia* Zometa example where this is the exact approach that was used. But, clearly, it takes a metaanalysis. There has to be substantial evidence of some benefit.

However, I would even say here the sample size may not be as horrendous as you would think because, if we are somewhat better, we can rule out we are somewhat worse. There was an noninferiority survival improvement and that was *docetaxel against *navalbine. In essence, the docetaxel median survival was a month longer. You can rule out that you are a month worse when you are a month longer without it being an extraordinary sample size.

Where it becomes extraordinary is if you truly are not any better and then you are having to rule out a small margin. Then it takes a big sample size.

I would hope we would learn from experience, and I think we are learning from experience. The temptation is to say, if I have an effective standard of care and I can come along with something that is less toxic, if the curves are overlapping, if their time-to-progression curves, survival curves, whatever, it is very

tempting to say, come on; efficacy is the same and safety is better.

It brings me back to March 14, 1986 when ODAC was meeting and we were looking at advanced breast cancer with adriamycin as the standard and mitexantrum was being considered and everybody was impressed by the fact that it was less nausea, vomiting, cardiotoxicity, myelosuppression. The committee voted 9 to 2 in favor of approval because there wasn't anything that was compellingly different in survival.

Yet, the fact that the curves are close together doesn't really mean we can rule out that it is worse. Fortunately, Bob Temple and others at the FDA came back and said, let's revisit this in a year. It was revisited in December of '87 and, at that point, the differences were significant favoring the control, now adriamycin, and the committee completely reversed its vote and it was 11-nothing against approval.

The relevance of what we learned fifteen years ago was it is important to understand what levels of rigor we have to have in order to judge that we can rule out that it is meaningfully worse. These margins are not just a statistician's

configuration of something to make clinicians' lives complicated. It does do that, but there is much more of an intention than that, and that is to be able to say, what is the difference between evidence that looks consistent with noninferiority versus evidence that really establishes noninferiority.

For superiority, if you had 30 patients on an arm and you had a two-month survival difference, we wouldn't claim that superiority if the p-value is 0.15. We have to be as rigorous, if not more rigorous, in a noninferiority setting.

So the conclusions that are actually derived and the points that are made in Paragraph 5 for Question 5 are relevant for Point 5 and Point 6. It is very important that we understand that we have active comparators that truly provide substantial benefit that is precisely estimated and where those estimates apply to the setting in which the noninferiority trial is going to be done. That is called the constancy assumption.

A lot of methods are out there. The Rothman method was referred to by Mark Scott in his open-session discussion. I would just point out, that method or any other needs to adjust for the

constancy assumption. Mark Rothman was mentioned that to me also at lunchtime. The method is now frequently being applied when it doesn't adjust for the validity of the constancy assumption which, again, clinically means, historically, I may have estimated my active comparator to have a certain level of effect, but it may not have that level of effect as an imputed placebo in my noninferiority trial if I have different sensitivities for efficacy, if I have different ways of measuring, if I have different supportive care.

The analysis that is being brought before this committee, I hope one question people would ask is, are we using rigorous methods to truly rule out meaningful differences and is that constancy assumption factor being factored in.

Moving to Question 6, we make our life far more complicated when we now try to do a noninferiority analysis on a surrogate endpoint. That is where we are in Question 6. If one is looking at ruling out a certain level of difference in time to progression--let's say you have got these combination regimens that have been established in first-line as standard of care on survival and we now want to look to see whether we

are not meaningfully worse in time to progression.

We are not even saying are we better. We are saying, are we not meaningfully worse. Then what we have to be able to say--I have registered concerns in using time to progression as a superiority because I haven't seen the evidence here presented that indicates that if we achieve a certain difference, beneficial effect in time to progression, that reliably means a treatment-induced effect in survival.

To answer Question 6 positively, you need far more information. You have to be able to know that if you give up a certain fraction of the benefit in time to progression, that will translate into the fraction of survival benefit that you are willing to give up. That type of functional relationship is extraordinarily hard to get at.

We talked about lipids as an example where FDA has used this as an acceptable surrogate. We have myriads of studies showing you can get a 10 percent reduction in cholesterol. It doesn't provide any kind of benefit. But a 30 to 40 percent does provide major benefit.

