Published in final edited form as:

Cancer Treat Rev. 2016 February ; 43: 74–82. doi:10.1016/j.ctrv.2015.12.008.

Clinical trial designs incorporating predictive biomarkers[☆]

Lindsay A. Renfro^{a,*}, Himel Mallick^{b,c}, Ming-Wen An^d, Daniel J. Sargent^a, and Sumithra J. Mandrekar^a

^aDivision of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA

^bDepartment of Biostatistics, Harvard School of Public Health, Boston, MA, USA

^cThe Broad Institute of MIT and Harvard, Cambridge, MA, USA

^dDepartment of Mathematics and Statistics, Vassar College, Poughkeepsie, NY, USA

Abstract

Development of oncologic therapies has traditionally been performed in a sequence of clinical trials intended to assess safety (phase I), preliminary efficacy (phase II), and improvement over the standard of care (phase III) in homogeneous (in terms of tumor type and disease stage) patient populations. As cancer has become increasingly understood on the molecular level, newer “targeted” drugs that inhibit specific cancer cell growth and survival mechanisms have increased the need for new clinical trial designs, wherein pertinent questions on the relationship between patient biomarkers and response to treatment can be answered. Herein, we review the clinical trial design literature from initial to more recently proposed designs for targeted agents or those treatments hypothesized to have enhanced effectiveness within patient subgroups (e.g., those with a certain biomarker value or who harbor a certain genetic tumor mutation). We also describe a number of real clinical trials where biomarker-based designs have been utilized, including a discussion of their respective advantages and challenges. As cancers become further categorized and/or reclassified according to individual patient and tumor features, we anticipate a continued need for novel trial designs to keep pace with the changing frontier of clinical cancer research.

Keywords

Clinical trial design; Adaptive trial design; Biomarker-based design; Bayesian adaptive design; Enrichment designs; Targeted therapies

---

[☆]*Research support:* This publication was supported by CTSA Grant No. KL2 TR000136 from the National Center for Advancing Translational Science (NCATS). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Funding was also provided by R01 CA174779/CA/NCI NIH HHS/United States.

*Corresponding author at: Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA. Tel.: +1 (507) 284 3202; fax: +1 (507) 266 2477. renfro.lindsay@mayo.edu (L.A. Renfro).

Conflict of interest statement

The authors have no conflicts of interest to disclose.

Introduction

Historical clinical trial paradigm and advent of targeted therapies

As cancer has become increasingly understood on the molecular level, therapeutic research has largely shifted from a focus on cytotoxic agents to newer drugs that inhibit specific cancer cell growth and survival mechanisms or that enhance immune responses to cancer cells. Increasingly common are trials of targeted therapies intended to show enhanced efficacy in patient subpopulations, such as those with a known biomarker value or genetic tumor mutation. For example, panitumumab and cetuximab have been indicated as treatment options for advanced colorectal cancer patients with KRAS wild-type tumors [1–2], and therapies targeting epidermal growth factor receptor mutation have improved outcomes in a subset of patients with advanced non-small-cell lung cancer [3–4].

New clinical trial design paradigm for therapies targeting patient subsets

In the past decade, a number of biomarker-based design solutions have been proposed to study treatments within possibly heterogeneous patient subpopulations. These can be broadly classified on several levels. First, clinical trials for targeted therapies may be generally classified as follows: “phase I” trials, where the marker and treatment are studied together in normal versus tumor tissue, the assay validated, and any relevant marker positivity thresholds tentatively selected; phase II trials, where interest lies in identifying and possibly validating a marker-based subpopulation where efficacy of a targeted therapy is most promising; and phase III trials, which generally entails a usual randomized treatment comparison in the population identified and believed to benefit from earlier phase II studies [5]. Marker-based trial designs may further be classified as retrospective (evaluation of the marker-treatment-outcome relationship after the trial has been completed) or prospective (formal incorporation of predictive markers in the design considerations), where the latter is typically required for clinical validation. A third classification of biomarker designs is a purely statistical one: frequentist or “classical” designs versus Bayesian designs, where differences between the two approaches lie primarily in the methods for hypothesis testing, decision-making, and use of prior (or historical) information.

In this review of biomarker-based trial methodology, we focus on prospective trial designs, both classical and Bayesian, with emphasis given to phase II and III studies where discovery, clinical validation, or subsequent use of a predictive biomarker are the primary objectives (Early literature on biomarker designs and A movement toward adaptive designs sections). Of importance but not covered in this review are the earlier stages of biomarker development, such as construction and assay validation of genomic signature classifiers or creation of diagnostic tests meant to detect patients with potentially enhanced treatment sensitivity. Selected case studies: implementation of biomarker-based designs in oncology section presents several recent or ongoing biomarker-based trials as case studies, and Going forward: future design challenges section concludes with a discussion of areas of future need for biomarker-based designs.

Early literature on biomarker designs

Targeted or enriched design

Among the earliest explorations of biomarker-based clinical trial designs were those of Simon and Maitournam [6–7], who compared conventional trials randomizing all patients with a particular disease to those in which only patients positive for a particular biomarker were randomized to experimental versus control treatments (i.e., “targeted” or “enriched” designs; Fig. 1A). Relative efficiency in terms of sample size was reported as a function of marker prevalence and differential treatment effects between marker-positive and marker-negative patients, taking the number of patients screened for eligibility into account. Using a genomic classifier to exclude patients from eligibility of a study requires a substantial level of confidence in the classifier, and a reproducible assay with a high level of sensitivity and specificity. In cases where equipoise is insufficient to ethically randomize marker-negative patients to a targeted therapy, enrichment designs may be the most ethical path forward for clinical development. An enriched non-inferiority trial design that considers misclassification error of the genomic classifier was described by Wang et al. [8], and enrichment strategies were discussed in more detail by Freidlin and Korn [9]. Examples of enrichment trials in practice include N9831 [10] and TOGA [11].

