# ASTRAZENECA ONCOLOGY

To the Chairman:

AstraZeneca Oncology would like to make the following contribution to the ODAC discussions at the December 16th meeting to discuss Endpoints in Cancer Trials generally and specifically those in Non Small Cell Lung Cancer.

## Symptomatic Improvement:

NSCLC is a disease of symptoms. There are well validated scales available for assessing disease related symptoms in NSCLC ( e.g. LCS Scale* ). Demonstration of relief of symptoms, as determined by well conducted, controlled patient-reported outcomes studies, should be acceptable as a sole basis for full approval in advanced NSCLC.

* Cella DF, Bonomi AE, Lloyd SR, Tulsky DS, Kaplan E, Bonomi P. Reliability and validity of the Functional Assessment of Cancer Therapy - Lung (FACT-L) quality of life instrument. Lung Cancer 1995:12:199-220

*Cella D, Eton DT, Fairclough DL, Bonomi P, Heyes AE, Silberman C et al. What is a clinically meaningful change on the Functional Assessment of Lung Cancer - Lung (FACT-L) questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) Study 5592. J Clin Epidemiol 2002;55:285-95

## Trials in Performance Status 2 Patients:

Inclusion/exclusion criteria for many clinical trials in NSCLC exclude PS2 patients both because of their short life expectancy and because many are considered unsuitable for cytotoxic chemotherapy. Novel agents with better tolerability may offer the chance to bring clinical benefit to this ill-served and under studied patient group.

FDA has recently granted fast track status for a compound to be investigated in a trial in PS2 patients*. Does the committee agree that a PS 2 population in advanced NSCLC is an identifiable population worthy of clinical study and for whom an indication can be written?

If not, how would they propose to define this population of patients often considered too unfit to tolerate chemotherapy and therefore excluded from many current clinical trials.?

**XYOTAX(TM) Receives Fast Track Designation for the Treatment of Advanced Non-Small Cell Lung Cancer from FDA**
Fast Track Designation Granted on Basis of Preliminary Anti-Cancer Activity SEATTLE, June 16 /PRNewswire-FirstCall/ -- Cell Therapeutics, Inc. (CTI) (Nasdaq: CTIC) received fast track designation from the U.S. Food and Drug Administration (FDA) for XYOTAX(TM) (CT-2103), its polyglutamate paclitaxel, for the treatment of advanced non-small cell lung cancer (NSCLC) in patients

with a poor performance status (PS2). Fast track designation was granted because NSCLC in PS2 patients is incurable with available therapy offering only modest benefit, and XYOTAX(TM) has the potential to demonstrate improvement over available therapy in these patients based on anticancer activity (tumor shrinkage) in phase I and II clinical trials. "We are extremely pleased that the FDA recognizes that this population of patients has few viable treatment alternatives and that XYOTAX(TM) has the potential to offer improvement over existing therapies," stated James A. Bianco, M.D., President and CEO of CTI. "Fast track designation of XYOTAX(TM) represents a major regulatory milestone in the development of this product candidate." XYOTAX(TM) is currently in two phase III clinical trials for PS2 patients with advanced NSCLC. A new drug application for this indication is targeted for the end of 2004. XYOTAX(TM) is also being studied in a phase III clinical trial among patients with non-small cell lung cancer who have relapsed following a single platinum-containing front-line treatment. According to the American Cancer Society (www.cancer.org), approximately 172,000 new cases of lung cancer are expected during 2003. Non-small cell lung cancer accounts for almost 80 percent of lung cancer cases and PS2 patients make up roughly 20 to 30 percent of the newly diagnosed advanced NSCLC patients. Fast Track Designation Fast track designation means the FDA will facilitate and expedite the development and review of the application for the approval of a new drug if it is intended for the treatment of a serious or life-threatening condition and demonstrates the potential to address an unmet medical need. An expedited review as defined by the FDA user fee performance goals provides for a review within six months.

**The Efficacy Standard:**

We would like to discuss the implications for oncologic drug development of the article by Rothmann et al published in the January 2003 special edition of SIM on non-inferiority trials [1]. The methods described in this article are increasingly used by regulators in the USA and Europe to evaluate the design and analysis of trials of new agents. The consequences for trial size are enormous. Shlaes and Moellering have expressed closely related concerns for anti-infective drug development [2].