You have got to understand the functional relationship that says how much time to progression

difference translates into the amount of survival difference I am willing to give up. I would argue that is wishful thinking. That level of insight and the data that we would need to be able to do that just doesn't exist.

DR. PRZEPIORKA: I think a key question here that he brought out was making sure that survival doesn't pay the price. If there is a way that you could keep the confidence intervals--or predict how much you have to keep the confidence intervals down so that you don't lose survival, if you know the correlation between the surrogate and survival, that would be one way to say, okay; it is kind of safe to do this since it is less toxic.

But if you can't predict, I think everybody would have a difficult time knowing the history of the drugs that we have seen in the long run to say yes, this would probably be okay to approve.

DR. TEMPLE: In some ways, probably the example we are more likely to see is where response rates may be a little better, time to progressions may be a little better than the control and we don't really have much data on survival. That would raise an interesting question about

accelerated approval, I think. That is probably more likely to face us.

It is not easy for me to imagine how we would be able to do successful noninferiority on time to progression if we didn't have a clue about survival. I am not sure how you could do that.

DR. PRZEPIORKA: I guess from our earlier discussion that if this little bit better is less than three months, as far as we are concerned, it is not inferiority, it is not superiority.

DR. TEMPLE: Right. Thanks.

DR. WILLIAMS: Perhaps we could go to the last area about symptoms again and have a little bit of discussion. We have heard all of Dr. Gralla's presentation about the merits of these endpoints, but what are we ready for now and how should they be used in the studies we are doing? Do we think that they are ready to be a primary endpoint? Is there a specific area we need to go with these endpoints? Do we need to include them in all the studies?

DR. PRZEPIORKA: We will go ahead and go through the second No. 7 and No. 8. But, before we do that, I just wanted to make a statement of concern that I had regarding the meaning of

validation in these quality-of-life tools that are used since they seem to be validated against other quality-of-life tools.

I work with these patients. I understand their quality of life needs to be good but what is the definition of quality of life. I sit in the chair under a cover and don't move but my pain is better or it is I can take the cover off, fold it up, do some laundry. So I am disappointed to hear that these are not validated against a functional scale which I think would be a meaningful clinical benefit.

DR. GRALLA: I'm sorry, but I think that is incorrect. These are not, but my pain is a little better, I am shivering. How much pain do you have? None at all or as much as it could be? These are validated in ways that are quite clear. If you, certainly, take an example of the FACT-L, there is looking at physical symptoms and how that affects functionality. So these are strongly validated. The Melzack-McGill scale looks at these issues and looks at the quality of pain.

We don't look at the quality of pain. So there is strong correlation with these if you look at how they are looked at. For instance, the

observer scale, as part of the LCSS, correlates with what type of pain medicine you now need. So, have you gone more or less down the WHO ladder or are you just taking tylenol?

So these are validated in ways that correlate with function, et cetera, but the main answer is do they tell you whether a person has pain or not. So a pain questionnaire answers the pain question. They have predictive validity for survival and maybe even for response as well.

But we don't ask that of survival, does this give us a function answer. We don't ask it of response, does it give us a function answer. Now we are asking of pain? I think the validity methods, the gold standards that are used are those that are used elsewhere and that, if you look at the function analysis, especially in the FACT, it really gives you a lot of information as to how people function. And they all correlate with performance status.

DR. PRZEPIORKA: That was not clear in your presentation, but we would certainly like to know more about how the quality-of-life scales predict function. I think that is really important.

DR. GRALLA: Again, if you just look at the validation study for the FACT-L, and it would have taken half an hour to discuss that alone, you can see that it is divided into social functioning, physical functioning, psychological functioning. All these areas are right there. So these address exactly the points that you wish to look at.

DR. PRZEPIORKA: Dr. Bonomi?

DR. BONOMI: In the FACT instruction, they have a thing Dr. Cella calls a Trials Outcome Index. It has 21 questions and it addresses the things that Dick just talked about. It has lung-cancer symptoms. It has functional symptoms. And it does get all of that stuff. In fact, David alluded to it earlier. That was the best predictor of survival in the study, 5592 study. It was better than performance as the initial Trials Outcome Index score was the best predictor.

