Clinical trial designs for targeted agents

Brigette B.Y. Ma, MBBS, FRACP\textsuperscript{a},
Carolyn D. Britten, MD, FRCPC\textsuperscript{b},
Lillian L. Siu, MD, FRCPC\textsuperscript{a,*}

\textsuperscript{a}Department of Medical Oncology, Princess Margaret Hospital, University Health Network, 610 University Avenue, Toronto, ON M5G 2M9, Canada
\textsuperscript{b}University of California Los Angeles, Jonsson Comprehensive Cancer Center, Los Angeles, CA 90095-7077, USA

Decades of intensive research in the molecular biology of human cancer have cumulated in a substantial paradigm shift in the treatment of this disease. The evolution of cancer is regulated by a complex network of aberrant genes and signaling pathways that control tumor proliferation, apoptosis, metastasis and angiogenesis [1]. Identification of these genes and molecules has provided an abundant source of potential therapeutic targets, resulting in the unprecedented variety and quantity of targeted agents available for clinical testing [2]. These compounds include monoclonal antibodies, antisense oligonucleotides, small molecule kinase inhibitors, ribozymes, and gene therapy. In general, these novel agents differ from conventional cytotoxic chemotherapy in two fundamental ways. First, targeted therapies may be predominantly antiproliferative, rather than cytotoxic. Second, often they produce less acute toxicity than cytotoxic agents, resulting in a broad therapeutic margin. These differences pose major challenges for drug development, necessitating modification of traditional clinical trial designs to accommodate these novel agents [3–6]. Failure to adopt the appropriate trial designs partly contributes to 90\% of new anticancer agents never making the transition from the laboratory to bedside [7]. This article discusses some of the strategies employed to evaluate contemporary anticancer agents, citing examples of some recently tested targeted agents to illustrate the advantages and disadvantages of various approaches.
Phase I trial design – limitations and novel strategies

The traditional phase I trial is a dose-finding study in which small cohorts of patients with advanced cancer are treated at escalating doses according to chronological entry into the study [8]. The principal scientific objective is to determine the maximally tolerated dose of the new agent using a particular dosing schedule, based on the assumption that the anticancer drug must be administered at relatively high doses. Recommended phase II dose has been regarded by some as being synonymous with the maximum tolerated dose but by others as being one level below the maximum tolerated dose. This semantic issue remains contentious [9]. Although cytotoxic agents generally have a significant dose-response effect, cytostatic agents may have a saturable target, such that dose escalation beyond a certain threshold may not result in increased efficacy but increased toxicity. Alternatives to traditional toxicity endpoints and modification of the conventional dose escalation schemes are being explored (Table 1).

Use of endpoints other than toxicity to guide dose escalation

The use of toxicity endpoints for evaluating targeted agents in a phase I trial can be problematic, given the variability of their toxicity profile. For agents that produce noticeable toxicity at doses close to the maximum tolerated dose, the use of toxicity as a primary endpoint remains appropriate. For instance, the

| Table 1 | Phase I trial—comparison between designs for cytotoxic and cytostatic agents |
|-----------------|---------------------------|-----------------------------|
| Phase I design  | Traditional /cytotoxic | Targeted / cytostatic |
| Treatment endpoint to guide dose escalation | Toxicity | Often not toxicity |
|                     |                         | Target inhibition |
|                     |                         | Surrogates for activity |
|                     |                         | Plasma level |
| Starting dose       | Based on dose required for tumor regression in animal models (eg, one tenth of lethal dose or MTD in mice; one-third of TDL in dogs). | Based on dose required to inhibit tumor growth in animal models, or to demonstrate molecular endpoints such as induction of apoptosis, inhibition of angiogenesis. |
| Number of patients explored per dose level | (1) Modified Fibonacci: three to six per level | More than six per level needed if dosing based on biologic or pharmacological endpoints. |
|                     | (2) Accelerated dose titration design; modified continual reassessment method; pharmacologically guided dose escalation: at low doses, one to two per level; if moderate toxicity occurs, expand to three or more per level. | |

