Changing methods to assess targeted therapies in oncology

Trevor Stanbury[1], **Xavier Paoletti**[2], and **Bernard Asselain**[3]

1 UNICANCER, Paris, France
2 OncoStat, Gustave Roussy, Villejuif, France
3 Université Paris Sud et Comité statistique de ARCAGY-GINECO, Orsay, France

Correspondence to:

Trevor Stanbury
UNICANCER
101 rue de Tolbiac
Paris, France 75654
+33 (0)1 44 23 55 67
t-stanbury@unicancer.fr

With permission from the journal and authors, this article was adapted and translated from the original French article: Asselain B, Paoletti X. Enjeux méthodologiques de la médecine de précision en oncologie. *Innov Ther Oncol.* 2017;3:33 –8. doi: 10.1684/ ito.2017.0071.

Abstract

New methods have been developed to evaluate targeted therapies, since the classic sequence – phase I, toxicity; phase II, efficacy; phase III, comparison with standard treatment – is no longer effective for evaluating these new treatments. In traditional cytotoxic chemotherapy trials, we observe a positive correlation between dose toxicity and dose efficacy. In targeted therapy trials, however, high doses can sometimes be well tolerated and increasing the dose beyond a certain level does not increase tumour response. Early clinical trials in targeted therapies therefore need to simultaneously assess toxicity and provide early signals of efficacy, based on biomarkers when available. Phase II primary endpoints have also been questioned, since the RECIST (Response Evaluation Criteria in Solid Tumour) is not well suited to functional modifications in tumours. New phase III trials, with more homogeneous targeted populations, are using more flexible designs, including interim analyses and adaptive designs. These flexible designs allow the sample size, and sometimes the trial design, to be modified during the trial. This article discusses these new methodological challenges for evaluating targeted therapies.

Targeted therapies, unlike traditional cytotoxic chemotherapies, block specific pathways/ mechanisms by which tumours grow and/or inhibit our immune system from responding. These therapies can target cancer cells or our immune cells (immunotherapies). Gefitinib and cetuximab, as examples, block the epidermal growth factor receptor (EGFR) signalling pathway on tumour cells. This interferes with the tumour signals that result in tumour growth, proliferation, and migration. Similarly, immunotherapies, such as pembrolizumab, bind to lymphocytes, interfering with the programmed cell death 1 (PD-1)/ programmed death-

ligand 1 (PD-L1) checkpoint signalling pathway, preventing tumour cells from deactivating the immune response.

The arrival of targeted therapies, including immunotherapies, in oncology has required a re-evaluation of the classical methods used over the last 50 years. The classical sequence consists of identification of the maximum tolerated dose (MTD) and the recommended dose during phase I, selection of the most promising molecules based on the activity (response rate) in phase II, and comparison with standard treatment in phase III. This sequence was developed for cytotoxic chemotherapies and is not optimal for targeted therapies.

In this article, we examine the changes adopted in each of the clinical study phases (I-III) and discuss these new methods for evaluating targeted therapies.

Phase I studies

The aim of a classical phase I studies, to assess cytotoxic chemotherapies, is to determine the MTD and a recommended phase II dose, and obtain an initial safety profile. These classical phase I dose escalation studies assume that an increase in the treatment dose will increases toxicity and activity (Figure 1).