The problem is, and this I would like to raise, it has 21 questions that the patients have to answer. I think that the things that are probably most valuable are the lung-cancer symptom scale or the FACT-L which is just seven questions about lung-cancer symptoms.

That is something you can get pretty

reliably. You start going to 21, it starts getting a little tougher. But maybe Dick has a comment about that.

DR. GRALLA: I agree. If you want detail--there are always tradeoffs. How much detail do you wish to have? If you will accept the fact that these validity studies that are done and published in the psychometrics that show all these outcomes that you want, and are boring as hell to read, these 20-page papers, or whatever, they go into these issues.

The question is when you get this ready for prime time, you don't want to be doing all those scales that they did because there is correlation with each one of these areas. So Dave Cella has developed this 7-question subscale which some people like, et cetera. The LCSS, which is supposed to address these, has only nine items to be done.

So these get to the questions, is there really pain relief, et cetera. The basis has already been done as to what this means to patients. There is a lot of information on that. It is like looking at a CAT-scan and saying, but how do I know it really works each time. There are

other studies that have shown what it really means, as far as that is concerned.

So I would have to say that if you want to look at the full scale which is what really Phil is talking about, and you take the T, O, I out of that, you get 21 items, et cetera. You can do these. But you can get answers that tell you that patients are improving in the areas that are most important to patients just by using the smaller areas. For the LCSS, it is whole instrument. For the FACT-L and for the EORTC, it is a subscale.

DR. PRZEPIORKA: Dr. Temple?

DR. TEMPLE: One of the areas in which we think we have made progress is we don't call these quality-of-life scales anymore. We call them patient-reported outcomes because quality of life captures--you have got to check the spiritual nature of it all and we are not so such cancer treatment fixes that.

But we think it is at least plausible that it might fix a good scale of lung-cancer symptoms. So the focus there is on those and they have a certain amount of face validity. They seem at least as valid as the typical questions a physician will be put to the patient, like, how is your

breathing or how are you feeling.

They are pretty solid. Those seem like the most promising things. Whether performance in the community for someone with advanced lung cancer is as relevant as how is your breathing these days, I think could be debated. But at least some of them seem very plausible on their face and we would be very happy to see effects on those things, I think.

DR. PRZEPIORKA: I think a question came up earlier regarding performance status II patients and whether or not there should be quality-of-life instruments are PROs as a primary outcome for studies in that subset of patients with lung cancer. Any comments?

DR. D. JOHNSON: You mean as separate studies altogether and is it something that is valid. I think the answer to that is yes. We have data from, again, prospective studies, one from Michael Cullen which I think is a really nice trial that was done in the U.K. in which they included patients with advanced disease who had performance status II, and they did patient-reported-outcome analyses.

What he demonstrated was what ECOG, SWOG

and CALGB and others have demonstrated, that the better your performance status at diagnosis, the greater is your "survival benefit." Again, just to give everybody some baseline data who aren't lung-cancer docs here, if you get a platinum-based therapy and you are Stage IV, your median survival will be nine months if you are 0 performance status, six months if you are 1, and three months if you are 2. That I call my Rule of 3s.

In Cullen's study, he showed really exactly the same thing. It was the exact reverse of that in terms of symptom benefits. Obviously, if you are asymptomatic, you can't get better. You can't get more asymptomatic.

The amount of benefit in terms of symptom improvement was greatest in the patients who were PS-2. So there was a balance. Their survival benefit was not as great. It is one-and-a-half months to two months with no treatment, three months to four months maximally with treatment.

But, by contrast, their improvement, however you chose to define that, was a higher percentage of improvement relative to the PS-1 patients although their survival, the PS-1s, was better than the PS-2s. That makes sense. The more

symptomatic you are, the more likely you are to improve.

DR. GRALLA: Dr. Przepiorka, in the validation studies, Dr. Holland, in Cancer in 1994, looked at "known groups." So we know that survival varies by each decline of the Karnofsky scale. So she looked at very low performance-status group patients, performance status 30 to 50. She found validity for the very low performance-status group, the median, the Karnofsky 50 to 70, and then the better 80 to 100. So part of the validation is looking at known groups and then seeing if this goes true.

