Design of clinical studies: Adaptive randomization and progression-free survival (PFS) as an endpoint in clinical studies of advanced non-small cell lung cancer (NSCLC)

Daniel Schrimpf¹, Christian Manegold² and Lothar R. Pilz²

¹Deutsches Krebsforschungszentrum (DKFZ), Department of Biostatistics, and ²University of Heidelberg, Medical Faculty Mannheim, Heidelberg, Germany

Introduction

Significant progress has been achieved recently in our biological understanding of advanced NSCLC and its tumor heterogeneity. The identification of novel tumor targets with different pathways and the search for specific antitumor agents requires new designs of randomized clinical trials. Individualized therapies based on predictive factors such as gene identification will help to stratify further the patient population in question. Even in multicenter studies the available sample sizes will be smaller for these study concepts. The appropriate choice of study endpoints and test procedures to enable an early decision to be made with as few patients as possible is crucial.

Methods

The classical endpoint overall survival (OS) in advanced NSCLC may explain the direct clinical effect of treatment. Efforts have been taken to substitute OS by non-classical endpoints like PFS [1]. Though OS is still accepted as the gold standard, PFS allows a more precise detection and attribution to the effects of the investigational treatment without being compromised by subsequent treatments. The concepts deal with the hidden information about OS in PFS and post-progression survival (PPS). In a meta-analysis with 69 studies [2] a high correlation between PPS and OS, and a moderate correlation between PFS and OS were found.

Simulations showed [3] that PPS is important to understand treatment effects especially if the post-progression period is short (~ 12 months). This knowledge can be used in connection with an adaptive randomization procedure in studies for advanced NSCLC. Application of the surrogate endpoint PFS and adaption of the patient’s allocation process has the aim to assign more patients to the probably better treatment. As adaptive method a customized Bayesian randomization process shown by Berry and Wathen [4] is used. The simulations for some realizations of the allocation process in our paper are carried out in the open source web-application RANDI2 [5].

Results

The process starts with a learning phase of 30 equally distributed randomizations between a new treatment and a control, followed by the adaptive process which changes the allocation probability of patients to treatments in function of the current available data. In our case the main parameter is PFS time and some additional influencing factors. Regulated is the process by the probability \( P_2 > P_1 \) that the success under the new treatment \( P_2 \) is higher than the success probability \( P_1 \) of the other treatment. The probabilities are calculated according to a method introduced originally by Thompson [6] and depend inter alia on the current success rates and the number of patients of the two treatments. In addition some influencing factors can be considered. Favorite candidates for supporting PFS as additional measures are quality of life (QoL) measures and treatment toxicity (TT). The calculation of probabilities is repeated for every new patient and \( P_2 > P_1 \) is used to assign the in-
coming patient to one of the treatments. A further refinement is the possibility to stop the randomization at a high or low value of $P_{p_2 > p_1}$. A high value means that the new treatment is much more successful than the other one. Accordingly, low values indicate the failure of the new treatment. The adaptation of the allocation process with and without a stop criterion was simulated with the aid of RANDI2.

Assumed is that for each individual OS is exponentially distributed. In the simulations of two treatment arms presented here the success of a treatment is defined by a PFS greater or equal to 90 days and a covariate influence of a positive QoL defined as a minimum of 120 days of “good life” after progression. OS is defined as the sum of PFS and post progression survival (PPS). Therefore, OS in case of success is correlated to the length of the time interval of PPS. If the PFS and QoL statements are answered positively (in the sense of success) new patients are assigned more often to the better treatment and, vice versa, in the negative case more rarely. Two cases were considered: the new treatment has a median OS of (i) 12 months and (ii) 16 months. In both cases a median OS of 8 months of the standard (control) treatment is assumed. A sample size calculation for a strict non-adapted situation of 8 vs. 12 months OS reveals a patient number of 354 using a two-sided log-rank test with a power of 90% and an $\alpha = 0.05$. Assumption of accrual time was 24 months with 140 events in the control arm and 118 events in the new treatment arm (PASS 11, version 11.0.7). This was set as the fixed basic number of patients to be recruited in the study. In this simulation part we were interested in the proportion of patients in the experimental and the control arm. Each simulation was performed with 1,000 runs, at first without a stop criterion, giving an approximated mean for the resulting patient figures per treatment arm. The result in case (i) was a population of 239 (68%) patients in the mean for the new treatment arm and 115 (32%) patients in the control arm. This should be compared to 177 in each arm (1 : 1 randomization) in case of a non-adapted study. In case (ii) the simulation results in a higher mean of 260 (73%) patients for the new treatment arm and some less 94 (27%) patients in the controls if a total of 354 patients as before were to be recruited. If in the second case the sample size for a non-adapted study is recalculated (OS of 8 months versus 16) this will end up with 66 patients per arm with 52 events in the new treatment arm and 38 events in the controls (all other assumptions as in case (i)). Simulations in this setting result in a mean of 89 (67%) for the new treatment and 43 (33%) for the controls.