Marker-by-treatment interaction design

Often there is insufficient evidence of a biomarker’s ability to predict treatment effect to justify exclusion of a subpopulation from randomization. In this case, a marker-by-treatment interaction design (sometimes referred to as a marker stratified design) or a trial randomizing patients to experimental versus control treatments within marker-defined subgroups is an alternative approach with many advantages [12]; see Fig. 1B. Specifically, such a design may be fully powered to detect a treatment effect within each subgroup, thereby precluding false negative results in a trial sized only to detect an effect in the overall population. A marker-by-treatment interaction design may be additionally powered to detect a statistically significant biomarker-by-treatment interaction effect in a regression model for the endpoint, thereby statistically confirming the predictive ability of the biomarker [13]. To achieve these benefits, however, a marker-by-treatment interaction design often requires a relatively large sample size, as its structure resembles multiple randomized trials conducted in parallel. For this reason, it is used selectively despite its theoretical advantages, although the INTEREST [14] and MARVEL [15] trials are two such examples.

Marker-based strategy design

A design that focuses specifically on the role of a biomarker in the treatment decision-making process is the biomarker strategy design [12,16–17]. In this design, patients are randomized at the time of screening to a treatment strategy (often standard of care) that ignores the biomarker versus a strategy taking biomarker status into account, through direct assignment to targeted therapies matched to the biomarker status of each eligible patient. Primary outcome analyses are then made between treatment strategies rather than specific treatments, with the hypothesis that better outcomes will be observed among those patients treated according to (versus independent of) their biomarker status. At the same time, questions regarding the best treatment for patient subgroups may remain unanswered as

treatment randomization within marker subgroups may not occur. Example strategy trials include SHIVA [18–19] and M-PACT [20]. A modified strategy design accounting for multiple potential marker-treatment pairs (similar to SHIVA) is shown in Fig. 1C.

Sequential testing designs

A reasonable compromise between the smaller targeted and larger marker-by-treatment interaction and biomarker strategy designs is an unselected randomized design with sequential hypothesis testing in the overall and marker-positive or sensitive subpopulations where the overall false-positive error rate is controlled at a pre-specified level [21–22]. The “adaptive signature design” proposed by Freidlin and Simon (2005) is one such example with two study stages, wherein a predictive signature that is developed in the first set of patients is used to evaluate the subset treatment effect in an independent second set of patients, in the event that the overall test based on all accrued patients is negative [21]. Within this sequential testing design, sometimes referred to as a “fall-back analysis plan”, the power to detect a treatment in the positive subset may be low if the trial size is based on powering the overall analysis alone, or if marker prevalence is low. One solution is to size the trial to achieve sufficient power for the marker subgroup analysis, thereby also ensuring adequate power to detect an overall effect, at the cost of an increased sample size.

Song and Chi (2007) refined the methodology for balancing the level of type I error between multiple tests [23], and Freidlin et al. (2010) improved on the efficiency of the adaptive signature design by replacing independent signature development and statistical validation datasets with cross-validation techniques [24]. Jiang et al. [22] also extended the sequential testing framework from binary to time-to-event outcomes, including testing procedures allowing for correlation between the overall and subgroup-specific test statistics.

Addressing the issue of low power to detect marker-subgroup effects in a sequential testing trial of fixed size when the marker prevalence is low, Zhao et al. [25] presented a strategy for “enriching” or artificially increasing the proportion of marker-positive patients enrolled to a trial relative to their existence in the general population, with appropriate hypothesis testing mechanisms. Riddell et al. [26] extended the biomarker-adaptive threshold sequential testing design of Jiang et al. [22] to the setting of a biomarker and outcome where each is a discrete count. Mackey and Bengtsson [27] extended the sequential testing framework to answer three sequential questions in the setting of a randomized trial with a time-to-event endpoint, with each subsequent question requiring an affirmative answer to the one before: (1) whether any patients benefit from treatment, (2) whether any patients do *not* benefit from treatment, and (3) what is the biomarker threshold.

Comparisons of early classical biomarker designs

Hoering and colleagues performed a simulation-based comparison of several early biomarker designs, including a targeted design, a biomarker strategy design where marker-negative patients are treated with the control therapy and marker positive patients randomized to experimental (targeted) versus control therapies, and a sequential testing design randomizing all patients [28]. They concluded that no one biomarker design fits all situations, and advocated for a thorough investigation of trial properties (e.g., via

simulation) before committing to a specific design. Mandrekar and Sargent [29] replicated some of these findings, and further performed a head-to-head comparison of the marker-by-treatment interaction design and the biomarker strategy design, finding the interaction design was superior to the strategy design in terms of required sample size in most cases. Numerous other discussions of the relative advantages and limitations of these early biomarker-based designs (and some more recent designs) have been published [15,30–39]. Features, capabilities, and examples of selected designs are shown in Table 1.

A movement toward adaptive designs

After the first few years of biomarker-based trial design literature yielded the fixed designs described above, a movement toward *adaptive* biomarker-based trial designs emerged. By “adaptive”, we refer to designs utilizing data accumulated from patients early in the trial to prospectively shift accrual, eligibility, or objectives later on in the trial. The estimation and decision rules for adaptive designs may be performed within either classical or Bayesian statistical paradigms.

Classical biomarker-adaptive trial designs

Adaptive enrichment designs—Wang et al. (2007) introduced one of the first biomarker-based clinical trial designs with allowed mid-trial adaptation based on the results of interim analyses [40]. The adaptive enrichment design initially randomizes an unselected patient population to experimental versus control treatment, and if the experimental treatment effect reaches a futility threshold in the marker-negative group at an interim analysis, accrual of marker-negative patients is terminated and the remaining sample size re-allocated to marker-positive patients (Fig. 1D). In that case, the primary hypothesis tested at the trial’s conclusions is the treatment effect in the marker-positive subgroup. Otherwise, if futility is not reached in the marker-negative group at an interim analysis, the trial continues unselected and performs both overall and subgroup-specific tests of treatment benefit at the final analysis time point with trial-wise type I error control. When compared with the fixed-randomization sequential testing design of Freidlin and Simon [21], the design of Wang et al. showed uniformly greater power for detecting a subgroup-specific treatment in simulations, a direct consequence of its ability to adaptively enroll a greater proportion of marker-positive patients. However, designs without mid-trial enrichment are capable of identifying and then validating predictive marker effects in separate patient cohorts, while an adaptive enrichment design loses this ability after it stops accrual to marker-negative patients. An extension of the adaptive enrichment framework to nested patient subsets was described by Wang et al. [41], and an adaptive enrichment design for the phase III setting was proposed by Hong and Simon [42].

