Clinical research in medical oncology is an essential key element for the transformation of new developments of preclinical and clinical research into the daily practice. With the high number of newly developed anticancer drugs, in particular new classes of compounds and treatments aiming at molecular targets, the role of properly designed and performed clinical trials is evident. Beyond several important structural and in particular ethical questions as discussed in this issue of ONKOLOGIE, there are two issues which are the key for a successful clinical trial: the quality and strategic importance of the scientific question of the study and the design and statistical structure of the trial. Given the large number of clinical trials required to address the rising number of new scientific questions and the limited number of patients available for clinical trials, the sample size is becoming a critical issue and is therefore under continued discussion. The article of Edler and Kopp-Schneider in this issue discusses several issues associated with planning of clinical trials in oncology and gives somewhat provocative considerations.

Because the topic of clinical trial design, sample size, and statistical considerations is highly important in medical oncology, this paper is accompanied by editorials from two experts: Dr. Punt, an expert in the field of medical oncology and clinical trials and Dr. Silvester, an expert in statistics in medical oncology, who is the chairman of the Statistical Department in the EORTC.

H.-J. Schmoll, Halle

Statistical Trial Designs and Clinical Practice: Are They Compatible?

In this issue, Edler and Kopp-Schneider try to unmask some myths around statistical principles of randomized clinical studies, and they present their case in a clear and straightforward manner [1]. Comments from a statistical viewpoint are presented elsewhere [2]; in this editorial comments are presented from a clinical viewpoint.

Firstly, the dogma that two independent randomized studies showing comparable favorable results are needed is criticized. Although in theory a single study may be sufficient indeed, the danger of a selection bias (which is inherent to all studies since any inclusion implies a selection) may render its results less applicable for the general population. Obviously, in case of known or unknown prognostic variables, this may be overcome by appropriate stratification or the inclusion of sufficient patient numbers, respectively. However, it is not unusual that prognostic criteria as defined within the trial are not considered in daily practice. For instance, when a phase III stage IV melanoma trial with a novel cytotoxic drug is open for all patients with M1a–c disease, but the majority of participating centers exclude M1a–b patients since they have a separate vaccine trial for this category, a negative result will usually be the end of this experimental cytotoxic in melanoma despite the fact that this was only shown for patients with the worst prognosis (M1c). Of course the statistics are not to blame for this, but performing a second trial in another part of the world may decrease such bias due to differences in clinical practice and therefore inclusion policies.

Also, with more effective drugs becoming available in several tumor types, the effect of a single novel drug will be more difficult to evaluate when overall survival is the primary end-point. In the case of stage IV colorectal cancer, positive results have been shown for irinotecan and oxaliplatin as first-line treatment, while the use of subsequent effective treatments was either unknown or the novel effective drug was not available for subsequent treatment in the control arm. This aspect does not invalidate the efficacy of the drug under investigation in any way, but it does invalidate the common interpretation that the novel drug has outright superiority in first-line treatment, while the use of subsequent effective treatments was either unknown or the novel effective drug was not available for subsequent treatment in the control arm. This aspect does not invalidate the efficacy of the drug under investigation in any way, but it does invalidate the common interpretation that the novel drug has outright superiority in first line.

The inclusion of more centers from different countries in a different trial may decrease this potential bias. Secondly, the authors state that it is not necessary to include all available patients into one trial, and that there are no statistical reasons against performing two competing trials within a single institution. This may, however, create the selection bias as discussed above in the case of melanoma. Although this should not invalidate the results of either trial, this again...
underscores that results should only be of consequence for the actual study population, which may differ from the population as defined by the inclusion criteria of the study, and that patients should be stratified by center.

Thirdly, the authors argue against the restriction to large centers for participation. However, the argument that they fail to present here is that centers, in which only a small number of eligible patients is available, will invariably have a) much less experience in treating this patient category, and b) will acquire much less experience with the drug under investigation during the course of the study. This implies that the quality of treatment may be lower in poorly accruing centers, which may be quite relevant when more complicated procedures are involved (e.g., total mesorectal excision (TME) surgery for rectal cancer), or drugs with more complicated toxicity patterns are investigated (i.e., high-dose methotrexate or irinotecan). Indeed, some studies were shown to have better outcomes when centers accruing only one or two patients were excluded [3], and the quality of procedures in centers that treat large number of patients has been shown to be superior to centers that treat low number of patients [4].

Lastly, the authors use two randomized studies on edrecolomab adjuvant therapy, by Riethmüller et al. with 166 patients, and [5] and by Punt et al. with 2,761 patients [6], as an example that a large study does not necessarily have to be better than a small study. A single study with only 166 patients may show a statistically significant difference, however, most oncologists would feel uneasy to incorporate the results of such a small-scale study in their daily practice, given the higher likelihood of bias. In the case of the Riethmüller study [5], the results are further confounded by the use of the Zelen randomisation method, which requires consent only for patients randomized to the treatment arm, as well as by the inclusion of both colon and rectal cancer. On the other hand, the optimum patient numbers for such evidence has never been defined. In any case, the example of the two edrecolomab studies for this argument is not appropriate, since these studies were designed for different questions (i.e., edrecolomab vs. observation and edrecolomab with/without chemotherapy vs. chemotherapy). The comment that the two trials on edrecolomab are not contradictory has already been made [6, 7]. There is no doubt that statistics form the backbone of any randomized trial, and no result of such trials should be interpreted without a careful evaluation of the statistical design. Yet, sometimes logistic problems do not allow an optimal design (as may be the case in rare tumor types which do not allow for a sufficient number of patients), and even in the most optimal design the statistics leave us with no more (and no less) than the probability that a given result is true. It is due to more humane aspects that when a survival curve shows a large difference, some colleagues tend to show less interest in the statistical design, while in fact we should have less interest in the survival curves when the design is suboptimal. Clearly, most of the comments made above concern the interpretation of results rather than the statistical design itself. In many clinical trials, the problems (if any) do not originate at the level of statistical design, but start when the results are being implemented in clinical practice. It would be helpful if a higher level of consensus would be achieved among oncologists on the requirements any new treatment should meet before being accepted as the standard of care. It is quite remarkable how differently some drugs are being used for a single tumor type among countries even within the European Union – especially when we take into account that we all base our decisions on the same study results. Given the great variety in the incidence of different tumor types, it is obvious that no general guidelines can be given. However, if greater care is given to learning more precisely what the result of a given study is telling us for a given patient population, much harm can be prevented. As clinical oncologists we owe our patients the optimal medical care, and this includes to protect them from the exposure of treatments that have become available as a result of overoptimistic interpretation of study results.

Finally, Edler and Kopp-Schneider state that the randomized trial was born halfway in the 20th century. This does not appear to be completely true, since already 250 years ago James Lind, a naval surgeon on the HMS Salisbury, randomized 12 patients from sickbay in 6 cohorts of 2 in order to find an effective treatment for scurvy [8]. The two patients who were given a diet of lemons and oranges quickly recovered, and although it took another 50 years before the availability of fresh fruit was made mandatory in the navy, this shows that it may only take 12 patients in a non-blinded, non-placebo controlled trial to set a new standard. In medical oncology we just need a drug with comparable efficacy to that of a lemon in case of scurvy.

C. J. A. Punt, Nijmegen

References