available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Position Paper

Design and conduct of phase II studies of targeted anticancer therapy: Recommendations from the task force on methodology for the development of innovative cancer therapies (MDICT)

Christopher M. Booth[a,1], A. Hilary Calvert[b], Giuseppe Giaccone[c], Marinus W. Lobbezoo[d], Elizabeth A. Eisenhauer[a], Lesley K. Seymour[a,], On behalf of the Task Force on Methodology for the Development of Innovative Cancer Therapies*

[a]*National Cancer Institute of Canada Clinical Trials Group, Queen's University, 10 Stuart Street, Kingston, ON, Canada K7L 3N6*
[b]*Northern Institute for Cancer Research, Medical School, Newcastle upon Tyne, UK*
[c]*Medical Oncology Branch, National Cancer Institute, Bethesda, USA*
[d]*NDDO Research Foundation, Amsterdam, The Netherlands*

ARTICLE INFO

Article history:
Received 26 July 2007
Accepted 31 July 2007
Available online 12 September 2007

Keywords:
Clinical trials/phase II
Drug therapy
Neoplasms
Receptors
Growth factors/antagonists
Inhibitors

ABSTRACT

The Methodology for the Development of Innovative Cancer Therapies (MDICT) task force considered aspects of the design and conduct of phase II studies for molecular targeted agents during their 2007 meeting. The task force recommended that multinomial endpoints and designs should be considered for phase II studies of targeted agents, that both single arm as well as randomised designs remain appropriate in certain settings, and that further assessment of novel endpoints (tumour growth kinetic assessment, biomarker or functional imaging) and designs (randomised discontinuation or Bayesian adaptive design) be encouraged. The MDICT cautioned on the use of small randomised trials which have a number of statistical pitfalls and dangers and strongly encouraged the complete reporting, including negative trials, in the scientific literature.

1. Background

The past decade has seen a significant shift in oncology drug development from the evaluation of traditional cytotoxic agents to molecular targeted anticancer therapies. While cytotoxic agents generally act at the level of cell division, novel molecular agents target specific proteins involved in tumour growth, angiogenesis, and/or metastases and may be selectively active. In addition, it has been suggested that these novel agents may not lead to tumour shrinkage. Classical drug development strategies have therefore been challenged, especially with regards to the utility of single arm phase II studies with objective response as the primary endpoint.

* *Corresponding author:* Tel.: +1 613 533 6430; fax: +1 613 533 2941.
E-mail address: lseymour@ctg.queensu.ca (L.K. Seymour).
[1] Research Fellow, NCIC Clinical Trials Group.
0959-8049/$ - see front matter
doi:10.1016/j.ejca.2007.07.031

The 'Methodology for the Development of Innovative Cancer Therapies' (MDICT) task force was established by the NDDO Research Foundation in 2006 to provide practical guidance on the development of anticancer targeted agents. The task force mission and membership has been described in a companion publication.[1] The second annual meeting of the MDICT task force was held on 7 March, 2007, in conjunction with the 5th International Symposium on Targeted Anticancer Therapies in Amsterdam. Participants included 16 experts from academic centres in eight countries, together with observers from industry and regulatory agencies. The mandate of the 2007 meeting was to review current knowledge and discuss and make recommendations regarding appropriate designs for phase II studies of targeted anticancer agents. We report here on the meeting as well as on the task force recommendations.

2. Scientific review and discussion

A review of the experience to date with molecular agents was presented. The design, endpoints and methodology of phase II studies for a cohort of 31 targeted drugs in six solid tumour sites was reviewed, based on published literature and meeting abstracts. Sixty-five phase II reports (89 trials) of 19 targeted agents were included (unpublished data, personal communication Dr. E. Eisenhauer). As a marker of eventual 'success' of the agents, the authors identified which drugs subsequently received US Food and Drug Administration (FDA) approval. Objective response was the primary (or co-primary) endpoint in the majority of studies (61 trials, 69%). Randomised designs were not common (14 of 65 reports; 22%) and generally involved different doses of the targeted agents rather than comparison with placebo or observation. Enrichment of the patient population for target expression was uncommon (18 studies, 20%). Evidence of objective response was seen in 38 studies (43%) although response rates were modest (3–28%). Agents with objective responses tended to have higher rates of non-progression with renal cell cancer being an exception to this observation. Higher response rates were found to be predictive of subsequent FDA approval ($p = 0.03$). Few studies correlated clinical with biological/molecular outcomes.

