Looking for efficiency rather than efficacy in randomized controlled trials in oncology

Randomized controlled trials (RCTs) are generally recognized as the gold standard for developing therapeutic agents and gaining definite data on their efficacy. When transferring the results of RCTs into practice, clinicians should consider the quality of the study design, and of the conduction, the extent of generalization of the results and the clinical relevance of the treatment effect. Considering the risk and cost–benefit profile of therapeutic agents when deciding on treatment of a patient has become even more compelling with the increasing cost of new drugs.

Although evaluating research results is a matter of ethical, scientific and public importance, systematic assessment of clinical relevance is still lacking. In oncology, previous reviews have reported that the numbers and sizes of RCTs have increased over time and the design has improved [1], although important deficiencies remain, particularly in assessments of the clinical impact of the results.

A recent analysis [2] of how often experimental cancer treatments result in successful new interventions showed that ~25%–50% of new cancer treatments prove successful. Interventions judged to be so beneficial that they should immediately become the new standard of care, or that had an effect size so large that they halved the death rate, or more, are reported to be discovered in 15% of trials.

However, considering the pooled results, which serve to compare the real effect of new and standard treatments in terms of patient outcomes, the hazard ratios (HRs) for overall survival (OS) and event-free survival were only 0.95 [99% confidence interval (99% CI) 0.93–0.98] and 0.90 [99% CI 0.87–0.93], respectively, suggesting that the majority of new treatments produced scant clinical benefit.

In a further analysis [3] of trends in methodology, sponsorship and outcomes of RCTs, based on >300 studies in patients with non-small-cell lung cancer (NSCLC), breast cancer and colorectal cancer and published from 1975 to 2004, Booth et al. found important changes over time. In terms of sponsorship, the more recent RCTs were more likely to be sponsored by industry, for-profit/mixed sponsorship increasing from 4% to 57% during the study period. In terms of design, the number and size of RCTs rose considerably and there was more frequent use of time-to-event measures (from 39% to 78%) and less use of response rates (from 54% to 14%) as primary end point. In terms of endorsement of results, authors of recent RCTs were more likely to accept novel therapies, even when the treatment effect remained the same size over the period considered. This might be because investigators are more likely to endorse even small clinical effects if obtained in studies of good quality; nevertheless, conflict of interest affecting the interpretation of results cannot be excluded since the majority of trials have a for-profit sponsor. A further possible explanation stems from the frequent difficulty in discerning clinical relevance from statistical significance.

The gap between statistical significance of results and perception of their clinical importance is confirmed by evidence from pivotal cancer trials. In a retrospective analysis of the influence of the tumor KRAS mutation status on the response to cetuximab in metastatic colorectal cancer, Van Cutsem et al. concluded that the addition of cetuximab to the standard chemotherapy combination of 5-fluorouracil and irinotecan (FOLFIRI) reduced the risk of progression, mainly in wild-type KRAS tumors [4]. However, although the KRAS mutation seemed to exclude the use of antibodies against the epidermal growth factor receptor, in patients with KRAS wild-type tumors cetuximab gave modest clinical benefit, with an increase of median progression-free survival (PFS) of <2 months and no effect on OS. Thus, the cost–benefit ratio for the KRAS status of the tumor can be considered negative, merely increasing the burden of cost for the public health system, with only a small benefit for patients, if any.

When there is a worthwhile effect on survival, the lack of thorough evaluation of the risk–benefit profile is a common pitfall of publications. As an example, in a report on chemotherapy plus bevacizumab for NSCLC in comparison to chemotherapy alone [5], Sandler et al. showed that the addition of bevacizumab to paclitaxel and carboplatin gave a significant survival benefit (HR 0.79, P = 0.003), even if there was an increase in the risk of treatment-related deaths. Clinically significant bleeding was observed in, respectively, 4.4% and 0.7% of patients treated with bevacizumab plus chemotherapy and chemotherapy alone. The authors did not present and discuss the results in terms of the risk–benefit profile, using appropriate summary statistics, such as the number needed to treat.

