Design of clinical trials for biomarker research in oncology

Clin. Invest. (2011) 1(12), 1627–1636

The developmental pathway from discovery to clinical practice for biomarkers and biomarker-directed therapies is complex. While several issues need careful consideration, two critical issues that surround the validation of biomarkers are the choice of clinical trial design (which is based on the strength of the preliminary evidence and marker prevalence) and the biomarker assay related issues surrounding the marker assessment methods such as the reliability and reproducibility of the assay. This review focuses on trial designs for marker validation, both in the setting of early phase trials for initial validation, as well as in the context of larger definitive trials. Designs for biomarker validation are broadly classified as retrospective (i.e., using data from previously well-conducted, randomized, controlled trials) or prospective (enrichment, allcomers or adaptive). We believe that the systematic evaluation and implementation of these design strategies are essential to accelerate the clinical validation of biomarker-guided therapy, thereby taking us a step closer to the goal of personalized medicine.

Keywords: adaptive design • allcomers design • biomarker • enrichment design • hybrid design • randomized controlled trial

Medical treatment for oncology patients is driven by a combination of the expected outcome for the patient (prognosis) and the ability for treatment to improve the expected outcome (prediction). Biomarkers aid this process through the estimation of disease-related patient trajectories (i.e., prognostic signatures) and/or by the prediction of patient-specific outcome to treatments [1–9]. Stewart *et al.* studied the impact of subpopulation characteristics on overall study outcomes through a series of simulation studies [10]. The authors concluded that although molecular profiling is expensive, not doing so can be far more expensive and can lead to incorrect conclusions.

The term 'biomarker' in oncology refers to a broad range of markers, including biochemical markers, cellular markers, cytokine markers, genetic markers, physiological results, radiological measurements, physical signs and pathological assessment. In the case of genetic markers, the pharmacogenetic determinants of efficacy and toxicity for many anticancer drugs remain unknown. A common approach to understand the genetic determinants of efficacy and toxicity is to look for molecular markers in the tumor itself. Another emerging area is the evaluation of allelic variants in genes coding for drug targets, transporters and metabolic enzymes. This pharmacogenetic approach is particularly important in the evaluation of drug toxicity, but it also has some utility in efficacy prediction. In this article we limit our discussion to tumor markers.

A prognostic marker is a single trait, or signature of traits, that separates a population with respect to the outcome of interest in the absence of treatment, or regardless of (standard) treatment. It is associated with the disease or the patient and not with a specific therapy [11]. Prognostic marker validation can thus be established using the marker and outcome data from a cohort of uniformly treated patients

Sumithra J Mandrekar* & Daniel J Sargent

Division of Biomedical Statistics & Informatics, Mayo Clinic, Rochester, MN 55905, USA *Author for correspondence: Tel.: +1 507 266 6724 Fax: +1 507 266 2477 E-mail: mandrekar.sumithra@mayo.edu



Review: Clinical Trial Methodology Mandrekar & Sargent

with adequate follow-up. A predictive marker, on the other hand, is a single trait or signature of traits that separates a population with respect to the outcome of interest in response to a particular treatment. Designs for predictive marker validation are inherently complex and are the focus of this review article [12].

The use of a randomized controlled trial (RCT) as opposed to a cohort or single-arm study is fundamentally essential for initial, as well as definitive, predictive marker validation, for the following reasons:

- RCTs assure that patients who are treated with the agent for which the marker is purported to be predictive are comparable to those who are not;
- Changes in patient population based on biologic subsetting and/or evolution in imaging technologies can make comparisons against historical controls inaccurate;
- RCTs are essential for making the distinction between a prognostic and predictive marker [13];
- RCTs provide the opportunity to assess multiple promising therapies (and multiple possible markers) for a given disease simultaneously in a Phase II setting.

In the absence of a RCT, it is impossible to isolate any causal effect of the marker on therapeutic efficacy from the multitude of other factors that may influence the decision to treat or not treat a patient. For instance, a cohort of nonrandomized patients was used to evaluate the predictive utility of tumor microsatellite instability for the efficacy of 5-fluorouracil-based chemotherapy in colon cancer. In this cohort, the median age of the treated patients was 13 years younger than those of the nontreated patients, thus rendering any meaningful statements about the predictive value of the marker impossibly confounded [14].

Another important component of biomarker validation relates to biomarker assay issues, including the choice of using a central facility versus local laboratories for patient selection [11,15]. This choice depends on three factors:

- The reliability and reproducibility of the assay;
- The complexity of the assay;
- The potential for a repeat assessment of the marker status (when feasible and ethically appropriate) if the results from the first assessment are questionable [11,15].