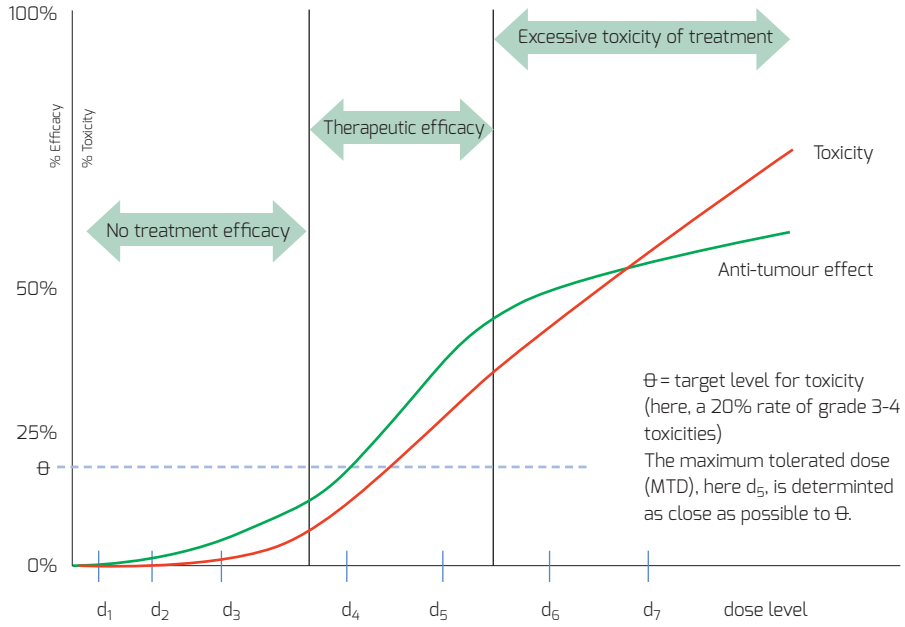

Figure 1. The relationship between dose-toxicity and effective dose in classical phase I clinical trials

This assumption allows us to identify a "therapeutic area" with acceptable toxicity and probable activity. We can identify this "therapeutic area" using different methods,[1] including:

- The "3+3" method that targets a MTD with a toxicity of 33% (i.e. with two patients out of six having a treatment-related side-effect);
- The continuous reassessment method that is more flexible. In this method, we establish a target toxicity level, usually 25% to 30%, before the study.

With the MTD identified, the recommended phase II dose is established, often corresponding to the dose level just below the MTD.

The following issues arise when we use this classical phase I methodology to evaluate targeted therapies. As mentioned, the targeted therapy dose does not always correlate with toxicity and activity. Because of the low toxicity of some targeted therapies, in about 25% of phase I targeted therapy studies, the MTDs are never reached despite using high dose levels.[2] The toxicity that does occur can also be relatively independent of dose. Concerning activity, once the drug's target is saturated, an increase in dose will not usually increase activity. The objective of a phase I targeted therapy study is to identify a biological active dose with minimum toxicity and not the MTD as for classical phase I studies.

Identifying the biologically active dose instead of the MTD at first seems interesting, but remains theoretical when the activity on the target is difficult to measure, particularly when biomarkers to measure the saturation of the drug target are not available. When biomarkers are available, their levels do not always correlate with a clinical benefit for patients, suggesting a more complex mode of action than expected. In addition, the use of biomarkers often requires repetitive tumour sampling, which may not be acceptable. Currently, less-invasive or non-invasive methods to monitor tumour evolution are being developed. These include functional imaging to assess tumour perfusion, heterogeneity, and texture, and to quantify angiogenesis (the development of blood vessels for tumours), and liquid biopsies (for example, blood sampling) to assess circulating tumour cells. This research may allow us to better evaluate drug activity independently of toxicity.

However, phase I targeted therapy studies in oncology simultaneously evaluate tolerance/ toxicity and activity.

Classical phase I studies for cytotoxic chemotherapies were not designed to assess activity. However, phase I targeted therapy studies in oncology simultaneously evaluate tolerance/ toxicity and activity. The targeted therapy studies have fewer dose levels with more patients at each dose level, compared to classical phase I studies. However, phase I targeted therapy studies need to limit the number of patients to a few tens and limit the number and size of extension cohorts. These cohorts, evalu-

ating targeted therapies, can reach hundreds, sometimes more than a thousand, patients without any *a priori* decision rules established.[3] Furthermore, the sample sizes used in these phase I expansion cohorts are not always justified.[4] A review of 522 phase I studies performed at the Dana-Farber/Harvard Cancer Center showed that 60% of studies with three or more expansion cohorts had response/activity as an objective without any justification of the sample size.[5] These phase I targeted therapy expansion cohorts should be designed with the same statistical rigour as classical phase II studies.