Abbreviations: MTD, maximum tolerated dose; TDL, toxic dose low.
tyrosine kinase inhibitor (TKI) against the epidermal growth factor receptor ZD1839 (Iressa) is known to cause consistent, dose-limiting moderate-to-severe (grade 3) diarrhea [10,11]. The caveats of this approach are two-fold, however. First, information about a drug’s toxicity profile often becomes evident only after completion of trial. Second, the traditional definition of “intolerable toxicity” usually is defined as grade 3–4 toxicity using the National Cancer Institute Common Toxicity Criteria (NCI-CTC). This definition may be insufficient for chronically and/or continuously administered targeted agents that produce noticeable, moderate (grade 2) toxicities, because such toxicities can be unacceptable if they become persistent. For example, chronic mild-to-moderate (grade 2) mucositis and diarrhea were regarded as dose-limiting in a phase I trial of the metalloproteinase inhibitor BAY 12-9566 [33]. For targeted agents that do not produce immediate or consistent drug-related toxicity, three categories of alternative endpoints can be considered. These are: (1) measuring target inhibition, (2) plasma drug levels that reflect achievement of biologically relevant concentrations, and (3) surrogate markers of biologic activity in non-tumoral tissues [3].

**Target inhibition**

Measuring target inhibition in tumoral tissues requires the selection of an appropriate molecular target, an accessible tumoral tissue, and a reproducible assay. The target should be involved directly in the development and/or progression of the cancer type, and ideally its function and/or expression is interfered selectively by the new agent. Tumoral tissues should be obtained and analyzed by an experienced team of interventional radiologists, pathologists, and scientists. The problems with this approach can be pragmatic and technical. First, tumor biopsy is invasive, costly, and subject to sampling error and molecular heterogeneity of malignancies [12]. Moreover, if the procedure is optional, patient refusal could result in smaller sample size, thus undermining the statistical power of the correlative study. One solution is to sample tumor cells in third space fluids such as pleural or peritoneal fluids or to administer the drug during a preoperative window [4]. Moreover, in a large single-center, 10-year review (n = 99 patients) of sequential tumor biopsies in phase I/II trials of anticancer agents, it was found that in the hands of an experienced team, the rate of patient refusal for repeated biopsies and the rate of procedure-related complications were low. The most common reason for failure was procurement of necrotic tumor or fibrous or normal tissue (8%) [13]. Another problem in using target inhibition as an endpoint is that target selection can be hampered, because either the selected target is “wrong” and not solely responsible for the malignant phenotype, or because the agent is “wrong” and does not interact with the target. Furthermore, choosing an arbitrary level of target inhibition to guide dose escalation can be a difficult task [3]. To illustrate the point of “missing the target,” farnesyl transferase inhibitors originally were developed to selectively target mutated Ras, present in 50% of human cancers. Subsequent research revealed, however, that the antitumor activity of farnesyl transferase inhibitors is independent of Ras
status, suggesting that this activity is caused by inhibition of farnesylated proteins other than Ras [14].

**Plasma drug concentration**

Another strategy involves identifying a target dose based on optimal pharmacological endpoints. In this method, preclinical models are used to identify a parameter of systemic exposure (such as area under the concentration-time curve [AUC], steady-state concentration [Css], or time above a threshold concentration) that correlates with antitumor activity. This approach is valid as long as the plasma drug concentration correlates with the tumor drug concentration, and there are no significant interspecies differences in dose-response curves [15,16]. Problems arise if the drug demonstrates significant protein binding in plasma, if oral absorption is saturable, if interpatient variability in drug metabolism is significant, or if too few patients are explored per dose level [16,17]. Furthermore, achieving a pharmacokinetic target does not assure a biologic effect.

**Surrogate markers in nontumoral tissues**

The third strategy involves the use of surrogate markers such as circulating growth factors (eg, angiogenic factors and cytokines [18]), peripheral leukocytes (eg, CD cell count in immunotherapy [4]), skin (eg, ZD 1839 and other anti-epidermal growth factor receptor therapies [19]), or buccal mucosa (eg, some Ras inhibitors). Serological tumor markers, such as Ca-125, Ca 19-9, CEA and PSA, also have been used to guide dosing in phase I studies with the metalloproteinase inhibitor BB2516 (Marimastat) in patients with ovarian, pancreatic, colon, and prostate cancer, respectively [20,21]. In addition, functional imaging techniques including MRI and positron emission tomography (PET) are utilized increasingly in the evaluation of tumor angiogenesis [22,23]. Despite recent attempts to include surrogate endpoints in phase I clinical trials, until these endpoints have been validated properly with biologic effect and clinical outcome, they should not be used alone to guide dose escalation. For instance, although skin biopsies of patients treated with ZD 1839 showed blockade of downstream signaling at doses well below the maximum tolerated dose, the correlation between changes of these biologic markers in skin (eg, apoptosis, p27) and clinical response, is largely unknown [24]. In practice, toxicity should remain an important endpoint in guiding dose escalation, and where applicable, the final dose for further testing should be determined with consideration of alternative endpoints [3]. Phase I trial design of targeted agents with dose-effect and dose-toxicity curves that span over a wide dose range (Fig. 1) should incorporate the “optimum biologic dose” [5] or “maximum target-inhibiting dose” [3] as an endpoint, in addition to the conventional maximum tolerated dose.