For the purposes of this review, we will assume that the issues surrounding technical feasibility, assay performance metrics and the logistics of specimen collection are resolved and that initial results demonstrate promise with regard to the predictive ability of the marker(s). This review is organized as follows:

- Review of the design strategies for initial marker validation (i.e., Phase II setting);
- Review of trial designs for definitive marker validation, along with a discussion of the relative merits and limitations of each design and a comparison of the designs;
- Anticipated future state of clinical trial designs for marker validation;
- Executive summary.

Examples of real clinical trials, where available, will be used to illustrate the design concepts.

Initial validation: Phase II testing

Phase II clinical trials are designed primarily to identify promising experimental regimens that are then tested further in definitive Phase III trials. Trial designs in the Phase II setting for initial marker validation can be classified under enrichment, allcomers or adaptive design categories, elaborated below.

Enrichment designs

An enrichment design screens patients for the presence or absence of a biomarker profile and then only includes patients who either have or do not have the profile in the clinical trial [12,16]. The goal of these designs is to understand the safety, tolerability and clinical benefit of the treatment within the patient subgroup determined by a specific marker status. This design is based on the paradigm that not all patients will benefit from the study treatment under consideration, but rather that the benefit will be restricted to a biomarker-defined subgroup of patients. N0923 is an example of a Phase II trial following an enrichment design strategy. This is a randomized double-blinded Phase II study of NTX-010, a replication-competent picornavirus, after standard platinum-containing cytroreductive induction chemotherapy in patients with extensive stage small-cell lung cancer (Figure 1).

Allcomers (stratified by marker status) designs

In this design, all patients meeting the eligibility criteria, which does not include the biomarker status in question, are entered [12,17]. The ability to provide adequate tissue may be an eligibility criterion for these designs, but not the specific biomarker result, or the status of a biomarker characteristic [12].

Adaptive designs

Adaptive design strategies are a class of randomized Phase II designs by which a variety of marker signatures and drugs can be tested under one umbrella protocol. In these designs, the success of the drug-biomarker subgroup is assessed in an ongoing manner, which allows either the randomization ratio to be altered to place more patients on the most promising arm(s) and/or the underperforming drugs and/or the biomarker subgroups are eliminated midway through the trial. Key requirements for adaptive designs include:

A rapid and reliable end point, which can be somewhat challenging in the oncology setting where time to event end points or end points that involve following a patient's status for a predetermined time period (such as the progression status at 2 years) are typically used;



Figure 1. Design of N0923, a Phase II trial following an enrichment strategy. SCLC: Small-cell lung cancer.

Real time access to all clinical and biologic data, which can be a

daunting task in multicenter trials at the current time, but may not be a rate-limiting step in the future, as outlined in the Future perspective section of this review.

Examples of Phase II trials that have utilized or are utilizing an adaptive design strategy are I-SPY 2 and The BATTLE trial [18,19]. I-SPY 2 is an ongoing neoadjuvant trial in breast cancer that is designed to compare the efficacy of standard therapy to the efficacy of novel drugs in combination with chemotherapy. All drugs will be evaluated within the biomarker-defined signature groups. Regimens that have a high predicted probability of being successful in a Phase III trial are moved forward to Phase III testing within subpopulations corresponding to the most promising biomarker signature(s). Regimens that have a low probability of efficacy for all biomarker signature subgroups will be dropped from further development [18].

The BATTLE trial is complete and used an outcomebased adaptive randomization design for randomizing patients to treatment choices based on multiple biomarker profiles in non-small-cell lung cancer. Patients had their tumors tested for 11 different biomarkers and subsequently categorized into one of five biomarker subgroups and then randomized to one of four treatment choices. The first 97 patients were assigned using a balanced randomization to one of the four treatments equally. All subsequent patients were adaptively randomized, where the randomization rate was proportional to the marginal posterior 8-week disease control rate. The results from the BATTLE trial showed, as hypothesized, that each drug works best for patients with a specific molecular profile [19,20]. Two successor trials, BATTLE 2 and BATTLE 3, are currently in development, both following an adaptive design strategy. More details on the BATTLE and I-SPY 2 trials can be found in Zhou *et al.* [19,20] and Barker *et al.* [18], respectively.

Table 1 lists some of the key considerations when deciding between enrichment versus allcomers versus adaptive designs in a Phase II setting [13]. The four main components include the marker prevalence, strength of the preliminary evidence, the assay reliability and validity and turnaround times for marker assessment [13]. These are discussed in more detail below.