The MDICT task force then explored critical issues in the contemporary design of phase II studies for targeted anti-cancer therapy, addressing each in a series of structured questions.

2.1. *Are phase II screening studies still relevant?*

The traditional aim of phase II studies has been to screen out ineffective agents and identify the most promising ones for randomised controlled phase III studies (RCT), based upon achieving a pre-specified response rate. Some novel agents have demonstrated improved survival in RCT despite low response rates, raising concerns that phase II studies are irrelevant; some researchers have suggested progression from phase I directly to phase III is appropriate.

Although recognising the challenges of designing phase II studies, the task force felt that overall, an appropriately designed phase II study remains an important mechanism for screening out ineffective agents. Given the large number of new agents entering clinical studies, it is unlikely that either financial or patient resources would be able to support definitive RCTs of every agent. Further, while observed response rates may be lower than those expected with classical cytotoxic agents, some targeted agents do have an appreciable response rate, and appear to remain predictive of FDA approval.

In addition, phase II studies include relatively homogeneous patient populations and usually include patients who are less heavily pre-treated than those accrued to phase I studies. Phase II studies may therefore be an ideal setting to test a number of additional hypotheses, including confirming the dose selected in phase I studies, optimising dose (dose escalation may be considered, or two doses compared), selecting the optimal schedule; testing and validating putative predictive markers or demonstrating proof of concept for selected agents (for e.g. poly-ADP-ribose-1 inhibitors in BCRA deficient tumours).[2]

Finally, phase II studies have led to conditional approvals in some indications when a large phase III may not be feasible or indicated. The phase II study of imatinib in advanced gastrointestinal stromal tumour (GIST) is one example in which phase II data alone has led to regulatory approval.[3]

2.2. *What are suitable endpoints?*

Given the well described concerns regarding objective response with targeted agents, the MDICT discussed: level of response; alternative methods of evaluating response such as waterfall plots and functional imaging; the inclusion of dual or multinomial endpoints (including prolonged stable disease or absence of progression), or progression free survival.

After careful consideration, the task force concluded that objective response as defined by response evaluation criteria in solid tumours (RECIST) criteria remained appropriate to include in phase II studies, but that the use of a multinomial design should be considered for all studies.[4] In addition, studies should be appropriately powered. This may be of particular relevance when a low response rate is anticipated and hypotheses and statistical power used to detect response rates of $\geqslant$20% may be inappropriate. Modern phase II studies may require larger sample sizes to detect lower response rates that may signal the agent is capable of improving survival.

The task force was particularly interested in further exploration of non-dichotomous response endpoints including tumour volume or kinetics measures. A number of such observations have been explored such as slowing of progression rates,[5] and the use of 'waterfall' or 'spider' plots[6] (Fig. 1). There was general agreement that while these are still exploratory, efforts to develop and validate these as potential endpoints should strongly be encouraged.