**New ideas in dose escalation scheme**

Phase I dose escalation schemes are designed to obtain the dose-toxicity profile of an investigational agent using a minimum number of patients.
Traditional dose escalation schemes with diminishing increments between successive dose levels, such as the modified Fibonacci method, were developed empirically. Although this approach to dose escalation is safe, few patients actually receive the optimal dose [25]. A recent consensus among phase I investigators stated that dose escalation based on the traditional modified Fibonacci method should no longer be the gold standard [26]. Given the increasing number of new agents entering phase I trials, several novel dose escalation designs have been developed in an effort to improve safety and efficiency. These include the pharmacologically guided dose escalation [27], statistically based methods such as the modified Continual Reassessment Method [28,29], and the accelerated titration method [30]. These methods often involve enrollment of only one patient per dose level, along with dose escalations of 50% to 100% increments over earlier dose levels. Reversion into a Fibonacci-type sequence with cohort expansion occurs if moderate toxicity is encountered, or if the plasma drug concentration reaches a target AUC. The target AUC may be derived from a pharmacokinetic model constructed from preclinical data or from a mathematical model based on data from previous
patients in the trial. In a retrospective, single institutional review, the modified Continual Reassessment Model was compared with the traditional Fibonacci scheme using data from 50 completed phase I trials. For drugs that have cytostatic activity and modest toxicity, the modified Continual Reassessment Model allowed more dose levels to be evaluated using smaller numbers of patients than the traditional method, even if the starting dose was much lower than the maximum tolerated dose. Implementation of the modified Continual Reassessment Model failed to accelerate trial completion, however, because of the rigorous safety monitoring required with a single patient per dose level [31]. In another survey of phase I designs, proposals on new anticancer agents (including targeted drugs) submitted to the Food and Drug Administration (FDA) over an 11-year period were reviewed, and the applicability of the pharmacologically guided dose escalation design was compared with the modified Continual Reassessment Model. Seventy-six percent of the trials proposing modified Continual Reassessment Model designs were implemented, whereas only 50% of the trials proposing pharmacologically guided dose escalation designs were implemented. Nonetheless, pharmacologically guided dose escalation trials were more often able to reach the recommended phase II dose with fewer than predicted numbers of patients, compared with modified Continual Reassessment Model trials [32]. Although novel dose escalation schemes may make more efficient use of patient resources than the traditional modified Fibonacci scheme, the enrollment of fewer patients per dose level prohibits reliable conclusions about biologic endpoints, because of insufficient statistical power [16]. Generally speaking, the total sample size required in phase I trials using biologic endpoints is often larger than those based on toxicity, especially if the difference in biologic response observed between dose levels is expected to be small [5,33].

Patient selection and pharmacokinetic study in phase I trials

Most phase I studies select patients with advanced cancer, resulting in a high dropout rate from progressive disease or declining health during study. This poses a problem when assessing the tolerability and biologic effect of targeted agents that may not have immediate toxicity and may also require prolonged administration for evidence of toxicity and antitumor activity [5]. Conducting phase I trials with pharmacokinetic assays in healthy volunteers is one solution [34,35], but there is concern that the results may not be applicable to sick cancer patients. Another problem of selecting patients with advanced cancer is that some agents may work better in earlier stages of disease when the tumor burden is lower. This can be overcome by exploiting the preoperative window and administering targeted agents to patients with earlier staged resectable cancer awaiting definitive surgery. This approach may not allow sufficient time to explore the efficacy and non-acute toxicities of the drug, however, and patients may find the idea of undergoing “unnecessary” experimental therapy prior to potentially curative surgery unappealing.
Phase II trial design – limitations and novel strategies