Enrichment designs are clearly appropriate when there is compelling preliminary evidence to suggest benefit only in a marker-defined subgroup(s) and/or when the marker prevalence is low (<10–20%). Under these circumstances, it is not feasible to use an allcomers strategy as the treatment effect in the overall population will be diluted, thus requiring a prohibitively large sample size. For enrichment designs, it is also essential to have an established assay with good performance and short turnaround times for marker assessment [12].

An allcomers design is appropriate when:

- The preliminary evidence is unclear and the marker prevalence is high (≥50%) and/or;
- The assay performance is not well established (i.e., no established cut-off point for marker status definition) and/or;
- The turnaround time for marker assessment is long (e.g., more than 1 week in second- or third-line treatment settings) [13].

Review: Clinical Trial Methodology Mandrekar & Sargent

Table 1. Criteria for choice of design for initial marker validation trials.			
Criteria	Design		
	Enrichment	Allcomers	Adaptive
Preliminary evidence			
Strongly suggest benefit in marker-defined subgroups.	Optimal	Not recommended	Appropriate (assess multiple treatments/ biomarker subgroups)
Uncertain about benefit in overall population versus marker-defined subgroups	Not recommended	Appropriate	Appropriate (learn and adapt as the trial proceeds)
Assay reproducibility and validity			
Excellent (high concordance between local and central testing; commercially available kits, and so forth)	Required	Not recommended	Required
Questionable	Not recommended	Appropriate	Not applicable
Turnaround times			
Rapid (2–3 days; without causing delay in the start of therapy)	Optimal	Optimal	Optimal
Slow to modest (1 week or more)	Not recommended	Appropriate (retrospective marker subgroup assessment)	Appropriate in some cases
Marker prevalence			
Low (<20%)	Optimal	Not recommended	Appropriate
Moderate (20–50%)	Appropriate	Appropriate (stratified by marker status)	Appropriate
High (>50%)	Appropriate	Appropriate	Appropriate

In most instances however, an allcomers design should incorporate a prospectively specified subgroup analysis of the treatment effect within biomarkerdefined subgroups. This is critical to ensure that the effect of the drug is tested both on the overall population as well as prospectively defined subsets of patients so as to not incorrectly conclude that the drug is ineffective, when it may be effective for a smaller subset of the population [12].

In cases where the prevalence of the marker in question is moderate (between 20–50%), then a possible strategy could be as follows: first, perform a single-arm enrichment trial (pilot) as a proof-of-concept that the treatment probably has a major effect within the marker subgroup. Second, based on the data from the pilot trial, perform an allcomers Phase II (randomized) trial, using either a trial stratified by marker status, with the primary hypothesis defined within the marker subgroup hypothesized to derive the most benefit. Accrue sufficient patients to the other subgroup(s) to demonstrate lack of benefit or an adaptive design where the relationship between markers to treatment success is assessed in an ongoing manner.

Definitive validation: Phase III setting

Prospectively designed, RCTs are the 'gold standard' approach to validating a predictive marker. In some cases, the possibility to test the predictive ability of a marker using data from previously well-conducted RCT comparing therapies for which a marker is proposed to be predictive can be a more feasible and timely option. Frequently, a complete understanding of the biology prior to the testing of a therapy (and even approval of the therapy in some cases) is not possible. Thus, therapies that benefit only a subset of patients may still result in an overall benefit; however, once a therapy is approved for common use, designs that randomize patients to not use that therapy become exceedingly difficult. Retrospective validation can aid in such situations by bringing forward effective treatments to marker-defined patient subgroups [12]. The important components of a retrospective validation are summarized in Box 1. In particular, a prospectively specified retrospective validation using data from multiple independent RCTs can provide strong evidence for a robust predictive effect [12].

An example of a successful retrospective validation is the establishment of mutant KRAS status as a predictor of lack of efficacy from panitumumab and cetuximab therapy in advanced colorectal cancer. This marker was first identified in single-arm trials after nontargeted Phase III RCTs had been completed [21-23]. A prospective KRAS analysis plan was specified and tested using the data from the multiple retrospective RCTs. The percentage of study populations for which KRAS status was assessed in these trials ranged from as low as 23%, to as high as 92%. The results consistently demonstrated that the benefit from panitumumab and cetuximab is restricted to patients with wild-type KRAS status, with mutant KRAS patients deriving no clinical benefit [23]. Based on this strong evidence, all ongoing clinical trials with these agents in colorectal cancer sponsored by the US National Cancer Institute (NCI) were amended to only include KRAS wild-type patients. Moreover, labeling changes have been implemented in the indications and usage, clinical pharmacology and clinical studies section of both panitumumab and cetuximab product labels by the US FDA. Specifically, the indications and usage labeling for these agents state that the use of cetuximab or panitumumab is not recommended for the treatment of colorectal cancer in patients with KRAS mutations in codon 12 or 13.

