

## Workshop Summary on Endpoints for Approval of Cancer Drugs for Lung Cancer

### Workshop participants:

Paul Bunn, MD  
Richard Pazdur, MD  
Laurie Burke, PhD, MPH  
Renzo Canetta, MD  
Martin H. Cohen, MD  
Janet Dancey, MD  
Thomas R. Fleming, PhD  
Richard J. Gralla, MD  
David Johnson, MD  
Richard Kaplan, MD  
Patricia Keegan, MD

Gerard T. Kennealey, MD  
Ellen Stovall  
Steve Piantadosi, MD, PhD  
Sheila Ross  
Scott Saxman, MD  
Deborah Y. Kamin, PhD  
Mark Somerfield, PhD  
Mary Lopez Wilson

(This document summarizes discussions among the workshop participants and does not necessarily represent the views of the FDA.CDER)

### Abstract:

The United States Food and Drug Administration (FDA) is embarking upon a process to evaluate endpoints used as the basis for approval of cancer drugs. In a series of public workshops, FDA will seek input from panels of experts on disease-specific endpoints to support drug approval. Issues identified by the workshops will be before the Oncologic Drugs Advisory Committee (ODAC). Subsequently, FDA will develop a series of guidance documents on endpoints for the approval of cancer drugs.

On April 15, 2003, a public workshop was held on endpoints for lung cancer drugs. Expert panelists, selected in consultation with ASCO and NCI included members from FDA, academia (including two statisticians), NCI, industry, and patient advocate organizations. The panel was charged with "identification and discussion of optimal endpoints for demonstrating clinical benefit of cancer drugs used in the management of lung cancer." This article summarizes the workshop discussions.

### 1. Regulatory Background

FDA is responsible for determining that drugs are safe and effective before marketing. FDA approves new drug applications by two different mechanisms, regular approval and (since 1992) accelerated approval (AA). Regular approval is based on evidence of clinical benefit or a surrogate endpoint that *reliably predicts* clinical benefit (e.g., blood pressure as a surrogate endpoint for risk of stroke). AA, which can only be granted when the new drug provides an advantage over available therapy, may rely on a less-established surrogate endpoint, a surrogate that is only *reasonably likely* to predict clinical benefit. After AA, the drug manufacturer is required to perform additional post-

marketing studies to evaluate whether treatment with the new drug provides clinical benefit.

### **1.1 Endpoints supporting cancer drug approvals**

In the early 1980's, upon the advice of the ODAC, FDA determined that objective response rate (ORR) would not generally be an acceptable endpoint for approval. The benefit associated with modest response rates did not necessarily outweigh use of highly toxic cancer drugs. Acceptable endpoints were determined to be survival or improvement in the quality of a patient's life (evaluated at the time by functional assessments or tumor-related symptoms). After promulgation of the 1992 AA regulations, ORR was determined to be a "reasonably likely" surrogate in selected settings to support AA under the 1992 Subpart H regulations.

FDA recently summarized the endpoints supporting 71 new cancer drug approvals over a 13-year period, from 1990 to 2002. Fourteen of these applications received accelerated approval, 18 received regular approval based on a survival improvement, while 39 applications received regular approval based on other direct or indirect evidence of clinical benefit. In some settings clinical inference allowed FDA to accept tumor endpoints as surrogates for symptom benefit even though those benefits were not directly measured (for instance, five drugs were approved for treatment of leukemic disorders based on prolonged complete responses). Occasionally effects on tumors were accepted as surrogates for clinical benefit. For instance, in studies evaluating hormone treatment of metastatic breast cancer, new hormone drugs were compared to tamoxifen, a long-accepted palliative agent. In this setting, with relatively non-toxic drugs and no proven effect of any hormone drug on survival, tumor response, response duration and TTP were accepted as surrogates for comparing new hormone drugs to tamoxifen. ORR and time to progression (TTP) were supportive endpoints for about half of the regular approvals (27 of 57). These 27 approvals were based on ORR alone (10/27), TTP alone (1), ORR plus TTP (7), or ORR plus a favorable effect on tumor-related signs or symptoms (9). Disease free survival in the adjuvant setting also supported approval of five drugs for breast cancer and for bone marrow transplantation for leukemia (Table 1). Morbidity assessments supporting drug approval included evaluation of pain, bone morbidity, cosmetic improvement in cutaneous lesions (Kaposi's sarcoma and cutaneous T-cell lymphoma), and symptomatic relief associated with improvement in pulmonary or esophageal obstruction.