Following completion of phase I testing of targeted agents, the options include proceeding directly to a phase III trial or performing a phase II trial alongside further correlative biologic studies. The objectives of phase II trials are to screen agents for efficacy and to further define toxicity. For cytotoxic agents, the traditional endpoint is the overall tumor response rate (sum of complete and partial responses) as defined by the standardized World Health Organization (WHO) or Response Evaluation Criteria in Solid Tumor (RECIST) criteria [36]. Patients are selected on stringent criteria based on cancer type, disease stage, and other characteristics. Commonly, phase II trials are conducted according to a two-stage design with incorporation of an early stopping rule after the first stage in case the agent or combination does not show a minimum response rate (P₀); this guards against excessive patient exposure to inactive agents [3,37]. At the end of the second stage, the agent or combination would be considered worthy of further evaluation if a target probability of response is achieved (P₁). The fundamental problem of applying this framework to most targeted agents is their propensity to induce tumor stabilization rather than regression. Therefore, the sole use of tumor response as a marker of efficacy may be inadequate. Other difficulties in the continual development of targeted agents after completion of phase I trials include selecting the most promising candidates, adopting the most appropriate dose and schedule of each compound, and combining an agent safely and effectively with other existent therapies. Potential strategies to overcome these problems include incorporation of novel endpoints of tumor stabilization, use of

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase II trial—comparison between designs for cytotoxic and cytostatic agents</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Endpoints</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Design</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
randomized phase II trial designs to expedite multiple comparisons, or even direct transition from phase I to phase III testing. (Table 2)

**Novel endpoints and designs in nonrandomized phase II trials**

Some investigators have advocated the use of alternative phase II endpoints based on the theoretical ability of a drug to slow tumor progression without inducing tumor shrinkage. These alternative endpoints include stable disease rate, time to progression, and progression rate (the proportion of patients whose disease has progressed over a predefined duration of therapy) [3,5,38]. Historical data derived from retrospective series are used sometimes to estimate the expected range of time to progression or progression rate, as a benchmark for hypothesis formulation [3,6]. These estimates must be interpreted cautiously, however, because of patient heterogeneity between trial candidates and historical controls. Application of these endpoints can be resource- and time-consuming, because it requires consistent and regular radiological and clinical follow-up. Furthermore, assessment of disease progression can be subjected to investigator bias and confounded by the mode and frequency of radiological or clinical evaluations.

To circumvent invalid comparisons using retrospective information, some investigators have compared the rates of progression within the same individual on different therapies, using patients as their own controls rather than historical data [5]. In one approach, the time interval of non-treatment between failure from prior therapy to trial entry (time to first progression) is compared with time to progression after trial therapy (time to second progression), with the two values expressed as a ratio [5]. A second approach assumes that the patient has failed a treatment recently before trial entry, compares the time to progression interval for the previous treatment (TTP1) and time to progression for the current trial (TTP2) in a pairwise analysis. The agent is considered active if the TTP1/TTP2 ratio, or the “growth modulation index” is above 1.33. Chi-square statistics are used to evaluate the paired failure time (TTP1/TTP2), and efficacy is expressed as a hazard ratio [39]. These methods are fraught with practical difficulties and require clinical evaluation [5].

Although these alternative phase II endpoints may allow a better assessment of the drug’s efficacy, tumor response still is considered important in practice, as many targeted agents show tumor response in preliminary phase I or II testing. For example, the tyrosine kinase inhibitor STI571 in chronic myelogenous leukemia, and ZD1839 and OSI774 in solid tumors have produced tumor shrinkage early in their course of development [40]. Against this background, a multinomial phase II stopping rule has been developed to incorporate tumor response and early progression as paired variables [41]. This strategy operates on several paired, possible observations that determine the decision to stop or continue the trial at the end of first stage in a two-stage phase II design. The multinomial design recently was evaluated retrospectively using data from 39 completed phase II trials conducted by two large international clinical trials groups [42]. Among those trials involving agents that later were deemed ineffective, the multinomial design
appropriately advised early stopping at the first stage more efficiently than traditional designs by Fleming or Gehan [43,44]. Because this analysis was performed exclusively using trials with cytotoxic agents, the applicability of this design to targeted agents has been questioned [45].