While retrospective validation may be acceptable as a marker validation strategy in circumstances such as those detailed above, the gold standard for predictive marker validation continues to be a prospective RCT. Several designs have been proposed and utilized in the field of cancer biomarkers for the prospective validation of predictive markers. These designs are discussed in further detail below and can be classified briefly as:

- Targeted or enrichment designs;
- Allcomers designs, which are further classified as hybrid designs, marker by treatment interaction designs and sequential testing strategy designs;
- Adaptive designs.

Targeted or enrichment designs

As discussed in the section 'Initial validation: Phase II testing', this design is based on the paradigm (when there is compelling preliminary evidence) that not all patients will benefit from the study treatment under consideration, but rather that the benefit will be restricted to a subgroup of patients who express (or do not express) a specific molecular feature [12,16]. Consequently, all patients are screened for the presence or absence of a marker profile and only those with (or without) the profile are included in the trial. Prior to the launching of a trial with an enrichment design strategy, the assay reproducibility, accuracy and turnaround times

Box 1. Requirements for a valid retrospective assessment of a predictive biomarker.

- Clinical and biomarker data from a well-conducted randomized, controlled trial.
- Established analytical and clinical validity of the assay.
- Availability of samples on a large majority of patients to avoid selection bias.
- Prospectively stated hypothesis, sample size and power calculations, analytical techniques and patient subpopulations.

for marker assessment must be well-established. As a general guideline, such designs are appropriate when:

- Therapies have modest absolute benefit in the unselected population, but cause significant toxicity;
- In the absence of selection, therapeutic results are similar whereby a selection design (even if incorrect) would not hurt;
- An unselected design is ethically impossible [12].

An enrichment design strategy of enrolling only HER2-positive patients (based on a local assessment of HER2 status) demonstrated that trastuzumab (i.e., Herceptin®) combined with paclitaxel after doxorubicin and cyclophosphamide, significantly improved disease-free survival among women with surgically removed HER2-positive breast cancer [24]. Subsequent analyses raised questions regarding the assay reproducibility based on local versus central testing for HER2 status [25,26]. As only patients deemed HER2-positive based on the local assessment were enrolled and tissue from patients deemed HER2-negative were not collected, the question of whether trastuzumab therapy benefits a potentially larger group than the approximately 20% of patients defined as HER2-positive in these two trials is the subject of an ongoing trial [27].

Another example of an enrichment design is the ongoing national cooperative group cancer trial N0577-Phase III intergroup study of radiotherapy versus temozolomide alone versus radiotherapy with concomitant and adjuvant temozolomide for patients with 1p/19q codeleted anaplastic glioma. In this trial, the 1p/19q status of the patient is assessed centrally (to address issues regarding standardization of assay techniques, reproducibility and interpretability of assay results) after which eligible patients are randomized to one of three treatment arms:

- Arm A: radiation therapy alone (the control arm);
- Arm B: temozolomide concomitant with radiation therapy followed by adjuvant temozolomide;
- Arm C: temozolomide alone (Figure 2).

fsg future science group

There is abundant evidence in the literature demonstrating that this subgroup of patients is more responsive to treatment and need to be studied separately from the cohort of patients without this co-deletion [28–30]. At the present time, it remains unclear whether the 1p and 19q deletions simply represent a molecular signature in this patient population and thus reflects a favorable natural biological behavior, or whether these markers are mechanistically related to response to therapy. This trial is designed to address the question of the optimal treatment strategy for the patients with this co-deletion.

Allcomers design

Hybrid designs

In this design strategy, only a certain subgroup of patients based on their marker status are randomized between treatments, whereas patients in the other marker-defined subgroups are assigned the standard of care treatment(s) [12]. This design is an appropriate choice when there is compelling evidence demonstrating the efficacy of a certain treatment(s) for a marker-defined subgroup, thereby making it unethical to randomize patients with that particular marker status to other treatment options. However, unlike the enrichment design strategy, all patients, regardless of the marker status, are enrolled and followed. This provides the possibility for future testing for other potential prognostic markers. At least three recent or ongoing oncology marker validation trials have utilized the hybrid design strategy [31-33]:

Phase III randomized study of oxaliplatin, leucovorin calcium and fluorouracil with bevacizumab versus





without in patients with resected stage II colon cancer and at high risk for recurrence based on molecular markers (Eastern Cooperative Oncology Group 5202);

- The TAILORx trial designed to evaluate the Oncotype Dx (Genomic Health, Redwood City, CA, USA), a 21-gene recurrence score in tamoxifen-treated breast cancer patients;
- The MINDACT trial for node-negative breast cancer patients designed to evaluate MammaPrint (Agendia, Amsterdam, The Netherlands), the 70-gene expression profile discovered at the Netherlands Cancer Institute.