Another adaptive enrichment design was proposed by Brannath et al. [43], wherein enrichment to a subgroup and sample size adjustment may occur following a first and second interim analysis, respectively, with possible early stopping for efficacy or futility at each time point. A more general testing framework for adaptive enrichment was described by Mehta and Gao [44]; specifically, a group sequential design may be modified to alter the number, spacing, and information times of subsequent interim analyses, with potential restriction of enrollment to a sensitive subgroup. A similar approach was described by

Mehta et al. [45] with specific focus on the challenges associated with time-to-event endpoints used in a sequential enrichment strategy, namely the complex tradeoff between power, sample size, number of events, timing of interim analyses, and study duration. A review of adaptive enrichment methods can be found in an article by Wang and Hung [46].

A two-stage adaptive enrichment design incorporating continuous marker threshold selection and independent marker evaluation and statistical validation cohorts was proposed by Renfro et al. [47]. This design assumed time-to-event endpoints and allowed for unequal randomization (e.g., for the case of a placebo-controlled trial where 2:1 randomization may be preferable). An interim analysis was used to identify a tentative marker threshold and determine whether an initially unselected trial should stop early for futility, continue unselected without a biomarker, or adapt accrual and resize according to the degree of promise shown by the biomarker. In application to an actual trial, it was noted that treatment effect as a function of a continuous marker's threshold may be noisy or non-monotone, which may require smoothing or advanced methods for threshold evaluation. Another adaptive enrichment design with cutpoint selection was proposed by Simon and Simon [48].

Phase II/III biomarker adaptive designs—Jenkins et al. [49] introduced a randomized seamless phase II/III design using early and late time-to-event endpoints with co-primary tests for overall and subgroup-specific treatment effects in the phase III portion. After an interim analysis between the phases, which uses the shorter-term endpoint, the trial can either continue to phase III in the co-primary overall and subgroup populations, continue in the subgroup only, continue in the full population without consideration of the subgroup, or stop for futility. Friede et al. [50] improved upon the designs of Jenkins et al. [49] and Brannath et al. [43] by identifying an optimal approach for controlling the family-wise type I error rate for the population and subgroup tests with uniformly greater power.

A phase II marker-based approach designed to adaptively inform phase III design recommendations was also developed by Freidlin et al. [51], wherein the overall and subgroup-specific results of the phase II trial lead to one of four phase III design recommendations: a randomized phase III trial in an enriched population, a marker-stratified phase III trial in an unselected population, a traditional randomized trial ignoring the biomarker in design, or no phase III trial when the phase II trial is futile overall.

Enriched biomarker-adaptive designs—To address cases where initially enriched (rather than initially unselected) randomization between treatments may be desired, An et al. proposed enriched designs with possible adaptation from randomization to direct assignment [52–53]. After an interim analysis separating two stages of patient enrollment, such a trial may stop for futility or efficacy, continue on as a randomized trial, or switch toward direct assignment of patients to the experimental treatment based on initially promising, but not definitive, results. Another enriched design with multi-stage treatment assessments was developed by Gao et al. [54], wherein it is assumed the biomarker for enrichment has imperfect performance (e.g., sensitivity and specificity). These authors compared their adaptive targeted design against an untargeted multistage, group-sequential counterpart, studying the relative impacts of marker sensitivity and specificity and population heterogeneity on efficiency (sample size), type I error, and conditional power. They found

that the proposed design showed greater efficiency and conditional power than the unselected design in many cases where good biomarker performance was assumed; however biomarkers with poor performance could lead to over-estimation of conditional power at earlier stages, resulting in invalid stopping decisions.

Adaptive strategy designs for biomarkers with measurement error—Wason et al. (2014) described an adaptive design similar to the biomarker strategy design described above [12,16–17], but where a second cheaper alternative biomarker exists that may be highly concordant with a more expensive gold standard marker [55]. The trial is comprised of two stages: in the first stage, patients are randomized to treatment driven by the gold-standard biomarker versus standard of care chemotherapy, while the secondary marker value is also recorded. In the second stage, the trial may switch toward use of the cheaper secondary marker if the two markers are highly concordant for predicting strategy benefit at an interim analysis between the stages. At the trial's conclusion, the primary objective is comparison of treatment strategies with or without use of the primary or secondary biomarker. This design was implemented with some modifications in the Optimal Personalized Treatment of Early Breast Cancer using Multi-Parameter Analysis (OPTIMA) trial for ER-receptor-positive, HER2-negative breast cancer [56].

Also assuming two biomarker assessment methods where one is a gold standard and the other subject to measurement error, Zang et al. [57] proposed two testing strategies for evaluating biomarker-based treatment effects in a two-stage framework similar to that of Wason et al. [55]: one that optimizes correct assignment of individual patients to the best targeted treatment, and one that maximizes the average treatment response rate of patients enrolled to the trial.

Group sequential design for biomarker-guided comparative effectiveness research—Lai et al. published a novel group sequential design for developing and testing biomarker-based treatment strategies in the area of comparative effectiveness research, where the relative efficacy of approved rather than new treatments are of interest [58]. With emphasis on attaining the best response rates for patients in the trial through adaptive randomization, the design addressed three objectives: (1) treating accrued patients with the best (yet unknown) available treatment, (2) developing a treatment strategy for future patients, and (3) demonstrating that the strategy developed has better outcomes than the historical mean effect of standard-of-care therapy plus some threshold indicating a clinically significant improvement.