Other investigators have moved towards measuring cancer-related symptomatic response or quality of life as trial endpoints. Two recent phase II trials of ZD 1839 in patients with advanced non-small cell lung cancer have provided methodological support for assessing symptom control as a useful trial endpoint. They showed that symptomatic improvement is associated with a longer progression-free survival and overall survival independent of objective tumor response [46,47]. Still, this approach requires validation in randomized phase III studies.

Novel designs in randomized phase II trials

One-stage randomized phase II design

Picking the “right” targeted agent from a diverse pool of single-arm phase II trials can be daunting. Variability between studies and across institutions in terms of patient selection, study endpoints (eg, conventional versus surrogate), treatment delivery, and disease evaluation makes it difficult to judge one phase II trial from another. The original randomized phase II trial has a single-stage design, where patients are allocated randomly to two or more experimental regimens in the presence of a control arm, which may be a placebo or “standard” regimen. The goals are to eliminate variability by screening multiple agents under one standardized trial condition, to increase the chance of detecting treatment activity, and to avoid discarding a potentially useful therapy [33,48]. If a placebo is used, then a crossover design or a 2:1 randomization favoring the experimental arm may be incorporated to attract patient accrual. As in a phase III trial, double blinding is applied whenever feasible in a randomized phase II trial to reduce bias. Unlike a phase III trial, however, a randomized phase II trial is not powered to determine whether one agent is superior or equivalent to another; hence a more economical sample size can be used. It is prone to misuse when it is seen as a substitute of, rather than simply a pilot for, a definitive phase III trial [49]. Several recent trials on targeted agents have employed this design. In one trial, patients with advanced renal cell cancer were randomized to three different doses of the mTOR inhibitor CCI-779 in a double-blinded fashion, using tumor response and time to progression as endpoints [50]. In another trial, patients with melanoma were randomized in a 2:1 fashion to a combination of weekly paclitaxel plus a metalloproteinase inhibitor or a placebo, using tumor response, progression-free survival, and overall survival as endpoints [51].

Two-stage sequential randomized phase II design

The original randomized phase II trial has a single-stage design, where the most active experimental regimen (the “winner” agent/dose/schedule/combination) will be selected for phase III testing. A variation of the randomized phase II design is the two-stage sequential trial, where the “winner” is picked...
among experimental arms in the first stage, then carried over to the next trial for comparison with a control. In this design, the control may be included in the second or in both stages and may represent a placebo or a standard regimen. The trial is terminated in the first stage if the regimen with the highest score of endpoint (e.g., tumor response, disease stabilization, or other surrogate markers) falls short of a cut-off level [52,53]. This schema is modified in another design, where the “winner” itself becomes the new control in the next trial and is compared with other experimental regimens [6,54]. The sequential design minimizes overall sample size and has a high rate of termination if the different arms are equivalent. It is best for screening agents when their activity is not substantially different from each other or from standard therapy [55]. An application of this type of trial design is exemplified by two sequentially designed randomized phase II trials in patients with hormone refractory prostate cancer. In the first trial, patients were randomized to two dose levels of thalidomide [20]. The dose that induced the higher rate of fall in the prostate-specific antigen (PSA) level was then used in a second trial, where patients were randomized to weekly docetaxel alone (control) or in combination with thalidomide [56].

**Enrichment design**

Enrichment design is another variant of the two-stage design. Rather than using traditional intent-to-treat analysis, the objective is to eliminate potential contamination of response data from patient noncompliance, or dropout from patients with low tolerance to toxicity [57]. Furthermore, it may “enrich” the cohort by selecting those patients who may have a higher chance of responding to targeted agents, based on the assumption that only a proportion of patients would express the necessary molecule(s) that the drug is designed to target. This is useful if the assay for screening the putative target is not available or applicable for practical use, or if the true molecular target(s) are not defined. An example of applying the enrichment principle is the randomized discontinuation trial design [58–60], where in the initial or “enrichment” phase all patients are treated openly with the same agent. At the end of a 2- to 4-month period of evaluation, those who have partial or complete response will continue treatment until disease progression, whereas those who have stable disease will proceed to the second phase and be randomized to continue the treatment or placebo. Those with progressive disease subsequently will discontinue the study [60]. Although this design may minimize exposure to placebo, the enrichment process has been criticized as simply another form of sampling bias [61]. A higher total sample size than that of standard design may also be required, since only a limited proportion of patients will be deemed stable and therefore eligible for entry into the second phase [5,6]. Others are concerned that an otherwise effective agent might be declared ineffective if its fails to maintain activity with continued use in the second phase, and that some patients may find it unacceptable to be changed to a placebo if they are responding [5]. A Cancer and Leukemia Group B (CALGB) trial using the antiangiogenic agent, carboxyamidotriazole, has adopted this design and is currently under accrual for patients with metastatic renal cell carcinoma [62].
the enrichment phase, patients receive oral carboxamidotriazole daily for 4 weeks for a total of 4 courses. Patients who experience partial or complete response will continue treatment until disease progression or unacceptable toxicity. Those with stable disease rate are randomized to continue treatment or placebo.