Marker by treatment interaction design

In this design, all patients meeting the eligibility criteria are entered into the trial [17]. The ability to provide adequate tissue may be an eligibility criterion, but not the specific biomarker result [12]. The marker by treatment interaction design uses the marker status as a stratifi -cation factor and randomizes patients to treatment choices within each marker-based subgroup. While this is similar to conducting two independent RCTs under one large RCT umbrella, it differs from a single large RCT in two essential characteristics. First, only patients with a valid marker result are randomized, and second, there is a prospective sample size specification for each marker-based subgroup.

The sample size planning for treatment-by-marker interaction design is based on the prespecified analysis plan. A separate evaluation of the treatment effect can be tested in the two marker-defined subgroups, or a

> test of interaction can be carried out first. Different sequential analysis plans can also be implemented. For example, when the primary test of interaction is not significant at a prespecified significance level, then the treatment arms can be compared in the overall population (ignoring the biomarker status). If the interaction is significant, then the experimental treatment can be compared with the control arm within the strata determined by the marker status.

Sequential testing strategy designs

Sequential testing designs are similar in principle to a RCT design [34-36]. These designs have a single primary hypothesis, which is either tested in the overall population first and then in a prospectively planned subset if the overall test is not significant, or in the marker-defined subgroup first and then tested in the entire population if the subgroup analysis is significant. The first is recommended in cases where the experimental treatment is hypothesized to be broadly effective and the subset analysis is ancillary. The latter (also known as the closed testing procedure) is recommended when there is strong preliminary data to support that the treatment effect is strongest in the marker-defined subgroup and that the marker has sufficient prevalence that the power for testing the treatment effect in the subgroup is adequate. This strategy is largely driven by three statistical parameters:

- α the type I error or probability of a false-positive result,
- β the type II error or probability of a false-negative result;
- δ the targeted difference or targeted effect size.

The sequential testing strategy designs differ in the choice of the values for these statistical parameters, which are dictated by the inference framework of the design. Both of these sequential testing approaches appropriately control for the type I error rates associated with multiple testing. A modification to this approach, taking into account potential correlation arising from testing the overall treatment effect and the treatment effect within the marker-defined subgroup, has also been proposed [36].

The closed testing procedure was utilized in the Phase III trial testing cetuximab in addition to FOLFOX as adjuvant therapy in stage III colon cancer (N0147) [37]. This trial initially randomized both KRAS mutant and wild-type patients and was amended later to randomize only patients with KRAS wild-type tumors, once the data on the use of cetuximab was restricted to KRAS wild-type patients. The primary analysis was therefore conducted within the KRAS wild-type patients with the provision in the design that if the treatment effect was significant in the KRAS wild-type group, a subsequent test would be performed on all patients.

Another class of designs that follow a similar sequential testing strategy is the adaptive threshold and the adaptive signature designs [38–40]. The former is used in situations where a marker is known at the start of the trial, but a cut-off point for defining marker-positive and marker-negative groups is not known. The latter is used when the marker and the threshold are both unknown at the start of the trial and the design allows for the 'discovery and validation' process of the marker within the realm of the single Phase III trial, using either a cross validation approach or the split-alpha approach [39,40]. The adaptive threshold design can be implemented one of two ways:

- The new treatment is compared with the control in all patients at a prespecified significance level and if not significant, a second stage analysis involving finding an 'optimal' cut-off point for the predictive marker is performed using the remaining alpha, or;
- Under the assumption that the treatment is effective only for a marker-driven subset, no overall treatment to control comparisons are made, instead, the analysis focuses on the identification of optimal cut points.

Both these approaches were concluded to be superior (in terms of the power and number of events required to detect an effect at a prespecified overall type I error rate) to the classic-nonadaptive design approaches in the simulation studies [38]. Two issues need further consideration with such designs:

- The added cost of a somewhat larger sample size and/ or redundant power dictated by the strategy of partitioning the overall type I error rate, and;
- Use of data from the same trial to both define and validate a marker cut-off point.