A comparison of fixed and adaptive biomarker designs: maximizing expected utility—An analytical and simulation-based comparison of popular fixed and adaptive biomarker-based designs was performed by Graf et al., with the objective of maximizing expected utility [59]. Specifically, utility functions were developed from both sponsor and public health points of view to quantify the relative gains and risks associated with rejected overall or subgroup hypotheses, accounting for possibly incorrect biomarker-based decisions such as inappropriate enrichment or maker-subgroup type I errors.

Bayesian biomarker-adaptive trial designs—Here, we present a class of designs that share some structural similarities with classical designs, but which utilize the Bayesian inference paradigm for decision making. For background on Bayesian clinical trials, we refer readers to the text by Berry et al. [60]. A feature of many of the following designs is response-adaptive randomization, i.e., the preferential randomization of patients at higher probabilities to treatments showing promise based on early outcome data. We briefly note that this practice is controversial and refer the reader to the position papers by Korn and Freidlin [61] and Berry [62]. A general schema for a response-adaptive, biomarker-based design is shown in Fig. 1E.

Multiple treatments and marker groups with adaptive randomization—Although some earlier Bayesian trial designs exist that could be generalized for use with biomarkers, e.g., hierarchical designs modeling tumor heterogeneity [63–64], one of the first Bayesian trials specifically designed to investigate differential biomarker-driven treatment effects was the BATTLE trial [65–67]. Within BATTLE, patients with advanced non-small cell lung cancer who had previously failed chemotherapy were assigned to 1 of 5 marker subgroups (comprised of 11 biomarkers) based on biomarker profiling at the time of enrollment and randomized to 1 of 4 treatments, resulting in 20 treatment-by-marker combinations. A key feature of this trial was response-adaptive adaptive randomization, where an initial learning period within each treatment arm was used to subsequently randomize patients with increasing probability to the treatment showing the most benefit (in terms of 8-week disease control rate) within his or her marker group. Additionally, if at any point during the trial current data showed that a treatment is unlikely to be beneficial to patients with a certain biomarker profile, randomization of subsequent patients with the same profile to that treatment was suspended. If by the end of the trial the posterior probability of the disease control rate exceeding an efficacy threshold within a particular marker-treatment subgroup was sufficiently high, a treatment was declared effective within that group.

Another often-cited Bayesian biomarker-based design is I-SPY 2, a phase II trial of neoadjuvant treatment of women with locally advanced breast cancer [68]. In this ongoing study, a baseline biopsy results in an assignment to 1 of many biomarker signature cohorts, wherein patients are randomized to one of several treatments. The primary endpoint is pathologic complete response (pCR), with relationships between MRI measurements over time modeled longitudinally. A drug performing well within a specific marker signature (in terms of Bayesian predictive probability) triggers adaptive randomization at higher probabilities for subsequent patients enrolled within the same signature, and definitively successful drugs are “graduated” to phase III study within the signature. Meanwhile, treatments not showing promise within a signature are removed from consideration, and drugs reaching futility across all signatures are dropped from the trial. This general framework allows for novel targeted agents of interest to continually enter and exit the trial protocol in an operationally seamless manner, taking advantage of established infrastructure and site participation. As noted by several authors [65–67,69–71], the utility of Bayesian adaptive randomization depends on quick marker assessment (so patients can be expediently randomized and treated), a relatively quick endpoint to inform the randomization algorithm,

and a slow-to-moderate accrual rate to ensure that early adaptations may benefit subsequent patients.

A Bayesian trial design incorporating covariate adjustment and response-adaptive randomization across multiple treatment arms was proposed by Eickhoff et al. [72]. In this design, predictive biomarker subgroups are determined adaptively during the trial (i.e., after an initial burn-in period of enrollment) rather than completely specified up-front, through a partial least squares logistic regression approach. This framework was outlined agnostic of endpoint type, including early stopping rules for efficacy within marker-treatment combinations, and showed greater power for marker subgroup detection than a marker-by-treatment interaction design with response-adaptive randomization in most cases (though the authors note that strong control of type I error was not a design objective).

Lee et al. [73] performed a simulation-based comparison of a simple randomize-all design, a marker-by-treatment interaction design [12], a marker strategy design [12,16–17], an enriched design [6–7], and a randomized design with Bayesian adaptive randomization similar to BATTLE [65]. In general, when a differential treatment effect existed within a marker group, the authors were able to demonstrate the efficiency advantages of the Bayesian adaptive design over the frequentist designs. Also exploring several biomarker groups and several potential treatments, Lai et al. [74] extended the general framework of Zhou et al. [65] to the special case where only a subset of the treatments are of interest within each marker group, but acknowledged that adaptive randomization can bias the treatment effect estimates and the resulting imbalances can complicate analysis. Barry et al. [75] described a version of the BATTLE design [65] where a seamless transition from equal to adaptive randomization was achieved via informative priors, with the objective of minimizing the impact of variable early data on treatment assignments.

Bayesian adaptive enrichment designs—A Bayesian version of the adaptive enrichment design originally proposed by Wang et al. [41] that allows for formal specification of prior confidence in a biomarker's predictive ability was described by Karuri and Simon [76]. Recently, Song [77] demonstrated application of the adaptive enrichment design of Wang et al. [41] and a Bayesian extension utilizing the prediction methods of Huang et al. [78] to second and first line trials in hepatocellular carcinoma, respectively, where the latter study utilized an earlier endpoint to predict longer-term responses. Krisam and Kieser [79] extended the adaptive enrichment approach of Jenkins et al. [49] to the setting of a binary endpoint, deriving optimal decision rules by considering Bayes' risk.

Bayesian early phase design for subpopulation detection—To address the scenario where a single randomized phase II design is desired that includes one potentially predictive marker on a graded scale, Morita et al. (2014) developed a Bayesian design that detects a naturally ordinal or continuous marker using posterior probabilities obtained from subgroup or regression-based models for a time-to-event endpoint [80]. In this design, monotonicity of the treatment effect as a function of the biomarker is assumed, but a dichotomizing threshold for defining a sensitive subpopulation need not be established prior to initiation of the trial. The result of such a study is a proposed threshold for focusing the population of study in phase III (which can range from the entire phase II population without

enrichment to no subpopulation, i.e. futility), and an estimate of treatment effect in the selected subpopulation. Through simulation studies, the authors suggest that the regression-based modeling approach is more powerful and achieves better operating characteristics than nested subgroup analyses.