Phase II studies

Phase II studies aim to establish whether a drug at the biologically active dose has clinical efficacy with sufficient tolerance to continue to phase III. The phase II studies of cytotoxic chemotherapies are often single-arm studies that assess response rates over a short time period. To measure treatment activity, we often use the change in the dimensions of the tumour lesions, using imaging (CT scan or MRI scan) assessed by response criteria. These response criteria have evolved over the last 20 years, from the World Health Organization (WHO) classification to the response criteria in solid tumours (RECIST) classification version 1.0, recently upgraded to version 1.1.

These criteria, based on tumour dimensions, are not ideal for evaluating targeted therapies. For example, with inhibitors of angiogenesis, the size of the targeted lesions can be unchanged, while the density and texture can change substantially, particularly due to intra-tumour necrosis.

In studies of new immunotherapies that inhibit immune checkpoints (PD-1/PD-L1) we can see an initial temporary increase in tumour size (pseudo-progression), most probably due to lymphocyte infiltration and tumour swelling.

Thus to evaluate tumour response with targeted therapies, we need to revise the established classifications. The new propositions, e.g., Choi criteria for the tyrosine kinase inhibitors, ir-RECIST and i-RECIST for immunotherapies, do not have an international consensus, and need validation before they can be widely used.

Most phase II oncology studies assess treatment activity based on the response rate (best response obtained or response at a given certain time). The hypotheses are generated using an already established response rate at a given time. The Simon and Fleming trial designs define a minimum response rate below which the treatment will be considered not effective (null hypothesis) and a targeted response rate (alternative hypothesis) that indicates sufficient activity to progress to phase III studies.

The Simon method was adapted by Bryant and Day[6] to simultaneously account for efficacy and toxicity. In the Bryant and Day method, the treatment is considered not of interest if the response is inadequate or if the toxicity is excessive.[7] The Bryant and Day method can be used to evaluate certain targeted therapies, e.g. immunotherapies, as well as combinations of targeted therapies or targeted therapies associated with cytotoxic chemotherapies and/or radiotherapy.

In phase II oncology clinical studies, progression-free survival (PFS) has become the preferred endpoint since the delay for analysis is substantially shorter than for overall survival (OS) and the interference by "salvage" therapies is limited.[7] In cancers where the patients' life expectancies are short, i.e. OS is very short, and/or salvage therapies are ineffective, OS may be the most appropriate endpoint. A single arm design can be considered when the endpoint either based on PFS or OS at a given time point, has a reliable historical control.

However, phase II studies evaluate targeted therapies in patients that will potentially respond to treatment. The patient selection is not only based on disease characteristics (clinical stage and histological type), but also on the tumour's molecular profile and the presence of biomarkers, if these biomarkers are predictive of response or are suspected to be.

In these studies with a highly selected population, it is very difficult to have a precise reference or historical control for response (e.g., PFS rate). Therefore, a control group is needed[8] to validate the hypotheses and assess treatment activity. However, we cannot directly compare the control and experimental groups since the statistical power is insufficient, due to the limited number of patients in phase II studies.

Sometimes, comparative phase II randomised studies based on phase III methodology are proposed. These studies accept a high false positive risk (alpha-risk), often in the range of 20%, but sometimes even 40%, to reduce the number of patients required. Accepting this risk means accepting that two out of five significant differences obtained will be by chance! However, when the treatment is extremely active, this strategy may prove to be more efficient. Marketing authorisation can be granted without doing phase III studies, which are long and expensive.[9] However, there is a large risk of obtaining not significant and unconvincing results, due to the relatively small sample size that lacks statistical power, which may stop the drug development of an active treatment.

Phase II studies, even on a limited number of patients, may identify biomarkers that could predict response, even if this research would be exploratory at this stage of drug development. These exploratory studies will facilitate drug development by increasing our understanding of the underlying biology of targeted therapies.