The adaptive signature design uses the first approach above, where the new treatment is compared with the control in all patients at a prespecified significance level. If this overall comparison is significant, then it is taken that the treatment is broadly effective. If, however, the overall comparison is not significant, a second stage analysis is undertaken for the development and use of a biomarker signature, using a split sample or a cross-validated approach [39,40].

Adaptive designs

Clinical trials utilizing adaptive design strategies in the Phase II setting are described in the section 'Initial validation: Phase II testing'. There are currently no NCIsupported definitive Phase III trials that utilize these adaptive strategies in oncology. A number of innovative statistical designs have recently been proposed that use either an adaptive strategy for analysis, or an outcomebased adaptive randomization. We review them briefly here for completeness.

The adaptive accrual design outlines a strategy to adaptively modify accrual to two predefined marker-defined subgroups based on an interim futility analysis [41]. Specifically, the trial follows the following scheme:

- Begin with accrual to both marker-defined subgroups;
- At the interim analysis, if the treatment effect in one of the subgroups fails to satisfy a futility boundary, terminate accrual to that subgroup;
- Continue accrual to the other subgroup until the planned total sample size is reached, including

accruing subjects that had planned to be included from the terminated subgroup.

This design has demonstrated greater power than a nonadaptive trial in simulation settings; however, this strategy might lead to a substantial increase in the accrual duration depending on the prevalence of the marker for the subgroup that continues to full accrual. In addition, the futility boundary is somewhat conservative and less than optimal as it is set to be in the region where the observed efficacy is greater for the control arm than the experimental regimen. Another design to adaptively modify accrual was proposed by Liu et al. [42]. In this design, only the marker-positive patients are accrued in the first stage. If the interim analysis shows promising results for the marker-positive cohort, then the second stage would continue accrual to the marker-positive cohort, but also include marker-negative patients. If the first stage shows no benefit in the marker-positive cohort, then the trial is closed permanently.

Future perspective

In this section, we speculate on the anticipated state of clinical trial designs for marker validation in the next 5–10 years. First, with technological advancement (mobile computing, electronic data capture, integration of research records with electronic medical records), we believe that real-time access to data will become a reality, even in multicenter trials, allowing adaptive designs to take on a much greater role in clinical trials. Second, a better understanding of the tumor biology (e.g., identifying patient subsets and rare tumor subtypes),

advancement in assay techniques and availability of commercial kits with rapid turnaround times will lead to the popularity of enrichment designs. Third, tailored treatments with effective biomarker-driven hypotheses will lead to smaller clinical trials targeting larger treatment effects. Finally, Phase II/III designs will grow in popularity as small patient subsets will require us to not 'waste' patients [43,44]. This class of integrated Phase II/ III designs (also known as the multiarm multi-stage designs) enable the simultaneous assessment of multiple experimental agents against the standard of care in the Phase II portion using an intermediate (or surrogate) end point. This eliminates the need to conduct separate (large-scale) Phase II trials to evaluate each experimental regimen. The Phase III portion will subsequently continue with the promising experimental arms from the Phase II portion, comparing them to the standard of care. GOG-182 is an example of an NCI-funded cooperative group trial that utilized the multiarm multistage design. This was a five-arm trial in advanced stage ovarian cancer or primary peritoneal carcinoma [45].

Financial & competing interests disclosure

Supported in part by the National Cancer Institute Grants: Mayo Clinic Cancer Center (CA-15083) and the North Central Cancer Treatment Group (CA-25224). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

- Biomarker identification is a critical component of targeted oncology drug development.
- A randomized, controlled trial is fundamentally essential for both the initial as well as definitive marker validation.
- Retrospective validation following the guidelines outlined in Box 1 can help to bring forward effective treatments to marker-defined patient subgroups in some situations.
- Prospective (initial and definitive) marker validation trials can be categorized into enrichment, allcomers (hybrid, marker by treatment interaction and sequential testing strategy designs) and adaptive designs.
- The choice of a clinical trial design for marker validation depends on the marker prevalence, strength of the preliminary evidence, the assay reliability and validity, and turnaround times for marker assessment.

References

Papers of special note have been highlighted as:

- of interest
- of considerable interest
- Sequist LV, Bell DW, Lynch TJ et al. Molecular predictors of response to epidermal growth factor receptor antagonists in non-small-cell lung cancer. J. Clin. Oncol. 25(5), 587–595 (2007).
- 2 Bonomi PD, Buckingham L, Coon J. Selecting patients for treatment with

epidermal growth factor tyrosine kinase inhibitors. *Clin. Cancer Res.* 13(15 Pt 2), \$4606-\$4612 (2007).