Characteristics, requirements, and examples of selected biomarker-adaptive designs are shown alongside classical biomarker designs in Table 1.

Selected case studies: implementation of biomarker-based designs in oncology

In this section, we highlight some practical considerations and challenges faced within selected recently completed or ongoing biomarker-based trials. For a comprehensive example detailing the marker design selection process for SWOG trial S0819, we recommend the article by Redman et al. [81]. Some of the following trials were previously reviewed by Mandrekar et al. [82].

BATTLE

The final results of the BATTLE trial and its inherent challenge were detailed in manuscripts [66–67,69–71]. Although several marker groups (each comprised of multiple markers) with enhanced benefit were identified, some groups were noted to be less predictive than their individual markers, resulting in weaker results than may have been observed otherwise. It was also evident that several pre-specified markers had little if any predictive ability for optimizing treatment selections, and the trial experienced a “drift” in the proportion of patients previously treated with erlotinib. Kim et al. [66] stated that in a follow-up trial, BATTLE-2 [67], a prospectively defined learning period would occur, from which only biomarkers showing sufficiently strong predictive ability would be subsequently used to guide patient assignments. BATTLE-2 is currently ongoing.

FOCUS4 and ALCHEMIST

Another ongoing biomarker-based trial is FOCUS4, in which previously treated metastatic colorectal cancer patients are screened for mutations or genetic features and subsequently enrolled to one of four targeted, biomarker-enriched, randomized, and placebo-controlled sub-trials (or to a sub-trial for all-wild-type patients) conducted in parallel [83]. As it guarantees availability of an appropriately personalized sub-trial for each patient with successful screening, this design was intended to be attractive to patients and avoid many of the pitfalls of individual targeted designs, including lack of efficiency in the number screened versus the number randomized. Another attractive design feature of FOCUS4 is randomization of marker-negative (wild-type) patients to targeted treatments showing promise in the targeted cohorts, so marker predictiveness can be assessed. Randomization against placebo rather than single-arm sub-trials within enriched marker-positive cohorts also ensures that any promising effects observed are not simply due to chance enrollment of patients with better prognostic features. FOCUS4 opened to enrollment in 2014, and is currently planning 4–5 years of patient follow-up. A similar ongoing study is ALCHEMIST, wherein patients with advanced nonsmall cell lung cancer tumors harboring EGFR or ALK

mutations are randomized to targeted therapy versus placebo within respective cohorts, while negative/wild-type patients are followed for survival [84].

LUNG-MAP

An ongoing trial with structure and objectives similar to FOCUS4 and ALCHEMIST but with randomization against standard of care within each of several marker-positive sub-studies is LUNGMAP, a phase II/III master protocol in patients with squamous cell lung cancer [85]. While the individual cohorts are technically separate protocols and no cross-cohort analyses are planned, randomization of targeted agents against standard of care requires promising markers/agents to demonstrate a truly predictive effect within cohorts to graduate to phase III study. An earlier disease endpoint (progression-free survival) serves as primary in phase II, while the phase III endpoint (overall survival) is collected during phase II and later analyzed with phase III data. There is no formal modeling of the relationship between the two endpoints, which Berry et al. [86] suggest could improve trial efficiency and be adaptively used to plan phase III trials tailored by phase II observations within promising cohorts. Trials such as FOCUS4, ALCHEMIST, and LUNG-MAP may be referred to as “umbrella” trials as they enroll many marker-defined cohorts in parallel under the “umbrella” of one disease area (Fig. 1F).

NCI MATCH

Another newly open marker-targeted master protocol is NCI MATCH, which uses Next Generation Sequencing to assign patients with advanced solid tumors or lymphoma to individual single-arm targeted studies where it is hypothesized that an enhanced tumor response will be achieved with targeted therapy [87]. The 25 subprotocols, often collectively referred to as a “basket trial”, are marker-specific but tumor agnostic and conducted in parallel without analyses across protocols (Fig. 1G). Because a single MATCH sub-protocol will enroll 30 patients from a number of different tumor types (e.g., breast and colon) that likely have different prognostic or treatment response profiles, it is possible that a sub-trial’s primary endpoint (tumor response rate) will vary across organ classes regardless of marker status and challenge the interpretation of results. Nonetheless, this trial is among the first of its kind, and reflects the rapidly changing paradigm of disease classification from one of organs and stage of disease to one of patient- and tumor-specific biology. The MATCH trial was activated in August 2015.

Going forward: future design challenges

The past decade has seen tremendous advances, both in molecular understanding of cancer and clinical trial design methodology to address biomarker-based objectives. Ultimately, for truly personalized treatment strategies in cancer to become the standard of care, additional work in both areas is needed. Specifically within clinical trial methodology, the need remains for a flexible design paradigm that incorporates both prospective identification of naturally continuous or combination biomarkers that are predictive of treatment effect, and on-study clinical validation of such markers and their thresholds or classifications, which continue to be typically performed in separate studies in an ad-hoc fashion. Furthermore, there is a need for more rigorous methodology and improved approaches for biomarker

threshold selection (i.e., classification of patients as marker-“positive” vs. marker-“negative”) when naturally continuous or combined biomarkers or signatures are utilized. As new advances on fronts of clinical research and trial methodology are made, we must also keep in mind the individual patients’ perspective, so that enrollment to biomarker-based clinical trials is an increasingly reasonable and ethical option [88]. In this current era of personalized and targeted medicine, systematic evaluation and development of new design strategies, both for early phase and definitive trials, will continue to be required for proper clinical validation of tailored tests and treatments.