### **1.2 Endpoints supporting approval of drugs for lung cancer**

Drugs approved in the past decade for first-line treatment of NSCLC are navelbine, paclitaxel, gemcitabine, and docetaxel. Docetaxel was also approved for second-line treatment of NSCLC. In most cases, the primary basis of approval was a statistically significant improvement in survival demonstrated in one or two randomized controlled studies. In most instances the survival advantage was 6-8 weeks at the median (Table 2). The first-line docetaxel application was approved based on survival benefit established though a non-inferiority comparison to navelbine. All of the approved first-line regimens

were cisplatin doublets except navelbine, which was also approved for single-agent use. Recently gefitinib received accelerated approval for third-line treatment of NSCLC (after failure of platinum and docetaxel regimens) based on a response rate of 10% and a median response duration of 7 months. Photofrin was approved for local treatment (intra-bronchial photodynamic therapy) of patients with symptomatic obstructing NSCLC. The endpoints supporting approval of Photofrin were the intra-bronchial tumor response rate and assessments on a pulmonary symptom severity scale (evaluating cough, hemoptysis, and dyspnea). Over the past decade, only one drug new drug application has been approved for the treatment of small cell lung cancer (SCLC). Topotecan was approved for second-line treatment based primarily on ORR (considered to represent benefit in that setting) with supportive evidence from pulmonary symptom assessments.

## **2. Lung cancer endpoints**

### **2.1 Objective response rate**

Effects on ORR have been used twice as the basis of lung cancer drug approval, once as a "reasonably likely" surrogate endpoint to support AA in NSCLC refractory to available therapy (gefitinib), and once as a full surrogate to support regular approval of second-line treatment in small cell lung cancer (topotecan). Recently RECIST criteria have provided a widely accepted common standard method for ORR measurement. Effects on ORR have not been demonstrated to reliably predict effects on survival in NSCLC, although some studies with higher ORR rates have reported higher levels of symptom benefit. Inclusion of stable disease in the endpoint definition has been suggested to improve the prediction of survival effects. For regulatory purposes, meaningful evaluation of such an endpoint would need to be done in randomized controlled studies.

### **2.2 Time to Progression**

TTP has often been poorly defined and not rigorously evaluated. In most NSCLC studies submitted for drug approval where a survival benefit was demonstrated, a TTP effect was also noted. However, more importantly, it has not been established that benefit on TTP reliably predicts benefit on survival. In Cancer Cooperative Group studies of NSCLC, median overall survival (OS) is usually about twice the median TTP. Simplified analyses of progression have been suggested, such as percent progression at a single pre-specified time. Retrospective analyses of a Cooperative Group data base suggest that such an progression endpoint may predict effects on OS. In NSCLC, the relationship between surrogates such as TTP and survival have been difficult to evaluate because drugs have produced only minimal effects on survival. With the advent of more effective agents, these relationships need to be re-examined.

### **2.3 Survival**

As discussed earlier, most lung cancer drug approvals have been based on a significant improvement in survival. Survival is an optimal endpoint because measurement is easy and accurate, and the value of a survival improvement is unquestioned. Given the

variation in survival among different study populations, survival must be assessed in randomized controlled studies. The most efficient design evaluates the potential superiority of the new drug regimen compared to a standard control regimen. In "add-on" designs (typically cisplatin plus or minus new drug) the new drug does not have to "beat" a standard drug, it must only demonstrate additional benefit in combination. Because multiple NSCLC treatment doublets are now marketed and are associated with small survival benefits, it may no longer be possible to have a platinum-alone control arm. To date no new 3 drug combination has shown constant superiority over 2 drug combinations in either SCLC or NSCLC and no approval has been applied for using this approach.