- 3 Amado RG, Wolf M, Peeters M et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. J. Clin. Oncol. 26(10), 1626–1634 (2008).
- 4 Augustine CK, Yoo JS, Potti A *et al.* Genomic and molecular profiling predicts response to temozolomide in melanoma. *Clin. Cancer Res.* 15(2), 502–510 (2009).
- 5 Riedel RF, Porrello A, Pontzer E et al. A genomic approach to identify molecular pathways associated with chemotherapy resistance. *Mol. Cancer Ther.* 7(10), 3141–3149 (2008).
- 6 Garman KS, Acharya CR, Edelman E et al. A genomic approach to colon cancer risk stratification yields biologic insights into therapeutic opportunities. Proc. Natl Acad. Sci. USA 105(49), 19432–19437 (2008).
- Bonnefoi H, Potti A, Delorenzi M *et al.* Validation of gene signatures that predict the

response of breast cancer to neoadjuvant chemotherapy: a substudy of the EORTC 10994/BIG 00–01 clinical trial. *Lancet Oncol.* 8(12), 1071–1078 (2007).

- 8 Kerr D, Gray R, Quirke P et al. A quantitative multigene RT–PCR assay for prediction of recurrence in stage II colon cancer: selection of the genes in four large studies and results of the independent, prospectively designed QUASAR validation study. J. Clin. Oncol. 27(15s), Abstr. 4000 (2009).
- Mandrekar SJ, Sargent DJ. Predictive biomarker validation in practice: lessons from real trials. *Clin. Trials* 7(5), 567–573 (2010).
- Stewart DJ, Whitney SN, Kurzrock R. Equipoise lost: ethics, costs, and the regulation of cancer clinical research. J. Clin. Oncol. 28(17), 2925–2935 (2010).
- Highlights the need for well-conducted biomarker-driven trial designs and the tradeoffs in terms of complexity, costs and patient care.
- 11 Mandrekar SJ, Sargent DJ. Genomic advances and their impact on clinical trial design. *Genome Med.* 1(7), 69 (2009).
- 12 Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. J. Clin. Oncol. 27(24), 4027–4034 (2009).
- Comprehensive review of trial designs for prospective validation of biomarkers.
- 13 Mandrekar SJ, Sargent DJ. Allcomers versus enrichment design strategy in Phase II trials. *J. Thorac Oncol.* 6(4), 658–660 (2011).
- 14 Elsaleh H, Joseph D, Grieu F et al. Association of tumour site and sex with survival benefit from adjuvant chemotherapy in colorectal cancer. Lancet 355, 1745–1750 (2000).
- 15 Moore HM, Kelly AB, Jewell SD *et al.* Biospecimen Reporting for Improved Study Quality (BRISQ). *J. Proteome. Res.* 119, 92–101 (2011).
- 16 Maitournam A, Simon R. On the efficiency of targeted clinical trials. *Stat. Med.* 24(3), 329–339 (2005).
- 17 Sargent DJ, Conley BA, Allegra C et al. Clinical trial designs for predictive marker validation in cancer treatment trials. J.Clin. Oncol. 23(9), 2020–2027 (2005).
- One of the first articles to introduce and discuss markers by treatment interaction design for prospective validation of biomarkers.

- 18 Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin. Pharmacol. Ther.* 86(1), 97–100 (2009).
- 19 Zhou X, Liu S, Kim ES. Bayesian adaptive design for targeted therapy development in lung cancer – a step towards personalized medicine. *Clinical Trials* 5, 181–193 (2008).
- Comprehensive overview of the BATTLE trial design.
- 20 Kim ES, Herbst RS, Wistuba II *et al.* The BATTLE trial: personalizing therapy for lung cancer. *Cancer Disc.* 1(44), 44–53 (2011).
- 21 Jonker DJ, O'Callaghan CJ, Karapetis CS et al. Cetuximab for the treatment of colorectal cancer. N. Engl. J. Med. 357, 2040–2048 (2007).
- 22 Karapetis CS, Khambata-Ford S, Jonker DJ et al. K-RAS mutations and benefit from cetuximab in advanced colorectal cancer. N. Engl. J. Med. 359(17), 1757–1765 (2008).
- 23 Van Cutsem E, Köhne CH, Hitre E et al. Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N. Engl. J. Med.* 360(14), 1408–1417 (2009).
- 24 Romond EH, Perez EA, Bryant J et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N. Engl. J. Med. 353(16), 1673–1684 (2005).
- 25 Perez EA, Suman VJ, Davidson NE et al. HER2 testing by local, central, and reference laboratories in specimens from the North Central Cancer Treatment Group N9831 intergroup adjuvant trial. J. Clin. Oncol. 24(19), 3032–3038 (2006).
- 26 Paik S, Kim C, Wolmark N. HER2 status and benefit from adjuvant trastuzumab in breast cancer. *N. Engl. J. Med.* 358(13), 1409–1411 (2008).
- Discusses the pitfalls when applying the enrichment design strategy to a less than optimal validated assay.
- 27 Hayes DF. Steady progress against HER2-positive breast cancer. N. Engl. J. Med. 365, 14 (2011).
- 28 Cairncross G, Berkey B, Shaw E et al. Phase III trial of chemotherapy plus radiotherapy compared with radiotherapy alone for pure and mixed anaplastic oligodendroglioma: Intergroup Radiation Oncology Group Trial 9402. J. Clin. Oncol. 24, 2707–2714 (2006).