References

1. Armado RG, Wolf M, Peeters M, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol.* 2008; 26:1626–1634. [PubMed: 18316791]
2. Jonker DJ, O’Callaghan CJ, Karapetis CS, et al. Cetuximab for the treatment of colorectal cancer. *N Engl J Med.* 2007; 357:2040–2048. [PubMed: 18003960]
3. Zhou C, Wu YL, Chen G, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicenter, open-label, randomized, phase 3 study. *Lancet Oncol.* 2011; 12:735–742. [PubMed: 21783417]
4. Maemondo M, Inoue A, Kobayashi K, et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N Engl J Med.* 2010; 362:2380–2388. [PubMed: 20573926]
5. Mandrekar SJ, Sargent DJ. Drug designs fulfilling the requirements of clinical trials aiming at personalizing medicine. *Chin Clin Oncol.* 2014; 3:1–11. [PubMed: 25842079]
6. Simon R, Maitournam A. Evaluating the efficiency of targeted designs for randomized clinical trials. *Clin Cancer Res.* 2004; 10:6759–6763. [PubMed: 15501951]
7. Maitournam A, Simon R. On the efficiency of targeted clinical trials. *Stat Med.* 2005; 24:329–339. [PubMed: 15551403]
8. Wang SJ, Hung J, O’Neill RT. Genomic classifier for patient enrichment: misclassification and type I error issues in pharmacogenomics noninferiority trial. *Stat Biopharm Res.* 2011; 3:310–319.
9. Freidlin B, Korn EL. Biomarker enrichment strategies: matching trial design to biomarker credentials. *Nat Rev Clin Oncol.* 2014; 11:81–90. [PubMed: 24281059]
10. Perez EA, Romond EH, Suman VJ, et al. Four-year follow-up of trastuzumab plus adjuvant chemotherapy for operable human epidermal growth factor receptor 2-positive breast cancer: joint analysis of data from NCCTG N9831 and NSABP B-31. *J Clin Oncol.* 2011; 25:3366–3373. [PubMed: 21768458]
11. Bang YJ, Van Cutsem E, Feyereislova A, et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet.* 2010; 376:687–697. [PubMed: 20728210]
12. Sargent DJ, Allegra C. Issues in clinical trial design for tumor marker studies. *Semin Oncol.* 2002; 29:222–230. [PubMed: 12063675]
13. Wang SJ, Hung HMJ. A regulatory perspective on essential considerations in design and analysis of subgroups when correctly classified. *J Biopharm Stat.* 2014; 24:19–41. [PubMed: 24392976]
14. Kim ES, Hirsh V, Mok T, et al. Gefitinib versus docetaxel in previously treated non-small-cell lung cancer (INTEREST): a randomized phase III trial. *Lancet.* 2008; 372:1809–1818. [PubMed: 19027483]
15. Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. *J Natl Cancer Inst.* 2010; 102:152–160. [PubMed: 20075367]
16. Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol.* 2005; 23:2020–2027. [PubMed: 15774793]
17. Mandrekar SJ, Grothey A, Goetz MP, Sargent DJ. Clinical trial designs for prospective validation of biomarkers. *Am J Pharmacogenomics.* 2005; 5:317–325. [PubMed: 16196501]

18. Le Tourneau C, Paoletti X, Servant N, et al. Randomised proof-of-concept phase II trial comparing targeted therapy based on tumour molecular profiling vs. conventional therapy in patients with refractory cancer: results of the feasibility part of the SHIVA trial. *Br J Cancer*. 2014; 111:17–24. [PubMed: 24762958]
19. Le Tourneau C, Delord JP, Gonçalves A, et al. Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicenter, open-label, proof-of-concept, randomized, controlled phase 2 trial. *Lancet Oncol*. 2015 published online.
20. [accessed 2.09.15] NCI-MPACT: Molecular profiling-based assignment of cancer therapy for patients with advanced solid tumors. <<https://clinicaltrials.gov/ct2/show/NCT01827384>>
21. 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*. 2005; 11:7872–7878. [PubMed: 16278411]
22. Jiang W, Freidlin B, Simon R. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *Clin Cancer Res*. 2007; 99:1036–1043.
23. Song Y, Chi GYH. A method for testing a prespecified subgroup in clinical trials. *Stat Med*. 2007; 26:3535–3549. [PubMed: 17266164]
24. Freidlin B, Jiang W, Simon R. The cross-validated adaptive signature design. *Clin Cancer Res*. 2010; 16:691–698. [PubMed: 20068112]
25. Zhao YD, Dmitrienko A, Tamura R. Design and analysis considerations in clinical trials with a sensitive subpopulation. *ASA Biopharm Res*. 2010; 2:72–83.
26. Riddell CA, Zhao Y, Petkau J. An adaptive clinical trials procedure for a sensitive subgroup examined in the multiple sclerosis context. *Stat Methods Med Res*. 2013 epub ahead of print (not published as of 26 Aug 2015).
27. Mackey HM, Bengtsson T. Sample size and threshold estimation for clinical trials with predictive biomarkers. *Contemp Clin Trials*. 2013; 36:664–672. [PubMed: 24064355]
28. Hoering A, LeBlanc M, Crowley JJ. Randomized phase III clinical trial designs for targeted agents. *Clin Cancer Res*. 2008; 14:4358–4367. [PubMed: 18628448]
29. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: one size does not fit all. *J Biopharm Stat*. 2009; 19:530–542. [PubMed: 19384694]
30. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. *J Clin Oncol*. 2009; 27:4027–4034. [PubMed: 19597023]
31. Mandrekar SJ, Sargent DJ. Predictive biomarker validation in practice: lessons from real trials. *Clin Trials*. 2010; 7:567–573. [PubMed: 20392785]
32. Buyse M, Michiels S, Sargent DJ, Grothey A, Matheson A, de Gramont A. Integrating biomarkers in clinical trials. *Expert Rev Mol Diagn*. 2011; 11:171–182. [PubMed: 21405968]
33. Goshio M, Nagashima K, Sato Y. Study designs and statistical analyses for biomarker research. *Sensors*. 2012; 12:8966–8986. [PubMed: 23012528]
34. Maitland ML, Schilsky RL. Clinical trials in the era of personalized medicine. *Ca Cancer J Clin*. 2011; 61:365–381. [PubMed: 22034206]
35. Baker SG. Biomarkers, subgroup evaluation, and clinical trial designs. *Discov Med*. 2012; 13:187–192. [PubMed: 22463794]
36. Chen JJ, Lu TP, Chen DT, Wang SJ. Biomarker adaptive designs in clinical trials. *Transl Cancer Res*. 2014; 3:279–292.
37. Simon R. Biomarker based clinical trial design. *Chin Clin Oncol*. 2014; 3:39. [PubMed: 25841465]
38. Mandrekar SJ, Sargent DJ. Drug designs fulfilling the requirements of clinical trials aiming at personalized medicine. *Chin Clin Oncol*. 2014; 3:14. [PubMed: 25414851]
39. Stallard N, Hamborg T, Parsons N, Friede T. Adaptive designs for confirmatory clinical trials with subgroup selection. *J Biopharm Stat*. 2014; 24:168–187. [PubMed: 24392984]
40. Wang SJ, O'Neill RT, Hung HMJ. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. *Pharm Stat*. 2007; 6:227–244. [PubMed: 17688238]
41. Wang SJ, Hung HMJ, O'Neill RT. Adaptive patient enrichment designs in therapeutic trials. *Biom J*. 2009; 51:358–374. [PubMed: 19358222]