It was appreciated that in some tumour types, where tumour measurements are notoriously difficult (e.g. mesothelioma, pancreatic cancer, ovarian cancer) the RECIST criteria may be suboptimal and validation of modified RECIST criteria may be required.[7] Similarly, it was felt that

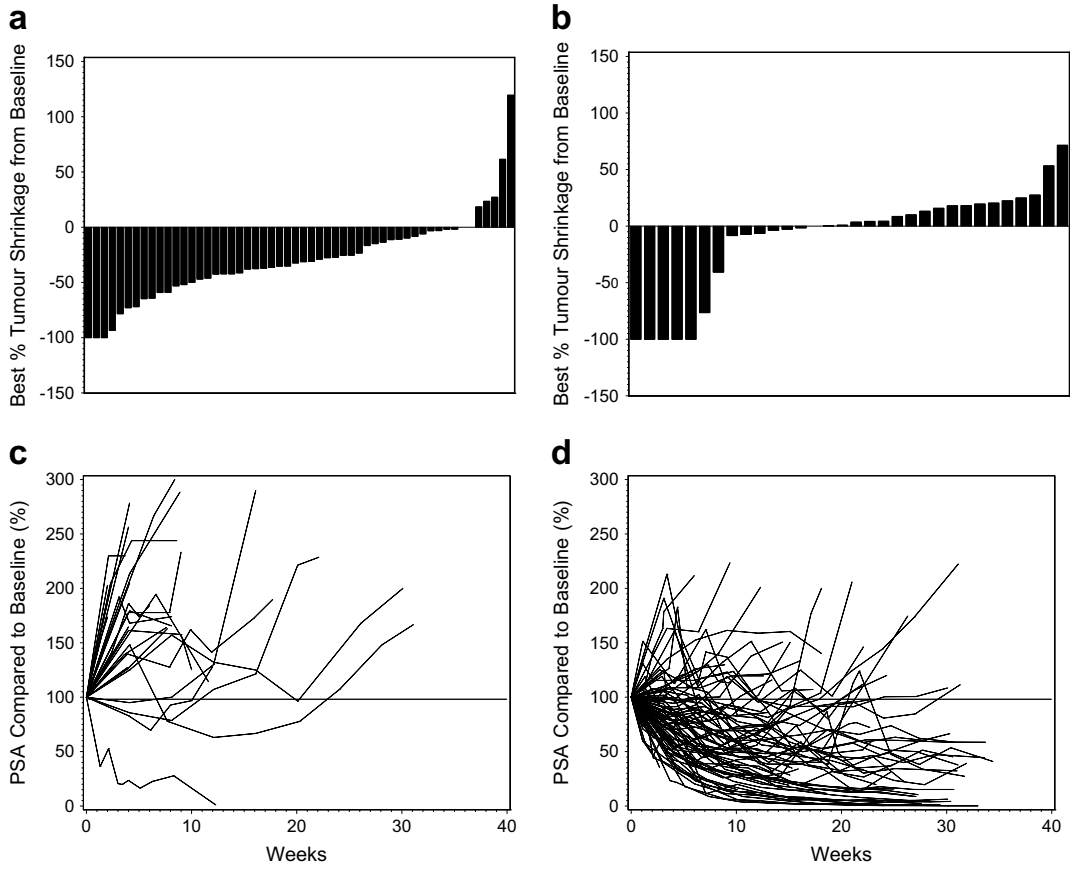

Fig. 1 – Simulated examples of waterfall and spider plots. Panels (a) and (b) are waterfall plots depicting the best percentage reduction in tumour size, as evidenced by the sum of diameters of target lesions. The agent depicted in (a) appears to have more antitumour activity than the agent depicted in (b). Panels (c) and (d) are examples of spider plots, showing changes in prostate-specific antigen (PSA) levels as a percentage compared to baseline (horizontal line). The agent depicted in (c) appears to be less active compared to the agent in (d).

surrogate indicators of response such as functional imaging are interesting, but largely still exploratory and that further work is to validate these as surrogate endpoints is encouraged.[8]

The inclusion of imaging and tissue based biomarkers in phase II studies was extensively discussed. The term biomarker has been used loosely used to describe measures of pharmacodynamic effect in phase I studies (e.g. change in phosphorylated epidermal growth factor receptor (EGFR) with EGFR inhibition; DCE magnetic resonance imaging changes suggesting vascular endothelial growth factor receptor (VEGFR) inhibition). For phase II and phase III studies, the term generally refers to predictive factors – i.e. markers (usually tissue/tumour based) that predict for benefit (e.g. oestrogen receptor expression predictive of response to tamoxifen), but may also be serial measures of surrogate effects such as changes in functional imaging. Predictive molecular factors are not surrogates of response or endpoints, but rather may be used to enrich the patient population (discussed below) in phase II studies. As discussed in MDICT 2006, all biomarkers to be included in clinical study should be tested and validated in preclinical models at a minimum prior to the initiation of a clinical study.