- 29 Jenkins RB, Blair H, Ballman KV et al. A t(1;19)(q10;p10) mediates the combined deletions of 1p and 19q and predicts a better prognosis of patients with oligodendroglioma. *Cancer Res.* 66(20), 9852–9861 (2006).
- 30 Cairncross G, Jenkins R. Gliomas with 1p/19q codeletion: a.k.a. oligodendroglioma. *Cancer J.* 14(6), 352–357 (2008).
- 31 Sparano JA, Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials. *J. Clin. Oncol.* 26(5), 721–728 (2008).
- 32 Cardoso F, Van't Veer L, Rutgers E *et al.* Clinical application of the 70-gene profile: The MINDACT trial. *J. Clin. Oncol.* 26(5), 729–735 (2008).
- 33 Bogaerts J, Cardoso F, Buyse M et al. TRANSBIG consortium. Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. Nat. Clin. Pract. Oncol. 3(10), 540–551 (2006).
- 34 Simon R, Wang SJ. Use of genomic signatures in therapeutics development. *Pharmacogenomics J.* 6, 1667–1673 (2006).
- 35 Bauer P. Multiple testing in clinical trials. Stat. Medicine 10, 871–890 (1991).
- 36 Song Y, Chi GYH. A method for testing a prespecified subgroup in clinical trials. *Stat. Medicine* 26, 3535–3549 (2007).
- 37 Alberts SR, Sinicrope FA, Grothey A. N0147: a randomized Phase III trial of oxaliplatin plus 5-fluorouracil/leucovorin with or without cetuximab after curative resection of stage III colon cancer. *Clin. Colorectcal. Cancer* 5(3), 211–213 (2005).
- 38 Jiang W, Freidlin B, Simon R. Biomarkeradaptive threshold design: a procedure for evaluating treatment with possible biomarkerdefined subset effect. *J. Natl Cancer Inst.* 99(13), 1036–1043 (2007).
- 39 Freidlin B, Simon R. Adaptive signature design: An adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clin. Cancer Res.* 11(21), 7872–7878 (2005).
- 40 Freidlin B, Jiang W, Simon R. The crossvalidated adaptive signature design. *Clin. Cancer Res.* 16(2), 691–698 (2010).
- Innovative approach to the discovery and validation of a biomarker signature in the context of a standard Phase III trial.
- 41 Wang SJ, O'Neill RT, Hung HMJ. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. *Pharm. Stat.* 6, 227–244 (2007).

Review: Clinical Trial Methodology Mandreka

Mandrekar & Sargent

- 42 Liu A, Liu C, Li Q, Yu KF, Yuan VW. A threshold sample-enrichment approach in a clinical trial with heterogeneous subpopulations. *Clin. Trials* 7(5), 537–545 (2010).
- 43 Hunsberger S, Zhao Y, Simon R. A comparison of Phase II study strategies. *Clin. Cancer Res.* 15(19), 5950–5955 (2009).
- 44 Parmar MK, Barthel FM, Sydes M et al. Speeding up the evaluation of new agents in cancer. J. Natl Cancer Inst. 100(17), 1204–1214 (2008).
- •• Comprehensive assessment of Phase II/III design strategies and its place in the current era of multiple targets and multiple treatments.
- 45 Copeland LJ, Bookman M, Trimble E. Gynecologic Oncology Group protocol GOG 182-ICON5. Clinical trials of newer regimens for treating ovarian cancer: the rationale for Gynecologic Oncology Group Protocol GOG 182-ICON5. *Gynecol. Oncol.* 90(2 Pt 2), S1–S7 (2003).