## 2.4 Non-inferiority survival studies in NSCLC

In many disease settings, efficacy can be established by showing that a new drug is "non-inferior" to an effective standard drug. As discussed below non-inferiority (NI) studies are difficult to perform in NSCLC at the current time.

One can never prove that two treatments are equal. One can, however, show that a treatment is not worse than a standard treatment by more than some specified acceptable amount (margin). A critical issue in determining this NI margin relates to identifying what treatment effect (TE) can reliably be attributed to the standard drug. The goal of the NI study in a regulatory setting is to compare the new drug to the standard drug and, through inference and statistical methodology, determine the *fraction* of TE that is *demonstrated* to be *retained* (FTEDR). Through clinical and regulatory judgment, one determines whether this is an acceptable fraction (AF) of the TE for the specific clinical setting. Some of the important considerations in this judgment are the level of improvement that the new drug provides in safety, tolerability or convenience relative to the standard treatment.

The field of NI analysis is still developing so that there are ongoing research and discussions about the most reliable and reasonable methods for determining FTEDR and regarding the most appropriate AF for different clinical settings.

NI trials are most readily conducted in settings in which the standard intervention provides a large TE, where the size of this TE has been precisely established in earlier randomized trials, and where these estimates of the TE from earlier trials are unbiased with respect to the effect the comparator regimen will actually have in the active control trial, (with this latter condition often referred to as the "constancy assumption"). In the NSCLC setting, NI designs are difficult because of the small and poorly documented survival benefit associated with the active control treatments (standard first-line chemotherapy). Evidence documenting the TE of currently approved treatments in NSCLC generally consists of one or two trials showing a 1-2 month median survival difference of marginal statistical significance. Taxotere is the only drug approved for second line treatment of NSCLC. The only data available to estimate the taxotere TE is a

104 patient study comparing taxotere to best supportive care. The small size of this study does not allow a precise estimate of treatment effect (HR .56, C.I. 0.35, 0.88).

## **2.5 Disease-free survival**

FDA has long stated that for adjuvant treatment, disease-free survival (DFS) would be an adequate approval endpoint in disease settings where most patients are symptomatic or where effects on DFS is a reliable predictor of effects on survival. Most lung cancer patients are symptomatic when disease recurs, so it seems reasonable for ODAC to discuss whether delay of symptoms is a reasonable basis to justify the use of DFS for lung cancer drug approval, and, if so, whether drug approval based on DFS should be restricted to less toxic therapies. Another approach would be to actually measure the treatment-related and/or tumor-related symptoms using HRQOL scales to determine whether treatment toxicity outweighed DFS benefit. The adequacy of DFS as a survival surrogate in lung cancer cannot be rigorously tested with existing data due to the lack of effective adjuvant treatments. Two small studies studies with a significant or near significant DFS benefit showed similar effects on survival and one study showed a significant survival effect but no effect on DFS.

## **2.6 Patient-reported outcomes (PRO)**

At presentation, over 90% of lung cancer patients with Stage III or IV extent report two or more disease-related symptoms. These commonly include pulmonary effects such as cough and dyspnea, and the general symptoms of fatigue, pain, and anorexia; additionally, patients have high degrees of psychological distress. Consequently, in addition to survival outcomes, information about treatment effects on the patient reported outcomes of health-related quality of life (HRQOL) and symptom benefit is important.

### **2.6.1 Symptom assessments**

To date, patient morbidity assessments used by FDA as direct support for cancer drug approval have consisted of measurements or observations that allowed FDA to infer symptom benefit, such as tumor responses paired with reported improvements in tumor-related symptoms or signs. The perceived advantages of a targeted assessment of symptoms include the potential relative brevity of instruments compared to some multidimensional tools, and the assumption that improvement in measurements could reflect true patient benefit. However, several problems exist for this approach: a) the definition of effective patient benefit is not always agreed upon, b) clinical correlates for changes in symptom measures are not always known, c) if the tool used is not validated, uncertainty in the accuracy of the endpoint could be a problem. Instruments assessing improvements in symptoms require that patients have the symptoms at study entry thus restricting patient eligibility.