This sort of paradoxical finding that David has explained to us seems to exist through that as well. So these instruments have all looked at those groups and these instruments, to some degree, have been looked at in the hospice population as well.

DR. PRZEPIORKA: Are there any other lung-cancer settings where the symptom-based endpoints can then serve as the primary endpoint for approval?

DR. B. JOHNSON: One of the things I would like to address is one of the reasons why--we work

quite a bit with mesothelioma patients. One of the reasons why they generated that--why the symptom scale--and we participated in that study where we assessed that--is that it is very difficult to assess responses in mesotheliomas because it is pleural based and you can't do Recist criteria. So you have either got to come up with a new way of doing it which has since been better validated.

But when those trials started, they didn't really exist. So they embedded that symptom scale in there. We got experience doing it and I agree with Dick. I think that was one of the first times we were really consistent about it and got it short enough so the patients could reproducibly do it.

And so mesothelioma would be a very good one to take a look at. But the thing that happened there is that the symptoms very closely paralleled what they saw radiographically which is what you see in almost every situation.

The other thing that happened that we learned in there is that, and this may be shocking to some people, but they don't always tell the doctor everything. If you took a look at what they filled out, they say, I feel great. Everything is going wonderful. And they have got it all

maximally symptomatic.

So it does collect information, no matter how thorough we try to be, that does not otherwise exist in the medical record.

DR. PRZEPIORKA: That is No. 7. Moving on to No. 8, discuss the role of quality of life as a drug-approval endpoint. Are quality-of-life results meaningful in single-arm studies? I think Dr. Gralla actually addressed that a little, if he wants to reiterate his opinion.

DR. GRALLA: My opinion on this would be that it is very interesting to see it is exploratory, but, for drug approval, I have real difficulty with it.

DR. PRZEPIORKA: Does anyone disagree with that? Okay. We also talked about blinding a little bit so I will skip b. and go to c.; should quality-of-life instruments be routinely included in lung-cancer studies and, if so, which ones.

DR. B. JOHNSON: If it is routine, then why would you have to pick them?

DR. D. JOHNSON: Actually, I am not sure I would mandate that they be included. There are circumstances where, if we are curing 100 percent of the patients and their quality-of-life drops a

little bit, I think they might accept that to some degree. I am being facetious, but I do think that there are circumstances where quality of life is really not going to be necessarily beneficial to the outcome of the trial.

Again, if you are powering for survival benefit, it seems to me redundant to look at the quality of life and then try to come in for a drug approval on the basis of that later on, as a secondary endpoint. Now, maybe FDA would feel differently about that, but, to me, if you want to use it, you should use it in the proper way.

DR. PRZEPIORKA: Dr. Bonomi.

DR. BONOMI: I agree with Dr. Johnson completely. I would not make it mandatory. You would pick it and, if it is your primary objective, great, and make it simple. It has got to be simple, lung-cancer symptom scale or something like that.

DR. PRZEPIORKA: Dr. Saxton.

DR. SAXTON: I agree with Dr. Johnson.

DR. PRZEPIORKA: Dr. Ettinger?

DR. ETTINGER: I agree.

DR. PRZEPIORKA: Do you have other questions you want us to look at?

DR. WILLIAMS: No. I just want to thank everybody for all their input. I think it has been a great discussion. It is a great way to sort of kick off the endpoints process.

DR. PRZEPIORKA: Ms. Ross.

MS. ROSS: Thank you, Madame Chair. Would it be in order for me to make a motion to have a vote on objective response rate as an acceptable endpoint for accelerated approval?

DR. WILLIAMS: We have already used it. The reason we didn't ask is because we already did it with Iressa. I guess you could. The only thing that could happen is that it would turn around that decision which isn't what you want, I don't think.

MS. ROSS: Drop the motion. Okay. Thank you.

DR. PRZEPIORKA: Just as a point of information, our next meeting will be March 4 and it will be one day. It might be one day, it might be two, but it is a different day than originally planned so please check your calendars and this meeting is now adjourned. Thank you.

[Whereupon, at 4:43 p.m., the meeting was adjourned.]

- - -