**Integrative phase II/III design**

Most recently, there is a trend towards adopting an integrative phase II/III trial design in evaluating targeted agents, as a means to avoid the lengthy administrative process and to overcome some of the problems associated with transition from phase II to phase III [63]. The integrative randomized phase II/III design uses survival as a common endpoint for a sequential phase II and III trial [64,65]. It has been evaluated recently using hormone refractory prostate cancer as a model [65]. The first stage is a typical randomized phase II study, where patients are assigned randomly to two or more experimental arms in the presence of a control or standard arm. The experimental arms can be single agents, combinations, or different dosages or schedules of a combination. At the end of the first stage, pairwise log-rank statistics is used to compare survival times between each experimental arm and the standard arm. Sample size in this stage is not powered to determine superiority of one experimental arm over another. Treatment arms with a pairwise log-rank statistic $S_j (n_i)$ of at least 1 denote a longer median survival and will be selected for the second stage, whereas those with $S_j (n_i)$ less than 1 will be dropped. The trial is terminated at this stage if none of the experimental arms show a positive value. The second stage is a typical phase III trial, where accrual will continue in a randomized manner to those arms that are still in the trial until a sufficient total sample size has been reached to determine the superiority of any particular arm. Unlike all of the randomized phase II designs discussed so far, this integrative approach uses a common survival outcome such as overall survival or time to progression in the phase II and III components. This design is suitable for therapies that prolong life predominantly by way of disease stabilization and in more indolent disease. This is because it avoids the difficulty of using short-term measures of response such as tumor response, which may not reflect the activity of the drug or the natural history of a disease (e.g., in prostate cancer, where bony metastasis is common but not measurable for response). The design is also suitable for a common disease where there are numerous completed phase II trials, but few resources to perform multiple large-scale phase III trials (e.g., breast cancer). From a patient’s perspective, the lack of a placebo and minimized exposure to inactive agents are attractive.

**Are phase II trials really necessary?**

Finally, the question of whether phase II trials are really necessary for targeted agents remains controversial, given that these agents may not have discernible toxicity or tumor response. One may consider omitting phase II trials in the presence of compelling preclinical data [4], significant activity in phase I trials, and in circumstances where there is no “standard” therapy
(eg, maintenance treatment following partial or complete response to standard chemotherapy). Another circumstance in which direct transition from phase I to phase III may be considered involves the comparison of chemotherapy alone versus chemotherapy plus a targeted agent. This is a reasonable approach if the targeted agent shows significant synergism with chemotherapy in the preclinical setting, and assuming that the combination will not be inferior to chemotherapy alone [3]. Overall, it may be risky to expose a large number of patients to an experimental agent in a phase III trial without sufficient data of its efficacy, short- and long-term toxicity, optimal delivery, and activity spectrum in different cancer types. The aborted development of the metalloproteinase inhibitor Bay 12-9655 and of the FTI R115777 [14] in pancreatic cancer illustrate the disadvantage of conducting phase III trials with insufficient phase II data. In the 2000 National Cancer Institute (NCI) State-of-the-Science Conference on non-small cell lung cancer, it was recommended that phase III trials should be built with consideration of phase II data, because the latter are more informative on efficacy and the safety of long-term exposure [14].