42. Hong F, Simon R. Run-in phase III trial design with pharmacodynamics predictive biomarkers. *J Nat Cancer Inst.* 2013 published online.
43. Brannath W, Zuber E, Branson M, et al. Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Stat Med.* 2009; 28:1445–1463. [PubMed: 19266565]
44. Mehta CR, Gao P. Population enrichment designs: case study of a large multinational trial. *J Biopharm Stat.* 2011; 21:831–845. [PubMed: 21516572]
45. Mehta C, Schäfer H, Daniel H, Irle S. Biomarker driven population enrichment for adaptive oncology trials with time to event endpoints. *Stat Med.* 2014; 33:4515–4531. [PubMed: 25130879]
46. Wang SJ, Hung HMJ. Adaptive enrichment with subpopulation selection at interim: methodologies, applications, and design considerations. *Contemp Clin Trials.* 2013; 36:673–681. [PubMed: 24076485]
47. Renfro LA, Coughlin CM, Grothey A, Sargent DJ. Adaptive randomized phase II design for biomarker threshold selection and independent evaluation. *Chin Clin Oncol.* 2014; 3:22–31. [PubMed: 25841418]
48. Simon N, Simon R. Adaptive enrichment designs for clinical trials. *Biostatistics.* 2013; 14:613–625. [PubMed: 23525452]
49. Jenkins M, Stone A, Jennison C. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. *Pharmaceut Statist.* 2010; 10:347–356.
50. Friede T, Parsons N, Stallard N. A conditional error function approach for subgroup selection in adaptive clinical trials. *Stat Med.* 2012; 31:4309–4320. [PubMed: 22865774]
51. Freidlin B, McShane LM, Polley MC, Korn EL. Randomized phase II trial designs with biomarkers. *J Clin Oncol.* 2012; 30:3304–3309. [PubMed: 22869885]
52. An MW, Mandrekar SJ, Sargent DJ. A 2-stage phase II design with direct assignment option in stage II for initial marker validation. *Clin Cancer Res.* 2012; 18:4225–4233. [PubMed: 22700865]
53. An MW, Lu X, Sargent DJ, Mandrekar SJ. The direct assignment option as a modular design component: an example for the setting of two predefined subgroups. *Comput Math Method M.* 2015 published online.
54. Gao Z, Roy A, Tan M. Multistage adaptive biomarker-directed targeted design for randomized clinical trials. *Contemp Clin Trials.* 2015; 42:119–131. [PubMed: 25778672]
55. Wason J, Marshall A, Dunn J, et al. Adaptive designs for clinical trials assessing biomarker-guided treatment strategies. *Br J Cancer.* 2014; 110:1950–1957. [PubMed: 24667651]
56. Bartlett J, Canney P, Campbell A, et al. Selecting breast cancer patients for chemotherapy: the opening of the UK OPTIMA trial. *Clin Oncol (R Coll Radiol).* 2012; 25:109–116. [PubMed: 23267818]
57. Zang Y, Liu S, Yuan Y. Optimal marker-adaptive designs for targeted therapy based on imperfectly measured biomarkers. *Appl Stat.* 2015; 64:635–650.
58. Lai TL, Liao OY, Kim DW. Group sequential designs for developing and testing biomarker-guided personalized therapies in comparative effectiveness research. *Contemp Clin Trials.* 2013; 36:651–663. [PubMed: 23994669]
59. Graf AC, Posch M, Koenig F. Adaptive designs for subpopulation analysis optimizing utility functions. *Biom J.* 2015; 1:76–89. [PubMed: 25399844]
60. Berry SM.; Carlin, BP.; Lee, JJ.; Müller, P. Bayesian adaptive methods for clinical trials. Boca Raton: Chapman & Hall/CRC Press; 2011.
61. Korn EL, Freidlin B. Outcome-adaptive randomization: is it useful? *J Clin Oncol.* 2011; 29:771–776. [PubMed: 21172882]
62. Berry D. Adaptive clinical trials: the promise and the caution. *J Clin Oncol.* 2009; 29:606–609. [PubMed: 21172875]
63. Thall PF, Wathen JK, Bekele BN, Champlin RE, Baker LH, Benjamin RS. Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Stat Med.* 2003; 22:763–780. [PubMed: 12587104]