There has been something of a paradigm shift in the approach to cancer treatment over recent years. Academia and industry alike are now fully engaged in the discovery, research and development of novel, well tolerated, biologically targeted (cytostatic) 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, non-inferiority trials to compare the new agent with a standard agent, the conventional aim being to show no clinically relevant loss of efficacy.

Such trials are often designed to demonstrate that the new treatment retains some fraction of the established effect of the standard, say at least ½. Note that this fraction is essentially arbitrary and no regulatory guidance currently mandates this as the minimum amount either to demonstrate clinical non-inferiority or to secure regulatory approval. If the standard treatment was previously shown to double survival in a particular disease setting (hazard ratio=0.50, p=0.02, say), and the goal for a new, better tolerated therapy is to retain at least ½ of this effect, a routine sample size calculation shows that a total of 350 events is required to provide 90% power at the 1-sided, 2.5% significance level.

There are several important issues associated with the design and analysis of non-inferiority trials, including 'constancy' – the extent to which the standard treatment performs as it did in previous trials – and 'assay sensitivity' – the ability of a non-inferiority trial to detect a real difference between the treatments compared. Much has been published in this area. The regulatory guidelines ICH E9 and E10 describe the issues in detail and provide some general guidance with respect to trial design and conduct [3, 4].

An issue not addressed in these guidelines arises from the fact that the standard effect is an estimate from earlier work and so is not known with certainty. Sample size calculations often ignore this uncertainty. Hung et al have shown that this approach increases the probability of erroneously accepting the efficacy of a truly inferior drug [5].

The approach offered by Rothmann tackles this issue. Assuming constancy of the effect of the standard and accepting assay sensitivity, Rothmann proposes a formal statistical comparison between the historical data characterising the standard effect and the data arising in the non-inferiority trial, thereby explicitly incorporating the uncertainty (i.e. SE) around the standard effect estimate. This is in fact akin to the putative or virtual placebo comparison approach described by Wang et al [6]. Operationally, Rothmann's approach is equivalent to conventional methodology using *not* the point estimate for the standard effect, but, rather, a lesser effect somewhere between the point estimate and its lower 97.5% confidence limit. This reduced effect is chosen so that the chance of falsely approving an inferior drug is exactly 2.5%, thereby managing the regulatory risk.

The key problem for researchers, physicians and patients alike is that, with Rothmann's approach, there is a dramatic increase in the size of the trial required, often rendering the trial completely infeasible. Applying this methodology to the example above increases the number of events required from 350 to 3082, a near 9-fold increase. This size of the increase derives from a combination of the (arbitrary) effect retention fraction (50% in this example) and the strength of prior characterisation of the standard effect, which is reflected in the (historical) p-value. As illustrated in the table below, the application of this methodology may actually require more events than there are patients with the disease:

Table 1. The number of deaths required to prove a new treatment retains ½ of the effect of standard treatment (HR=0.50) using Rothmann's methodology [true HR new:standard is unity, 90% power, $\alpha$ 2.5% 1-sided]

| (Historical) P-value for Standard v placebo | Upper 95% CI for HR of New to Standard must be less than: | Approx no. deaths required to prove 50% retention |
|---|---|---|
| 0.049 | 1.004 | 3,000,000 |
| 0.02 | 1.12 | 3082 |
| 0.01 | 1.18 | 1563 |
| 0.001 | 1.27 | 735 |
| 0.0001 | 1.31 | 572 |
| 0.00001 | 1.33 | 505 |
| 0.000001 | 1.35 | 459 |
| «0.000001[#] | 1.41 | 350 |

[#]Equivalent to the conventional approach i.e the standard effect is known with (virtually) complete certainty.

This serves to illustrate that even with a highly significant standard effect estimate, p~0.001 say, Rothmann's approach can double the size of a non-inferiority trial. Importantly, if the standard treatment has only just reached statistical significance, this approach implies that no new drug can ever be approved via the non-inferiority route in that group of patients on the basis of clinical benefits other than efficacy; superiority in efficacy to the standard treatment would have to be shown.