However, it can be calculated that 12 patients needed to be treated with the combination of bevacizumab and chemotherapy to prevent one death at 1 year, at the price of an excess of one death from toxic effects for every 24 patients treated. The balance between benefit and risk is clearly questionable, but nevertheless the use of bevacizumab was endorsed in this setting.

In breast cancer, the benefit of bevacizumab is even less impressive. In combination with paclitaxel, bevacizumab was reported to prolong PFS significantly more than paclitaxel alone (median PFS 11.8 versus 5.9 months; HR for progression 0.60; P < 0.001) [6]. But the benefit in PFS not only did not translate into a significant increase in OS, but...
there was no improvement in quality of life, as suggested by the lack of appreciable changes in the scores on several validated scales. Even so, the Food and Drug Administration approved bevacizumab for the treatment of metastatic breast cancer.

In pancreatic cancer, the addition of erlotinib to gemcitabine improved the median OS by only 10 days (6.24 versus 5.91 months) [7]. Objective response rates did not substantially differ and patients receiving both drugs more frequently experienced rash, diarrhea, infection and stomatitis. In renal cell carcinoma, an OS advantage for sorafenib was only demonstrable on comparing sorafenib-treated patients with those receiving placebo, who did not cross over to receive sorafenib, who might form a subgroup with a poorer prognosis. Moreover, the survival advantage was obtained at a substantial cost in terms of both toxicity and expense [8].

These examples show that establishing the minimum clinical benefit needed to adopt a therapy as the new standard of treatment and the measures to be used for its quantification poses a major challenge for the oncology community. This obviously calls for a multidisciplinary approach by methodologists, investigators, scientific societies and regulatory authorities.

From a methodological point of view, an important issue contributing to biased information regarding the clinical importance of study results is the tendency to rely mainly on statistical significance and on relative measures of treatment effect, rather than on absolute therapy benefits. Biased interpretation due to selective information has been a topic of debate for many years [9–12]. In 1994, Bobbio showed how different ways of reporting results affected physicians’ recommendations and that the method of reporting trial results and the completeness of information in case of controversial results affected physicians’ willingness to prescribe. The willingness to prescribe was closely related to the presentation of results using measures of relative effect. Despite repeated publication of Consolidated Standards of Reporting Trials recommendations [13], still only 5% of authors express their findings in terms of absolute benefit (i.e. absolute risk reduction and/or number needed to treat) in RCTs with statistically significant treatment effects published in major journals [14].

Unlike dichotomous end points for which it is straightforward to calculate the absolute measures of effect, for time-to-event end points it is more troublesome. They are usually measured at a specific point (5-year OS or median survival), thus ignoring the results for the whole period considered. As an alternative, absolute benefit can be established as a difference between the areas under time-to-event curves up to a given time; this uses all the data from the time-to-event curves and might be more informative than the usual method [15]. Seruga et al. [16] investigated trends in absolute benefit in terms of time-to-event end points, using the ‘classical’ difference in absolute benefit at specific time points for the adjuvant setting, the differences in median time for advanced disease and the difference in area between the two curves up to a predetermined time for all studies. The area used was a percentage of the area for the entire graph up to the same time; this adjustment permitted comparison of absolute benefits determined by the area method in different RCTs. This is analogous to assessment of absolute benefit by the standard method, while estimates by both methods do not depend on the outcome in the control arm. The results showed that median absolute benefits of experimental adjuvant treatments decreased over time, while outcomes in control arms improved. For RCTs evaluating metastatic disease, there were no changes in absolute benefit over time, but incremental monthly costs of new approved treatments rose 100-fold. Analysis also confirmed that the authors’ endorsement of experimental medical therapies did not change much over time in the adjuvant setting, while in RCTs evaluating treatment of metastatic disease, endorsement of experimental therapies rose significantly with time even though there was no gain in the absolute benefit.