Phase III studies

Phase III studies are essential to compare targeted therapies to standard treatment. Targeted therapies can be evaluated as monotherapies, in combination with other targeted therapies, or in association with standard treatments. The comparison with standard treatment is recommended in these studies, while comparing different targeted therapies or different dosages of the same targeted therapy without a control arm is not recommended. The control arm also allows us to evaluate biomarkers that may predict activity.

> The comparison with standard treatment is recommended in phase III targeted therapy studies.

When a biomarker is known, or suspected, and the biomarker status available at randomisation, the study should be stratified according to the biomarker. However, this is rarely the case. No biomarker (EGFR or KRAS) analyses were planned in the studies assessing gefitinib in lung cancer[10] and cetuximab in colorectal cancer (the CRISTAL and PETTAC8 studies).[11,12] Actions were taken retrospectively or during the study on a portion of the population, probably not representative. In these studies, the benefits of randomisation were probably lost and the results difficult to interpret.

Although requiring more patients than in phase II studies, phase III targeted therapy studies can be limited to a few hundred patients. The clinical gain, for instance the decrease in the

risk of progression or relapse, observed with targeted therapies is far superior to that observed with chemotherapy. We could predict a reduced risk of progression or relapse of 40% (hazard ratio [HR] of 0.6), and even 60% (HR of 0.4). This is an important parameter for calculating the number of patients required. These gains can be even more important considering the highly selected population, for example selecting patients with a specific tumour mutation to evaluate the corresponding targeted therapy. Traditionally, phase III studies evaluated treatments in broader populations, however, phase III targeted therapy studies are often performed in a biologically homogenous subpopulation. In studies comparing targeted therapies to placebo, an unequal randomisation can be used, for example including two-thirds of the patients in the experimental arm and the remaining third in the placebo arm (randomisation 2:1). This minimises the number of patients exposed to the placebo. At equal power, overall about 10% more patients are needed in these studies, but with about 10% to 20% fewer patients receiving placebo. This unequal randomisation allows us to more precisely evaluate the toxicity and efficacy of the targeted treatment in the experimental arm.[8]

These phase III studies will initially concern patients with advanced staged disease, as with classical phase III chemotherapy studies, and the new therapies will have to prove efficacy at the advanced and non-advanced disease stage before being evaluated on patients with a better prognosis – in the adjuvant or neoadjuvant setting.

Phase III studies, like phase II studies, often use the PFS as the primary endpoint to show the advantage of the treatment in delaying disease progression or by stabilising the disease. However, a gain in PFS does not always correlate to gain in overall survival, due to interference caused by salvage treatment after the failure of the experimental treatment.

Furthermore, with targeted therapies we cannot be certain that the results in patients with metastatic disease will extrapolate to those at a less advanced disease stage, e.g., after surgery (adjuvant treatment), as was the case in the studies for the antiangiogenic molecules in localised colorectal cancer. In phase III targeted therapy studies, it is important that the intermediate analyses of efficacy and futility be done according to strict rules, under the control of independent committees of experts. This may eventually allow us to reach an early conclusion with publication of the results, either positive or negative.

What methodology is appropriate for evaluating targeted therapies?

The early phases

We have seen that the methodology of the early phase (I and II) have been questioned: A joint evaluation of the tolerance and the efficacy has become necessary as soon as the optimal dose is established.

Despite toxicity being relatively independent of the localisation and histological type of the tumour, we cannot dissociate these disease characteristics from treatment efficacy. We therefore need to treat a sufficient number of patients with the same type of tumour at an early phase.

We could consider adaptive methodologies, where we randomise patients in a number of treatment arms (with different dose levels) with a control arm, in which we equilibrate the types of tumours in each arm to have an initial idea of the treatment efficacy. The intermediate analyses will allow the selection of one or more of the experimental arms based on the tolerance and biological criteria of efficacy. Patient inclusions could continue in the two or three arms showing the most promise and eventually in the tumour types that seem to be most sensitive. These extension cohorts of a reasonable size will allow a decision on whether or not to proceed to a randomised phase II study with a more "robust" criterion, such as the PFS.