2.3. *What are appropriate designs for screening studies?*

Phase II design is of critical importance and the discussion included considerations such as randomisation (non-comparative reference arm, phase II/III designs, selection/pick the winner designs), enrichment strategies (randomised discontinuation design, molecular enrichment strategies) as well as adaptive Bayesian designs.

Although some have suggested that a randomised design is mandatory for evaluation of molecular targeted agents, there are a number of valid methodologic concerns regarding the design and interpretation of small randomised studies especially regarding high false-positive and negative rates.[5,9] The task force felt that single-arm phase II studies with traditional statistical designs remain appropriate in a number of scenarios. This was felt to be particularly true when testing a targeted agent in tumour types for which robust historical databases exist, such as pancreatic cancer and glioblastoma. Single-arm studies are also appropriate when the hypothesis of the study is merely to demonstrate target effect, or in late disease or salvage settings for which no standard treatment exists. Similarly, for agents where response is expected, traditional two-stage designs may be appropriate.

However, randomisation may be relevant in when the use of a non-treatment arm is not appropriate, the natural history of the tumour being studied is not known, or the target population is unclear.[10,11] A randomised design may also be appropriate when it is necessary to select the optimal dose or schedule (when more than one dose or schedule has been tested in phase I), when no historical data are available to guide the statistical design of the study, when the patient population to be studied is very heterogeneous, when it is likely that outcomes have changed significantly due to a change in standard practice, or when testing combinations of agents. The task force urged caution in the design of such studies and interpretation of their results; especially when endpoints more typically included in phase III studies (such as progression-free survival) are selected. Small randomised studies bring with them unavoidable issues such as high false negative and positive rates, unstable *p* values, and the risk that these studies, whether negative or positive, may (incorrectly) preclude later definitive studies.[5,9] The task force recommended a retrospective review of published randomised phase II studies be undertaken to define their positive predictive value.

Adaptive Bayesian methods (Fig. 2) are being increasingly suggested as appropriate designs for phase II studies and are of interest, although as yet unproven.[12,13] The task force did acknowledge that not all academic institutions have sufficient statistical resource to implement these designs, and that such designs may limit the collection of validation of biomarkers.

Although randomised discontinuation designs are a well described method of patient enrichment, molecular enrichment is increasingly discussed as an important phase II design criterion. Since studies of molecular agents have an underlying hypothesis regarding a molecular target, enriching the study population to include only patients expressing the relevant molecular phenotype seems intuitively sound. While this methodology is appropriate for agents for which a strong preclinical rationale exists (such as PARP-1 inhibitors in BRCA1/2 deficient patients with ovarian or breast cancer), for many agents the evidence that activity is limited to patients with specific molecular phenotypes or high levels of target expression is tenuous in early stages of development of the drug. The MDICT felt that in general, a strategy of enrolling all patients with a particular tumour type, while collecting tumour specimens (archival if necessary) at entry to the study allowing measurement of possible predictive biomarkers is more appropriate. If the molecular hypothesis is sufficiently compelling, accrual of an appropriately sized cohort of patients with the molecular phenotype can be planned.