Phase III trial design — new strategies

Several refinements to the conventional phase III design have been made to develop targeted agents optimally. The first strategy is to identify those patients with the highest chance of responding to an agent. Prescreening of patients for the overexpression and/or dysregulation of a novel agent’s molecular target has been utilized with success in the phase III development of STI 571 in chronic myelogenous leukemia [66] and trastuzumab in metastatic breast cancer [67]. This approach may not be widely applicable if target expression by the tumor does not predict for drug response because of the presence of multiple molecular abnormalities, as in the case of most advanced solid tumors [14]. Another strategy involves phase III testing in patients with earlier disease stage and lower tumor burden such as in the adjuvant or neoadjuvant setting, or in cancer chemoprevention. This strategy has been advocated for some metalloproteinase inhibitors and antiangiogenic agents [68,69]. It also has been applied in the development of the monoclonal antibody, edrecolomab, where it has been tested against placebo in an adjuvant phase III trial in patients with node negative stage II colon cancer [70].

Examples of successful and not-so-successful trial designs

Suboptimal clinical trial design may contribute to the failure of a drug to deliver its promise. There are excellent reviews in the literature that summarize the positive and negative experience in the development of several targeted therapies, such as the TKIs, farnesyl transferase inhibitors, metalloproteinase inhibitors and antiangiogenic agents (Table 3) [14,17,40,71]. In this section, two
examples will be cited to illustrate some of the pitfalls and successes of novel clinical trial designs.

**STI 571**

STI 571 has been cited as one of the triumphs in the development of a rationally designed targeted therapy [71]. Unlike the development of other TKIs in solid tumors, STI 571 has the advantage of striking an ideal molecular target: the bcr-abl tyrosine kinase, and an ideal disease: chronic myelogenous leukemia. Bcr-abl TK is expressed in all patients with chronic myelogenous leukemia and is linked directly to the pathogenesis of chronic myelogenous leukemia. The disease status of chronic myelogenous leukemia patients is defined by relatively accessible tumor-related hematological (eg, white blood cell (WBC) count in blood) and cytogenetic endpoints (eg, presence of the Philadelphia chromosome in bone marrow). The preclinical evidence is compelling and typical of a cytostatic agent, with STI 571 being a potent, selective, and sustained inhibitor of chronic myelogenous leukemia progenitor cell growth in vitro, associated with little toxicity to normal cells [71]. The optimal scheduling (continuous) and dosing...
of this drug were determined in animal models prior to clinical testing. Traditional tumor response and novel but validated pharmacokinetic indices (ie, plasma levels) and target inhibition were used as endpoints [72]. Like most cytostatic agents with wide therapeutic margins, no dose-limiting toxicity was encountered in the phase I studies of STI 571, and the maximum tolerated dose was never reached. Instead of performing randomized phase II trials to select the recommended phase II dose, it was estimated from the pharmacokinetic data. Significant antitumor activity (with complete response) was observed in phase I trials among patients with chronic and blastic phases of chronic myelogenous leukemia. Even with the compelling phase I results, phase II trials were performed subsequently, thus allowing previously unrecognized toxicity such as cytopenia, skin, and gastrointestinal disorders to be described [73]. The phase I/II data eventually led to its FDA licensing because of the impressive single agent activity in chronic myelogenous leukemia [74]. In a large, multicenter randomized trial, STI 571 was compared with conventional therapy (interferon and cytarabine) in patients with newly diagnosed chronic myelogenous leukemia. Using time to progression as an endpoint and allowing treatment crossover, STI 571 conferred a significantly superior time to progression than the control arm [66]. Using a similar algorithm, STI 571 also has been applied successfully in the treatment of gastrointestinal stromal tumors because of its inhibitory effect on c-kit, which is overexpressed, mutated, and constitutively activated in this disease [75].