Adaptive phase III methods

The intermediate analyses, evoked for the phase III studies, are a first step towards more flexible methods. The new adaptive methods will allow modification during studies.

Adaptive randomisation allows us to modify the treatment allocation ratio based on intermediate results during the study. Thus, if the treatment administered in Arm B of a study proves to be more active than that of Arm A during an intermediate analysis, the study could begin randomising more patients in Arm B than A. However, even though this method appears to be promising, it is controversial among statisticians since these modifications may extend the study duration and introduce biases.[8]

These studies continually select a population of patients for which the treatment may have greater efficacy. For example, in a study of an immune checkpoint inhibitor, we could decide to include only patients with a strong PD-L1 expression in the stroma, or we could reinforce this subpopulation in calculating a specific power. In this situation, we could use a procedure of sequential testing (closed testing procedure), where we rank the statistical tests to be done in a hierarchical way (e.g., we only test the effect of the treatment in the enriched subpopulation only if the benefit of the treatment is globally significant). These methods could be based on biomarkers that we strongly suspect to be related to the efficacy of the treatment. However, there are a number of cases where we do not have these associated biomarkers. We could also propose a "therapeutic test" by treating all the patients with the targeted therapy, but only randomise patients who respond or who are stable in a second phase. The comparison of the responders and non-responders in the initial phase may allow the identification of new biomarkers.

If allowed for in the protocol at study design, the number of patients required could be adjusted depending on the intermediate results, reviewed by a committee of independent experts. The use of the estimation of treatment effect in the intermediate analysis most frequently leads to an increase in the number of patients to be included.

The phase II-III study design is another option, which allows initiation of a randomised phase II study that can be extended to a phase III, if the phase II results are positive. The patients included in the phase II would be included in the phase III analysis.

Finally, certain trials go beyond all the traditional classifications of cancers by localisation and histological type, using molecular anomalies to classify patients and to propose for each patient a targeted therapy that is most adapted to their profile. These pilot studies, like the SHIVA study,[13,14] pose new methodological challenges, raising such questions as how many targeted treatments, what combination of strategies, what stopping rules, and how to introduce new treatment arms? Moreover, how can we analyse these data to be able to extract knowledge that we can have confidence in?

Conclusion

The development of "precision oncology" based on known specific molecular anomalies of tumours and the corresponding therapies targeting these anomalies has caused substantial modification of the methodologies that were developed to evaluate cytotoxic chemotherapies. We can no longer base the evaluation of targeted therapies on the parallel between dose-toxicity and dose-efficacy. Early phase trials need to be adapted to simultaneously evaluate tolerance and initial efficacy of therapies. In addition, some targeted therapies are so well tolerated that the MTD was never reached in phase I studies.

The first-in-human studies are approaching phase II studies, with fewer dose levels but including more patients per level. Extension cohorts established for the most promising dose levels, give an idea of the tolerance and efficacy of difference cancer localisations. These extension cohorts need however to remain at a reasonable size, including 10 to 20 patients, with clear statistical decision rules. These early phases need to rapidly establish the therapeutic dose and provide initial efficacy information. The randomised phase II studies will evaluate the degree of treatment activity and allow us to design smaller phase III studies. The flexible methodology used in randomised phase III studies allow for the re-evaluation of the initial hypotheses and modification of the sample size and inclusion criteria to select patients more likely to benefit from the treatment during the study. The intermediate analyses for futility allow the early termination of studies that have an extremely small chance of showing treatment efficacy. However, if these more flexible methods are now permitted, the study conception and the application rules must be clearly defined in the protocol at study design.

It is only by respecting a strict methodology, based on early randomised studies, starting from phase II studies, that we can optimise the investigational methodology to evaluate the large number of targeted therapies tested, with their associations, so that each patient can benefit from the treatment best adapted to the biological profile of their tumour.

Disclaimers

The opinions expressed in this article are the authors' own and not necessarily shared by their employers or EMWA.