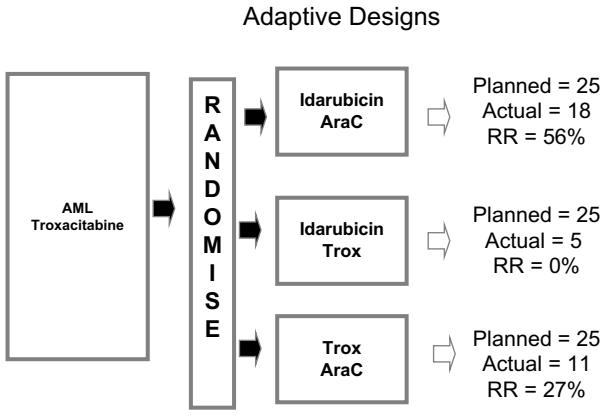

Fig. 2 – Example of an adaptive design. A Bayesian design was used in this phase II study to adaptively randomise patients to treatment. Although there was initially an equal chance for randomisation to IA, IT, or TA, treatment arms with a higher success rate progressively received a greater proportion of patients.[13]

Although definitive clinical validation of biomarkers which are surrogates of efficacy or predictive of outcomes are most efficiently done in phase III and post-registration studies, there may be a sound rationale for their inclusion in phase II studies, such as to gain mechanistic information, as an early endpoint (i.e. changes in positron-emission tomography (PET) scan) or as interim validation of a biomarker for use in RCTs.

2.4. *What needs to be reported on phase II studies?*

The task force discussed the reporting of phase II studies and commented that many manuscripts failed to adequately describe important data such as patient demographics and the objectives as well as planned endpoints of the study. Of concern was the failure to publish studies that failed to meet the planned endpoints or which closed prematurely. Journal editors and investigators were urged to ensure publication of all clinical study results.

3. MDICT task force recommendations

Based on the scientific review and discussion, the MDICT task force developed a series of recommendations for the design, implementation, analysis and output of phase II studies of targeted agents.

Are phase II studies screening still relevant/appropriate?

1. Phase II studies continue to have an important role to play in the development of molecular targeted agents. Phase II studies enroll more homogeneous and less heavily pretreated patient populations than phase I studies and are an appropriate setting to optimise dose and schedule, test proof-of-concept for selected agents and test predictive markers.

   What are suitable endpoints?
2. Multinomial endpoints and designs should be considered for phase II studies of targeted agents especially when response rates are anticipated to be low.
3. Tumour growth kinetic assessments (slowing of rate of progression, spider and waterfall plots) and biomarker or functional imaging endpoints are generally not sufficiently validated to be considered primary endpoints at present, but continued assessment and validation, including in preclinical models prior to initiation of clinical studies, is strongly encouraged.

   Is randomisation mandatory?
4. Single arm phase II studies continue to be appropriate in certain settings, especially when the likely outcomes in the population studied are well described.

5. Small randomised studies have a number of statistical pitfalls and dangers. Nonetheless, in select circumstances, randomised phase II clinical studies are helpful to define the best dose or schedule, or to test combinations. Endpoints should be carefully considered so as not to render this an underpowered phase III trial. Further examination of the positive predictive power (vis-à-vis eventual drug success) of completed randomised phase II studies should be considered.
6. The role of other designs such as randomised discontinuation or Bayesian adaptive design is not yet clear. Such designs are of interest but remain exploratory at this time.

Is molecular enrichment and/or biomarker testing mandatory?

7. In the presence of a compelling rationale, phase II studies may be conducted in enriched patient populations; however, more frequently, these studies should test possible predictive biomarkers and ensure collection of baseline tissue for consenting patients.
8. The clinical validation of surrogate markers of efficacy through serial tissue collection or functional imaging may be best suited to studies after clinical benefit has been demonstrated in phase II, such as phase III or post-registration studies setting as such investigations are costly to perform and complex to conduct.

Reporting of phase II studies

9. Complete reporting, including negative studies or studies that have been prematurely terminated, is important, and publications should include all relevant information such as patient characteristics, design, endpoints and predefined go/no go criteria.