Thus, the use of Rothmann's approach, coupled with the arbitrary 50% effect retention requirement, would result in impracticable trials in many settings, notably those where the standard was approved on the basis of a relatively small evidence base. Assuming that direct comparisons with placebo are unethical, we are forced to contemplate non-inferiority trials too large ever to be mounted, effectively removing non-inferiority as a viable tool in the evaluation and ultimate approval of new cancer medicines. This outcome is identical to the one faced by those developing anti-infective agents which has contributed toward a decline in the number of companies investing in antibacterial research and development and, consequently, in the number of new antibiotics to treat serious infections [2]. Some fundamental re-thinking in this area is called for to avoid the obvious adverse impact on the future development of new cancer medicines.

One way forward is to argue that there should be no difference between the standards of evidence required in a superiority setting and in a non-inferiority (or active-control) setting. Hence the first purpose of the non-inferiority trials in question should be to prove that a new agent would have been better than placebo if placebo had been included. The second purpose should be to estimate (indirectly) the size of the effect of the new agent relative to placebo. . Both of these aims are achievable with Rothmanns' approach and assumptions, with some small modifications. In well-conducted trials with hard endpoints, little non-compliance and complete follow up, there should be no need to require 50% retention of effect to demonstrate superiority to placebo. When estimating the size of the effect, attention could focus on the point estimate in the usual manner, and not on the lower confidence limit alone. An approach along these lines has been nicely illustrated by Fisher [7] and does not require pre-specification of a percentage effect retention though, having obtained the data, the result can easily be displayed in relation to the likely fraction of the standard effect retained. Concentrating on estimating the degree of benefit over placebo, albeit through indirect measures, seems more in line with the efficacy standards required by

US law, which calls for substantial evidence of efficacy to be established, with no requirements on relative efficacy with respect to existing agents.

The scientific and statistical debate on how best to draw inferences from active-control, non-inferiority trials should not be considered complete. Rothmann's approach serves to highlight that considerable statistical, methodological and philosophical issues remain. Failure to consider these issues constructively will, at the very least, lead to ever increasing drug development times and, thus, 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 becoming available to patients. We sincerely hope that the scientific community together with regulatory bodies worldwide will give this important area further careful thought.

1. Rothmann M, Li N, Chen G, Chi GYH, Temple R, Tsou H-H. Design and analysis of non-inferiority mortality trials in oncology. *Statistics in Medicine.* 22. 239–264 (2003).

2. Shlaes, DM and Moellering, RC. The United States Food and Drug Administration
and the End of Antibiotics. *Clinical Infectious Diseases,* 34. 420–422, (2002).

3. ICH Topic E9. Note for Guidance on Statistical Principles for Clinical Trials. ICH Technical Coordination, EMEA, London, 1998.

4. ICH Topic E10. Note for Guidance on Choice of Control Group for Clinical Trials. ICH Technical Coordination, EMEA, London, 2000.

5. Hung, J, Wang, S-J, Tsong, Y, Lawrence, J and O'Neil, RT. Some fundamental issues with non-inferiority testing in active controlled trials. *Statistics in Medicine.* 22. 213–225 (2003).

6. Wang, S-J, Hung, J and Tsong, Y. Utility and pitfalls of some statistical methods in active
controlled clinical trials. *Controlled Clinical Trials.* 23. 15–28 (2002)

7. Fisher, LD, Gent, M, Büller, HR. Active-control trials: How would a new agent compare with placebo? A method illustrated with clopidogrel, aspirin, and placebo. *American Heart Journal.* 141. 26-32 (2001).

**Novel Progression Endpoints for Approval:**

Progression free survival (PFS) is an important endpoint in oncologic drug development. Apart from the psychological trauma of knowing that their disease has relapsed or is worsening, in many settings progression will be associated with increasing symptoms, especially pain from new or progressing lesions. It would seem axiomatic therefore that a delay in disease progression or lowering of the risk of progression represents a clinical benefit to the patient in and of itself and should be acceptable as a sole basis of approval.