Bay 12-9566

In contrast, the clinical development of the metalloproteinase inhibitor, Bay 12-9566, in pancreatic and small cell lung cancer was prematurely terminated because of methodological problems [76]. Unlike the targeting of tyrosine kinase in hematologic malignancy, the metalloproteinase system is extremely complex in solid tumors, such that the subtypes appropriate for therapeutic targeting remain to be elucidated [17]. Preclinically, BAY 12-9566 showed significant promise as a potent inhibitor of metalloproteinases -2, -3, and -9. It inhibits cell migration, angiogenesis, proliferation, and demonstrates a broad spectrum of antitumor activity in vivo against various human cancers. The pharmacokinetic profile was defined, showing a saturable absorption typically seen with some targeted agents. It was tolerated well among rats and healthy human volunteers [34], and results from five phase I trials have demonstrated stable disease (but not tumor shrinkage) in a significant proportion of patients [17]. In a phase I study of BAY 12-9566 [15] where plasma concentrations were used to determine the recommended phase II dose, however, the number of patients explored per dose level was considered too small for the results to be conclusive [16]. Another criticism in the development of BAY 12-9566 is the lack of sound phase II efficacy data on specific cancer types when making the transition from phase I to III testing in advanced pancreatic, ovarian, small cell lung, and colon cancer. For instance, in a multicenter phase III trial, patients with advanced pancreatic cancer were randomized to gemcitabine or BAY 12-9566. The study was designed to
demonstrate equivalence between the two arms. Using overall survival as the primary endpoint and quality of life, clinical benefit, and progression-free survival as secondary endpoints, two interim analyses were planned to terminate the trial earlier in the event that the experimental arm was ineffective. The early stopping rule allowed termination of the trial at the first interim analysis if fewer than 6 out of 30 patients on BAY 12-9566 were progression-free at 2 months. Yet, the first analysis failed to demonstrate this, and after 227 patients had been enrolled, at a second interim analysis, the gemcitabine arm was found to be significantly better than BAY 12-9566, prompting an immediate trial termination [77]. Similarly designed trials in ovarian and small cell lung cancer also were closed prematurely based on this result [78]. This trial has been criticized for the inadequate number of patients analyzed in the first interim stage, which led to a false sense of security for the investigators to continue the trial. Moreover, in the absence of phase II data, the antitumor activity (or the lack thereof) of BAY 12-9566 might not have been apparent during the relatively short evaluation period (2 months) leading to the interim analyses. Hence, alternative designs like the integrative phase II/III trial, or a pilot randomized phase II trial might have been more appropriate. Additionally, a metalloproteinase inhibitor such as BAY 12-9566 might have demonstrated some benefit had it been tested in patients with low tumor burden rather than in patients with advanced disease [77]. Finally, since BAY 12-9566 is cytostatic and has no single-agent cytoreductive activity, a more appropriate trial design would have compared the compound with gemcitabine versus gemcitabine alone.

Summary

With an increasing number of targeted agents available for testing, clinical trials must be rationally designed based on sound knowledge of the molecular mechanisms linking target and disease, fortified by strong preclinical data demonstrating how this relationship is modified by the targeted agent. Patients and resources are precious and should be expended judiciously on clinical trials that are well planned. Although traditional trial designs and endpoints may not be adequate for developing contemporary targeted drugs, transiting directly from phase I to phase III testing should be avoided except in distinct circumstances. Increased research efforts should be spent on the prospective evaluation and validation of novel biologic endpoints and innovative clinical designs, such that promising targeted agents can be effectively developed to benefit the care of cancer patients.

References

[3] Eisenhauer EA. Phase I and II trials of novel anti-cancer agents: endpoints, efficacy and exist-

1281–7.


[7] Von Hoff DD. There are no bad anticancer agents, only bad clinical trial designs—twenty-first

development guide. Preclinical screening, clinical trials and approval. Vol. 3. Totowa, New


[10] Mendelsohn JM, Baselga J. The EGF receptor family as targets for cancer therapy. Oncogene

receptor-tyrosine kinase inhibitor, is well tolerated and active in patients with solid, malignant

[12] Betensky RA, Louis DN, Cairncross JG. Influence of unrecognized molecular heterogeneity on


[14] Sebti SM, Hamilton AD. Farnesyltransferase and geranylgeranyltransferase I inhibitors and
cancer therapy: lessons from mechanism and bench-to-bedside translational studies. Oncogene


[16] Hidalgo M, Eckhardt SG. Matrix metalloproteinase inhibitors: how can we optimized their

escalation in a phase I study of the antiangiogenic aVb3 and aVb5 integrin antagonist EMD

receptor inhibitor ZD1839 in skin from cancer patients: histopathologic and molecular conse-

inhibitor, in patients with androgen-independent prostate cancer. Clin Cancer Res 2001;7:
1888–93.

3 doses of Marimastat™ in patients with hormone naïve, biochemically relapsed (M0) prostate


[22] Neeman M, Provenzale JM, Dewhirst MW. Magnetic resonance imaging applications in the

[23] Albanell J, Rojo F, Baselga J. Pharmacodynamic studies with the epidermal growth factor