4. Conclusions

Molecular targeted agents pose substantial challenges in contemporary drug development. The MDICT task force believes that phase II studies remain an important screening tool in drug development, but require thoughtful design appropriate for each individual agent. Consideration should be given to the use of multinomial trial designs. Unless there is a compelling rationale, molecular enrichment is not essential but the identification, piloting and validation of a putative predictive marker is an important endpoint of phase II studies. Randomisation may be appropriate in select circumstances, with recognition of the potential statistical limitations. Important areas for future research and validation include assessment of tumour growth kinetics, functional imaging, and adaptive trial designs as these are not yet sufficiently validated to include routinely in the design of a phase II trial. Topics in contemporary drug development which merit discussion at future meetings of the MDICT task force were identified and include the utility of phase 0 designs, and methodology of phase I studies for combinations of targeted and/or cytotoxic agents.

Conflict of interest statement

None declared.

Acknowledgements

The authors wish to acknowledge the contribution of the 2007 MDICT task force membership to the writing of the recommendations and manuscript. Task force members present at the 2007 meeting include: Steinar Aamdal (Norway), Susan Bates (United States), Hilary Calvert (United Kingdom), Johann De Bono (United Kingdom), Gail Eckhardt (United States), Tito Fojo (United States), Giuseppe Giaccone (United States), Marinus Lobbezoo (The Netherlands), Peter O'Dwyer (United States), Eric Raymond (France), Jan Schellens (The Netherlands), Cristiana Sessa (Switzerland), Lesley Seymour (Canada), Jean-Charles Soria (France), John Wright (United States), Heinz Zwierzina (Austria).

The NDDO Research Foundation (Amsterdam, The Netherlands) is acknowledged for its administrative and financial support to the MDICT task force.

REFERENCES

1. Booth CM, Calvert AH, Giaccone G, Lobbezoo MW, Seymour LK, Eisenhauer EA. Endpoints in phase I studies of targeted anticancer therapy: report of the task force on methodology for the development of innovative cancer therapies (MDICT). *Eur J Cancer* 2007, in press.
2. Bryant HE, Schultz N, Thomas HD, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* 2005;**434**(7035):913–7.
3. Demetri GD, von Mehren M, Blanke CD, et al. Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. *N Engl J Med* 2002;**347**(7):472–80.
4. Dent S, Zee B, Dancey J, Hanauske A, Wanders J, Eisenhauer E. Application of a new multinomial phase II stopping rule using response and early progression. *J Clin Oncol* 2001;**19**(3):785–91.
5. Rubinstein LV, Korn EL, Freidlin B, Hunsberger S, Ivy SP, Smith MA. Design issues of randomized phase II trials and a proposal for phase II screening trials. *J Clin Oncol* 2005;**23**(28):7199–206.
6. Ratain MJ, Eisen T, Stadler WM, et al. Phase II placebo-controlled randomized discontinuation trial of sorafenib in patients with metastatic renal cell carcinoma. *J Clin Oncol* 2006;**24**(16):2505–12.
7. Byrne MJ, Nowak AK. Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. *Ann Oncol* 2004;**15**(2):257–60.
8. Benjamin RS, Choi H, Macapinlac HA, et al. We should desist using RECIST, at least in GIST. *J Clin Oncol* 2007;**25**(13):1760–4.
9. Redman M, Crowley J. Small randomized trials. *J Thorac Oncol* 2007;**2**(1):1–2.
10. Rosner GL, Stadler W, Ratain MJ. Randomized discontinuation design: application to cytostatic antineoplastic agents. *J Clin Oncol* 2002;**20**(22):4478–84.
11. Freidlin B, Simon R. Evaluation of randomized discontinuation design. *J Clin Oncol* 2005;**23**(22):5094–8.
12. Berry DA. Bayesian clinical trials. *Nat Rev Drug Discov* 2006;**5**(1):27–36.
13. Giles FJ, Kantarjian HM, Cortes JE, et al. Adaptive randomized study of idarubicin and cytarabine versus troxacitabine and cytarabine versus troxacitabine and idarubicin in untreated patients 50 years or older with adverse karyotype acute myeloid leukemia. *J Clin Oncol* 2003;**21**(9):1722–7.