However, concerns remain with respect to the use of PFS as an endpoint to demonstrate drug effectiveness. Statistically, concerns relate mainly to the fact that a large proportion of oncology trials are necessarily open and the exact time of disease progression is rarely documented. Open trials raise the possibility of bias both in the ascertainment and assessment of radiographic scans. In trials of novel agents, for example, it is possible that patients randomised to the comparator arm may be prematurely declared as having progressed so that they can gain access to the experimental therapy. However, independent radiological review of both patients deemed to have progressed and at least a sample of non-progressing patients would largely address such concerns. With respect to the time of disease progression, events that occur between scheduled clinical assessments are often assigned to the scheduled visit at progression was detected, leading to over estimation of median time to progression. This can also lead to bias in the comparison of treatments if differential assessment schedules have been employed, as may be likely with a novel therapy vs. standard chemotherapy.

AstraZeneca has undertaken an investigation to compare the utility of measuring the proportion of patients with disease progression at a single time-point with the more usual practice of performing a 'Time to Event' analysis. Investigation of the fundamental relationship of the relationship between treatment effect measures over time and measures at a specific time point, including multiple simulations yields the following conclusions.

- The common practice of assigning the true, unknown time of progression to the time of the clinic visit at which it was detected and then analysing via the conventional log rank test tends to bias the treatment effect estimate toward the null and reduce power.
- If a standard interval censored analysis of time to progression has a nominal power of 90%, then a comparison of the percentage of patients progressing at a single time point such as the median follow-up will retain 85-87% power. This loss of power could be regained by longer follow-up.
- The single time point analysis is unbiased.
- The single time-point analysis offers an unbiased alternative when differential visit structures between the arms of a trial would prevent the satisfactory employment of a time to progression analysis. This may be particularly important where a novel agent, which is being dosed continuously is being compared to a cytotoxic which is being given at three-week intervals.
- In many cancer trials the curves do not separate immediately, but there is an initial lag phase where they are super-imposable. This can have a detrimental effect on the power of a study if a time to progression endpoint is being used. In this setting a single time-point analysis of the percentage of progressors actually becomes more powerful than the traditional time to progression analysis.

It is clear from the simulations that, unless disease assessments are frequent, the routine practice of assigning the time of progression to the visit it was detected can introduce a slight bias toward the null and reduce power. If disease assessment schedules are different, the bias increases and the adverse affect on power is greater.

Interval censored analysis is the most powerful, unbiased alternative to a log-rank analysis of PFS, retaining full nominal power under proportionality. However, interval censored analysis demands a common and well adhered disease assessment scheduled. In practice, this is often not realised. The next best alternative is comparison of treatments based on the % alive and free from progression at median follow-up. This approach is also unbiased under proportionality and has reasonable power. Further, this approach is likely to be more powerful than an interval censored approach in the common occurrence of late separation of Kaplan Meier curves.

A single time point binary analysis of progressors offers a simple approach without the complications and issues of time assignment. Such analyses are of reasonable power and can provide for more flexible trials, where patients are managed not by schedule but by clinical practice.

Trials using an comparison of treatments based on the difference in percent alive and free from progression at median follow-up, allow a means of overcoming some of the concerns expressed about the accuracy and possibility of bias involved in estimating time to progression.. The loss in power is modest and can be regained by following for a further one or two disease assessments; simulations should be employed to ascertain the exact period of follow-up required to provide the desired level of power.

In order to avoid any potential for ascertainment bias, all patients who have not previously progressed should undergo a disease assessment at the end of minimum follow-up. Potential for investigator bias in the reading of radiological scans can be addressed by using a fully independent radiological review panel.

A subsidiary, binary analysis based on a count of all progression events should be undertaken in order to reassure that time based analyses are unbiased.

Given the above findings, demonstration of a reduction in the proportion of patients who have either died or have disease progression at a single time-point should be acceptable as a basis for approval.

Acceptance that a reduction in the risk of progression is a clinical benefit would mean that this endpoint could be appropriately used as the sole basis for a full approval.

Brent M. Vose, Ph.D.
Vice President and Head
Oncology and Infection Therapeutic Area
AstraZeneca Pharmaceuticals LP

George Blackledge, M.D.
Vice President and Medical Director
Oncology and Infection Therapeutic Area
AstraZeneca Pharmaceuticals LP