Conflicts of interest

The authors have no conflicts of interest to disclose.

References

1. Hansen AR, Graham DM, Pond GR, Siu LL. Phase 1 trial design: is 3 + 3 the best? Cancer Control. 2014;21(3):200–8.
2. Postel-Vinay S, Gomez-Roca C, Molife LR, Anghan B, Levy A, Judson I, et al. Phase I trials of molecularly targeted agents: should we pay more attention to late toxicities? J Clin Oncol. 2011;29(13):1728–35.
3. Robert C, Ribas A, Wolchok JD, Hodi FS, Hamid O, Kefford R, et al. Anti-programmed-death-receptor-1 treatment with pembrolizumab in ipilimumab-refractory advanced melanoma: a randomised dose-comparison cohort of a phase 1 trial. Lancet. 2014;384(9948):1109–17.
4. Menis J, Litiere S, Tryfonidis K, Golfinopoulos V. The European Organization for Research and Treatment of Cancer perspective on designing clinical trials with immune therapeutics. Ann Transl Med. 2016;4(14):267.
5. Dahlberg SE, Shapiro GI, Clark JW, Johnson BE. Evaluation of statistical designs in phase I expansion cohorts: the Dana-Farber/Harvard Cancer Center experience. J Natl Cancer Inst. 2014;106(7).
6. Bryant J, Day R. Incorporating toxicity considerations into the design of two-stage phase II clinical trials. Biometrics. 1995;51(4):1372–83.
7. Rubinstein L. Phase II design: history and evolution. Chin Clin Oncol. 2014;3(4):48.
8. Saad ED, Paoletti X, Burzykowski T, Buyse M. Precision medicine needs randomized clinical trials. Nat Rev Clin Oncol. 2017;14(5):317–23.
9. Ledermann J, Harter P, Gourley C, Friedlander M, Vergote I, Rustin G, et al. Olaparib maintenance therapy in platinum-sensitive relapsed ovarian cancer. N Engl J Med. 2012;366(15):1382–92.
10. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med. 2009;361(10):947–57.
11. Van Cutsem E, Lenz HJ, Kohne CH, Heinemann V, Tejpar S, Melezinek I, et al. Fluorouracil, leucovorin, and irinotecan plus cetuximab treatment and RAS mutations in colorectal cancer. J Clin Oncol. 2015;33(7):692–700.
12. Taieb J, Tabernero J, Mini E, Subtil F, Folprecht G, Van Laethem JL, et al. Oxaliplatin, fluorouracil, and leucovorin with or without cetuximab in patients with resected stage III colon cancer (PETACC-8): an open-label, randomised phase 3 trial. Lancet Oncol. 2014;15(8):862–73.
13. Paoletti X, Asselain B, Kamal M, Servant N, Huppe P, Bieche I, et al. Design and statistical principles of the SHIVA trial. Chin Clin Oncol. 2015;4(3):32.
14. Le Tourneau C, Kamal M, Tredan O, Delord JP, Campone M, Goncalves A, et al. Designs and challenges for personalized medicine studies in oncology: focus on the SHIVA trial. Target Oncol. 2012;7(4):253–65.

Author information

Trevor Stanbury, PhD, is a medical writer at UNICANCER, based in Paris, specialising in writing for oncology clinical studies, including protocols, clinical study reports, and publications. He has a background in organic chemistry. For the last 11 years he has been working in clinical studies and exclusively in oncology for the last 5 years

Bernard Asselain, MD, PhD, is the former head of Biostatistics department at the Institut Curie in Paris and is now working in the field of ovarian cancers in the GINECO group in Paris. He has been involved in the methodology of oncology clinical trials for more than three decades.

Xavier Paoletti, PhD, is a senior biostatistician, working in the Oncostat Department at Gustave Roussy Institute in Villejuif. He is involved in the field of early phase clinical trials in oncology and has published many papers on phase I trials. He is also involved in the field of meta-analyses and he recently organised a symposium on adaptive methods in clinical trials.