In a second section of simulations the stop criterion $P_{p_2 > p_1} = \Pr(p_2 > p_1)$ is used, giving the probability that the treatment is better than the control with the numerical expression

$$P_{p_2 > p_1} = \sum_{i=0}^{n_i} \left( \frac{n_i + r_i - \alpha}{n_i} \right) \times \left( \frac{s_i + s_2 + 1 + \alpha}{s_i} \right) \left( \frac{n_i + n_2 + 2}{n_i + 1} \right)$$

where $i = 1, 2$ are the treatment arms, $n_i$ – patients per treatment, $r_i$ – successes per treatment, and $s_i = n_i - r_i$. The following cases will be looked at $P_{p_2 > p_1} \geq 0.85/0.9/0.95$. Provided that the total patient number was assumed to be 354 in the non-adaptive study, the randomization ends in case (i) under the Thompson stopping rule with a total mean of 234, 306, and 343 patients for the two treatment arms, respectively. Thus, if the stop criterion $P_{p_2 > p_1} \geq 0.95$ in the simulation example with a median OS of 12 months is used, typically 3% (343 instead of 354) fewer patients will be enrolled in the study.

With a basic number of 132 patients of a non-adaptive study in case (ii) 86, 109, and 124 patients are assigned to the two treat-
ments on the mean using the Thompson stopping of 0.85/0.9/0.95, respectively (Figure 1).

A high enough value of the stop criterion prevents an early end of the study, minimizes the risk of under powering the study with the consequence of not reliably analyzable data.

If the stop criterion $P_{p_>\geq p_\geq 0.9}$ in the simulation example with a median OS of 16 months is used then 17% fewer patients (in comparison with the basic number of 132 patients) will be enrolled in the study using PFS and the covariate QoL without making any assumptions about the power of the study. In case of using the stopping rule at a level of 95% the gain will only be 6% showing that smaller sample numbers can only be derived with a lower stopping limit. In this numerical experiment the 90% stopping translates into a time benefit of about 20% since only the recruiting time of the study will be shortened.

The consideration of OS as the sum of PFS and PPS and with using the sum success criterion in PFS suggests applying PFS as a surrogate for OS, with the inherent considerable advantage that subpopulations with longer PFS intervals can be detected more early. Taking into account the used risk-benefit profiles of treatments, the concept can be considered for decision making. If accompanied by some independent measures like QoL or TT, PFS should be able to cover the clinical benefit achieved by treatment. The implementation of the adaptive procedure and the simulations can be reasonably realized in RANDI2. A covariate adjusted procedure within this system can be used to shorten the development-time and number of patients involved to test new treatments.

Conclusions

NSCLC studies for individualized therapies with elaborated patient selection processes can be designed in a more flexible manner in implementing adaptive randomization. PFS in addition with some measures like QoL and/or TT is a good candidate to adapt the randomization process. Since recruiting time matters design has also to be adopted in this factor. Randomization adaption performed with a web-based system like RANDI2 gives the chance of an earlier decision making with ethically preferable fewer patient numbers. In typical population samples the simulation shows clearly an imbalance of patient distribution in favor of the superior treatment and its subgroups.

References