## **2.6.2 Health Related Quality of Life Instruments (HRQOL)**

When evaluating the role of HRQOL instruments in the lung cancer drug-approval process, important issues are the instrument's relevance and validity in that setting. Psychometric properties are important to consider. Do the instruments available for evaluating lung cancer meet the well-established criteria for acceptability? Several existing measures applicable to lung cancer patients have undergone extensive psychometric evaluation. Because HRQOL attempts to assess the impact of treatment and disease on multiple dimensions of importance to patients, these instruments vary in length. Typically, quality of life scales are longer than scales that only evaluate single or a few symptoms. This greater length must be considered in the context of the problems of missing data. Additionally, differentiating between cancer-related symptoms, side-effects of treatment and symptoms or problems not related to cancer can often be difficult and perhaps the impact of these can be best assessed using a multi-dimensional instrument.

Regulatory context will affect the appropriate use of instruments. HRQOL instruments have been proposed as primary efficacy endpoints determining whether a new drug is approved, as co-primary endpoints supporting drug approval, or as secondary endpoints to be described in drug labeling or to guide future research. For drug approval, FDA must find that the drug is both safe and effective. Therefore, assessments that reflect primarily a difference in drug toxicity cannot support approval without separate demonstration of effectiveness. PROs that blend together assessments of efficacy and safety may be acceptable as primary endpoints when the comparator drug is relatively non-toxic. When the comparator drug is toxic, however, it may be necessary to separately assess tumor-related PRO benefit and toxicity-related PRO benefit, or to measure effectiveness by non-PRO endpoints.

There are currently three lung cancer-specific instruments with published and acceptable psychometrics in peer-reviewed literature, using previously established criteria. All three are in common use, the EORTC-LC13, the FACT-L, and the LCSS. They share several factors, but also have some differences, which have been discussed in some detail in a recent comprehensive review (Drs. C. Earle and J. Weeks for the NCI sponsored Clinical Outcomes Working Group (COMWOG)). Although these questionnaires differ in the number and format of their scales, all ask patients to measure the impact of lung cancer specific symptoms and treatment related complications on several dimensions of quality of life; all are brief, easy to administer, and have acceptable psychometric properties (feasibility, reliability and validity). They are likely to be able to measure accurately the positive and negative impact of disease and treatment, as expressed by patients, on the various dimensions of quality of life. All three of these instruments have undergone fairly extensive field testing, and have been used in many trials in many countries. In trials using more than one instrument, the different instruments tend to show convergent results. It appears that the EORTC instrument is more frequently used in Europe, while the LCSS and FACT-L predominate in the US.

- EORTC QLQ-LC13. The EORTC QLQ-C30 consists of 30 general cancer-related questions in Likert and numerical analogue scale (NAS) formats, covering the week leading up to its administration. The LC13 adds 13 lung cancer related questions (thus 43 in total). The core instrument combined with the lung cancer subscale is estimated to take about 11 minutes to complete. It has been translated into 23 major languages.
- LCSS. The LCSS is a lung cancer-specific instrument. It concentrates on the symptoms of lung cancer, capturing overall quality of life only by a global question. It does not have a “general cancer” component and does not attempt to assess the toxicity of treatment directly. It consists only of 9 visual analogue scales (VAS) and 6 optional items for an observer to fill out for further context if desired, and asks about HRQOL in the previous 24 hours. Of the three lung cancer-specific HRQOL instruments, the LCSS has the most published literature documenting its psychometric properties.
- FACT-L. The FACT-G (the general component) consists of 34 questions, while the FACT-L (for lung cancer) currently adds 7 questions. The FACT has well-documented content validity. As with the LCSS, it was developed using patient input as well as that of medical professionals for item generation and review. The FACT-L emphasizes social and emotional well-being, enhancing its multidimensional in scope. The FACT may be best in situations where patients are not as ill. It does not have as comprehensive an assessment of symptoms as the other two lung cancer-specific instruments and, therefore, has been most successful in monitoring patients receiving supportive care rather than aggressive anti-cancer treatment.