It thus seems essential to improve methodology in order to close the perception gap between statistical significance and clinical relevance. Recent work has helped define clinically important results. Sobrero and Bruzzi [17] suggested paradigm changing with different increments in the time-to-event end point, according to the prognosis of the disease, as a condition of approval for advanced disease where the median survival time (MST) is <1 year (e.g. pancreatic cancer, gastric cancer or NSCLC), a paradigm changing agent should give at least a 50% increment in MST or 2-year survival rates. For diseases where the MST is in the order of ≥2 years (e.g. breast, colorectal and ovarian cancer), a 30% increment in MST or in 2- to 3-year survival rates should also be considered paradigm changing. Agent giving significant results but not reaching the clinical cut-off for difference might be considered only to provide proof of therapeutic principle. They would not be licensed for sale but proposed for further trials in molecularly selected patient populations as part of new drug combinations with other incremental agents or in the setting of adjuvant therapy.

Agreeing to raise the clinical bar for decision implies adopting a different statistical framework. According to standard statistical methodology, the design of an RCT should ensure that the trial has an adequate probability (i.e. statistical power) to reject the hypothesis of no difference between treatment arms (H0, the null hypothesis) at the prespecified level of significance (usually 0.05) if the null hypothesis is false and the alternative hypothesis H1 is true. Accordingly, results are usually presented as statistically positive or negative depending on whether they do (P < 0.05) or do not (P > 0.05) allow one to reject H0. If a new treatment is to be introduced into clinical practice, it should not be enough to show that it is ‘better’ than the standard therapy, regardless of the size of its effect. Instead, it should be necessary to demonstrate that the effect is clinically worthwhile, meaning as large as or larger than a specified threshold representing the minimal clinically worthwhile effect.

It is perhaps worth recalling here the role and responsibility of regulatory agencies, cancer researchers and professional societies in defining standards of care. Regulatory agencies should be more selective in their approval process when examining trials demonstrating no survival advantage or prolonging survival only by a few months. Since acceptance and place of presentation have considerable weight in the spread of information, at national and international meetings, which are
an important vehicle for publicizing recent results, scientific societies should give less space to study results with marginal effect and require that a thorough discussion of risk–benefit profile should be conducted.

A treatment with marginal benefit should not be immediately declared a new standard and other results should merit debate in plenary sessions. According to Fojo and Grady [18], a minimum standard for presentation should be required. The benefit the study wants to achieve, the measures used to quantify this benefit and its impact in application should be clearly discussed and displaying the results in terms of absolute rather than relative difference should be encouraged. Societies also need to back oncologists when they decide that the marginal benefit is not worth the cost of treatment of a given patient or a group of patients.

Instrument such guidelines should be produced emphasizing the quality of the data and providing recommendations based on clinical relevance of results. The approach of GRADE methodology is a good example of how this approach could help medical decisions [19]. Many organizations and clinical societies, with the important exception of cancer societies, have endorsed the GRADE approach and an emerging consensus is being observed [20]. In conclusion, for many new therapeutic agents the survival gain is so little to be almost of no clinical interest. To be considered as a true ‘therapeutic breakthrough’, and to justify the dramatic raise in the costs, new drugs should produce relevant improvement in clinical outcome and improve the risk–benefit profile. Methods to assess clinical relevance are available but are not implemented in the key situations where consensus is formed and adoption is decided. It is time that oncologic community shoulder his responsibilities for changing this trend.

I. Floriani*, S. Garattini & V. Torri
Laboratory of Clinical Trials, Department of Oncology, Mario Negri Institution, Milan, Italy (*E-mail: irene.floriani@marionegri.it).

references
15. Ajani JA. The area between the curves gets no respect: is it because of the median madness? J Clin Oncol 2008; 25: 5531.