More power to trials for non-small-cell lung cancer

The median survival times for locally advanced non-small-cell lung cancer (NSCLC) are only 8 to 10 months with 2-year survival rates ranging from 10% to 20% [1–4]. Despite the use of modern radiotherapy techniques, these figures have remained essentially unchanged for the past two decades [5], resulting in an overall 5-year survival of only 3%–6% in randomized trials. There is evidence from a recent meta-analysis that radiotherapy combined with cisplatin-based chemotherapy may improve the survival in NSCLC when compared to patients receiving radiotherapy alone even though the survival benefit is very modest indeed [6].

With respect to the level of statistical significance, varying results have been reported. The most recent meta-analysis, cited above, revealed that the combination of chemotherapy containing cisplatin plus radiotherapy was associated with a hazard ratio of 0.87 (13% reduction in death risk; absolute benefit of 4% at two years) compared to patients receiving radiotherapy alone (p = 0.005) [6]. The relatively small number of randomized patients may explain the varying levels of significance obtained in the individual studies included in this meta-analysis, resulting in lack of statistical power in the comparison of the different treatments.

Determination of the appropriate number of patients to be included is probably the most critical statistical issue in the design of a clinical study. The optimum sample size is the minimum number of subjects necessary to achieve the study aims, minimizing the number of patients subjected to the risks associated with unproved therapies. In this issue of *Annals of Oncology*, Planting et al., on behalf of the EORTC Lung Cancer Cooperative Group, report on a randomized phase II study of high-dose, split-course radiotherapy preceded by high-dose chemotherapy versus high-dose radiotherapy alone in locally advanced NSCLC [7]. The final objective with respect to numbers was to include 192 patients to achieve an 80% chance of detecting a true increase of 50% in the median survival rate (alpha = 0.05 – two-sided test) using a stepwise design with an interim analysis after inclusion of 80 patients to evaluate feasibility. Unfortunately, the study had to be closed before fulfillment of even the latter goal because of the insufficient patient accrual of only 74 patients, in spite of the fact that the study was open in a 3-year period for patients from at least 2 European countries with thousands of patients being candidates for such a study. The overall response rates were 44% among 34 patients receiving the combined treatment and 60% among 30 patients receiving radiotherapy alone. This difference was not statistically significant and neither were the differences in median survival or percentage of patients surviving 2 years. However, the number of patients actually entered into the study enabled only a power of 0.8 to detect a doubling of the median survival associated with the addition of chemotherapy and less than 0.5 power to detect a true increase of about 50%. This illustrates some of the questions at issue with respect to the design of randomized phase II trials, extending into phase III trials.

One of the major objectives of a randomized phase II trial design in NSCLC, with standard treatment as control, is to facilitate the interpretation of a poor response rate to an investigative treatment [8]. If the investigative regimen yields an observed response rate of, for example, more than 40%, the treatment would be identified as having antitumor activity regardless of the magnitude of the response rate to the standard treatment. In contrast, if the investigative treatment yields a low response rate, for example, less than 20%, the definite interpretation concerning activity will depend on the response rate observed in the group receiving standard treatment. If the latter response rate is sufficiently large and the sample size adequate, the conclusion is that the investigative treatment is inactive. If the response rates to both the investigative treatment and the standard treatment are poor, the trial is inconclusive due to inappropriate patient selection. The purpose of randomization in this setting is not to determine whether the investigative treatment is better or worse than the standard treatment.

The use of randomized phase II trials in a setting such as described above, i.e., with a standard treatment arm for comparison, does not seem to offer a broadly useful approach to classical phase II testing [8]. The design is potentially useful in circumstances where an adequate response rate to the standard treatment is not assured, but the identification of a standard treatment may be difficult and is currently not universally agreed upon in locally advanced NSCLC. However, randomized phase II trials without a control arm offers a rapid method for evaluation of new treatments. Such studies may be performed with sample sizes conventionally used for non-randomized phase II studies.

Another important issue is the determination of the sample size sufficient to obtain clinically relevant differences between an investigative treatment and a standard regimen. A statistical significance based on p-values may not always be indicative of clinical significance and thereby herald therapeutical progress. The sample size is based on how small a difference one
wants to be able to detect, how certain we want to be of detecting a difference of that size, if it exists, and the level of significance of the test. The significance level is conventionally 0.05, whereas the power may vary, usually from 0.8 to 0.9 corresponding to a type II error of 20% or 10%, respectively. If we anticipate that radiotherapy alone may result in a 50% response rate and that addition of chemotherapy may increase the response rate with a minimum of 25% (to a 75% response rate) approximately 60 patients would be needed in each treatment arm in order to achieve a significance level of 0.05 with a power of 0.8. Smaller anticipated differences between the two regimens would require a higher number of patients, whereas a lower number of patients diminishes the power of the comparison, i.e., increases the risk of overlooking clinically significant benefits. Similar calculations are essential for comparison of survival rates.

Given these statistical problems, it is important to define carefully which clinical studies are the most relevant and how to organize these among various institutions at an international level in order to assure an adequate number of patients for reaching conclusive results. This may result in fewer studies being performed, but if those made are based on firmer ground, this may eventually lead to the progress in the treatment of NSCLC which is still so desperately needed.

J. B. Sørensen & H. H. Hansen  
Department of Oncology  
Finsen Center  
National University Hospital  
Copenhagen, Denmark

References