64. Berry SM, Broglio KR, Groshen S, Berry DA. Bayesian hierarchical modeling of patient subpopulations: efficient designs of phase II oncology clinical trials. *Clin Trials*. 2013; 10:720–734. [PubMed: 23983156]
65. Zhou X, Liu S, Kim ES, et al. Bayesian adaptive design for targeted therapy development in cancer—a step toward personalized medicine. *Clin Trials*. 2008; 5:181–193. [PubMed: 18559407]
66. Kim ES, Herbst RS, Wistuba II, et al. The BATTLE trial: personalizing therapy for lung cancer. *Cancer Discov*. 2011; 1:44–53. [PubMed: 22586319]
67. Liu S, Lee JJ. An overview of the design and conduct of the BATTLE trials. *Chin Clin Oncol*. 2015 published online ahead of print.
68. Barker AD, Sigman CC, Kelloff GJ, et al. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther*. 2009; 86:97–100. [PubMed: 19440188]
69. Berry DA, Herbst RS, Rubin EH. Reports from 2010 Clinical and Translational Cancer Research Think Tank Meeting: design strategies for personalized therapy trials. *Clin Cancer Res*. 2012; 18:638–644. [PubMed: 22298897]
70. Marchenko O, Fedorov V, Lee JJ, et al. Adaptive clinical trials: overview of early-phase designs and challenges. *Ther Innov Regul Sci*. 2014; 48:20–30.
71. Lara PN Jr, Redman MW, Kelly K, et al. Disease control rate at 8 weeks predicts clinical benefit in advanced non-small cell lung cancer: results from Southwest Oncology Group randomized trials. *J Clin Oncol*. 2008; 26:463–467. [PubMed: 18202421]
72. Eickhoff JC, Kim K, Beach J, et al. A Bayesian adaptive design with biomarkers for targeted therapies. *Clin Trials*. 2010; 7:546–556. [PubMed: 20571131]
73. Lee JJ, Gu X, Liu S. Bayesian adaptive randomization designs for targeted agent development. *Clin Trials*. 2010; 7:584–596. [PubMed: 20571130]
74. Lai TL, Lavori PW, Shih MC, et al. Clinical trial designs for testing biomarker-based personalized therapies. *Clin Trials*. 2012; 9:141–154. [PubMed: 22397801]
75. Barry WT, Perou CM, Marcom PK, et al. The use of Bayesian hierarchical models for adaptive randomization in biomarker-driven phase II studies. *J Biopharm Stat*. 2015; 25:66–88. [PubMed: 24836519]
76. Karuri SW, Simon R. A two-stage Bayesian design for co-development of new drugs and companion diagnostics. *Stat Med*. 2012; 31:901–914. [PubMed: 22238151]
77. Song JX. A two-stage patient enrichment adaptive design in phase II oncology trials. *Contemp Clin Trials*. 2014; 37:148–154. [PubMed: 24342820]
78. Huang XL, Ning J, Li Y, et al. Using short-term response information to facilitate adaptive randomization for survival clinical trials. *Stat Med*. 2009; 28:1680–1689. [PubMed: 19326367]
79. Krisam J, Kieser M. Optimal decision rules for biomarker-based subgroup selection for a targeted therapy in oncology. *Int J Mol Sci*. 2015; 16:10354–10375. [PubMed: 25961947]
80. Morita S, Yamamoto H, Sugitani Y. Biomarker-based Bayesian randomized phase II clinical trial design to identify a sensitive patient subpopulation. *Stat Med*. 2014; 33:4008–4018. [PubMed: 24820639]
81. Redman MW, Crowley JJ, Herbst RS, et al. Design of a phase III clinical trial with prospective biomarker validation: SWOG S0819. *Clin Cancer Res*. 2012; 18:4004–4012. [PubMed: 22592956]
82. Mandrekar SJ, Dahlberg SE, Simon R. Improving clinical trial efficiency: thinking outside the box. *ASCO Educational Book*. 2015 published online.
83. Kaplan R, Maughan T, Crook A, et al. Evaluating many treatments and biomarkers in oncology: a new design. *J Clin Oncol*. 2013; 36:4562–4570. [PubMed: 24248692]
84. Gerber DE, Oxnard GR, Govindan R. ALCHEMIST: bringing genomic discovery and targeted therapies to early-stage lung cancer. *Clin Pharmacol Ther*. 2015; 97:447–450. [PubMed: 25677079]
85. LUNG-MAP. A groundbreaking collaborative approach to clinical trials. <<http://www.lung-map.org>>.

86. Berry DA. The brave new world of clinical cancer research: adaptive biomarker-driven trials integrating clinical practice with clinical research. *Mol Oncol.* 2015; 9:951–959. [PubMed: 25888066]
87. [accessed 3.09.15] NCI Molecular Analysis for Therapy Choice Program (MATCH) & Pediatric Match. <<http://www.cancer.gov/clinicaltrials/noteworthy-trials/match>>
88. Meric-Bernstam F, Brusco L, Shaw K, et al. Feasibility of large-scale genomic testing to facilitate enrollment onto genomically matched clinical trials. *J Clin Oncol.* 2015 published online ahead of print.

Fig. 1. Schemas for biomarker-driven designs: (A) enrichment design; (B) biomarker-by-treatment interaction design; (C) multi-target, multi-agent biomarker strategy design; (D) adaptive enrichment design; (E) biomarker design with response-adaptive randomization; (F) umbrella trial; and (G) basket trial.

Table 1

Requirements, features, and examples of most common categories of biomarker-driven trials.

Design	Requires strong predictive marker evidence	Requires excellent assay performance	Requires fast assay turn-around time	Requires moderate to high marker prevalence	Enrolls and treats all eligible patients	Includes many markers and drugs	Adapts sample size during trial	Ability to validate individual predictive marker	Example trials [reference]
Marker enriched	Yes	Yes	Yes	Maybe	No	Maybe	No	No	N9831 [10] TOGA [11]
Marker-stratified or interaction	No	Yes	Yes	Maybe	Yes	Maybe	No	Yes	INTEREST [14] MARVEL [15]
Marker strategy	Yes	Yes	Yes	No	Yes	Maybe	No	No	SHIVA [18–19] M-PACT [20]
Adaptive enrichment	Yes	Yes	Yes	Yes	Maybe	No	Yes	No	–
Outcome-adaptive	Yes	Yes	Yes	Maybe	Maybe	Maybe	Yes	Maybe	BATTLE [65–67] I-SPY2 [68]
Umbrella	No	Yes	Yes	No	Maybe	Yes	Maybe	Yes	FOCUS4 [83] ALCHEMIST [84] LUNG-MAP [85]
Basket	Yes	Yes	Yes	No	Maybe	Yes	Maybe	No	NCI MATCH [87]