### **2.6.3 Conclusions regarding PROs in lung cancer**

There is a clear need to evaluate patient-reported outcomes (PRO) in patients with lung cancer. While evaluation difficulties remain, recent trials indicate that initiatives to overcome these problems are meeting with some success. Studies using symptom endpoints and quality of life have been successful in selected indications, but present difficulties. Attention to the areas specified in the companion article (Gralla et al) is necessary if trials are to overcome common problems in PRO evaluation. Education of investigators in the importance and conduct of PRO research is needed prior to the initiation of trials. Patients must understand the study requirements as part of the consenting process. Steps must be taken to ensure that as little data as possible are missing. The endpoints and analysis plan need to be specified prior to the initiation of the trial. Trials need to be properly powered and adequately controlled for the specified endpoints.

Further research in the evaluation and analysis of PROs will enhance this important component of the cancer drug evaluation. Concordant evidence of anti-tumor activity (either as survival data, response rates or as prolongation of TTP) can be desirable. Indeed PROs should be viewed as components of the total value of a treatment, and together with these other cancer endpoints, provide a mutually enhanced picture of the benefits and risks of anticancer therapies.

Table 1: Approved treatment indications based on disease free survival (DFS), 1990-1992

Drug	Indication	Significant findings other than DFS	Year of approval
Anastrozole	Adjuvant therapy of post menopausal breast cancer		2002
Busulfan	Induction therapy for bone marrow transplantation in CML	Time to engraftment	1999
Paclitaxel	Node positive breast cancer	Survival benefit	1999
Epirubicin	Node positive breast cancer	Survival benefit	1999
Tamoxifen	Node negative breast cancer		1990

**Table 2: NSCLC Approved First and Second Line Treatments  
Advanced or Metastatic Disease**

Treatments	# Trials	# of Pts	Endpoint	Result
<b>First-Line</b>				
Navelbine vs. 5-FU/LV	1	211 2:1 rand	Survival	Median surv. 30w vs 22w, p=.06 1 year surv. 24% vs 16% RR 12% vs 3%
Navelbine/cisplat vs cisplat &	2	432	Survival	Median surv. 7.8m vs 6.2m, p=.01 1 year surv. 38% vs 22% RR 19% vs 8%, p<.001
Navelbine/cisplat vs navelbine vs vindesine/cisplat		612	Survival	Median surv. <u>9.2m vs 7.2m</u> vs 7.4m, p=.05 1 yr surv. 35% vs 30% vs 27% RR <u>28% vs 14%</u> vs 15%, p<.001
Gemzar/cisplat vs cisplat &	2	522	Survival	Median surv. 9.0m vs 7.6m, p=.008 TTP 5.2m vs 3.7m, p=.009 RR 26% vs 10%, p<.001
Gemzar/cisplat vs VP16/cisplat		135	Survival	Median surv. 8.7m vs 7.0m, p=.18 TTP 5.0m vs 4.1m, p=.015 RR 33% vs 14%, p<.01
Paclitaxel 135mg/m2 or 200mg/m2/cisplat vs VP16/cisplat	1	599	Survival	Median surv. 9.3m vs 10.0m vs 7.4m, NS TTP 4.3m vs 4.9m, vs 2.7m, p=.05, .08 RR 25% vs 23% vs 12%, p=.001, <.001
Docetaxel/cisplat vs Navelbine/cisplat vs Docetaxel/ carboplat	1	1218	Survival	Median surv. 10.9m vs 10.0m vs 9.1m, NS Efficacy established by a non-inferiority analysis. Docetax/ carboplat did not demonstrate preservation of 50% of the survival effect of Navelbine/cisplat.
<b>Second Line</b>				
Docetaxel vs Best Supportive Care &	2	104	Survival	Median surv. 7.5m vs 4.6m, p=.01 TTP 12.3w vs 7.0w, p<.05 RR 5.5%
Docetaxel vs Investigator choice		248	Survival	Median surv. 5.7m vs 5.6m, NS 1 year surv. 30% vs 20%, p<.05 RR 5.7% vs 0.8%

m=months, NS=nonsignificant, RR= response rate, TTP=time to progression, w=weeks