Cancer Treatment Reviews 43 (2016) 74–82

Contents lists available at ScienceDirect

Cancer Treatment Reviews

journal homepage: www.elsevierhealth.com/journals/ctrv

Hot Topic

# Clinical trial designs incorporating predictive biomarkers ☆

Lindsay A. Renfro [a,*], Himel Mallick [b,c], Ming-Wen An [d], Daniel J. Sargent [a], Sumithra J. Mandrekar [a]

[a] Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA
[b] Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA
[c] The Broad Institute of MIT and Harvard, Cambridge, MA, USA
[d] Department of Mathematics and Statistics, Vassar College, Poughkeepsie, NY, USA

ARTICLE INFO

*Article history:*
Received 21 September 2015
Received in revised form 26 December 2015
Accepted 29 December 2015

*Keywords:*
Clinical trial design
Adaptive trial design
Biomarker-based design
Bayesian adaptive design
Enrichment designs
Targeted therapies

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.



## 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

☆ *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.
*E-mail address:* renfro.lindsay@mayo.edu (L.A. Renfro).

http://dx.doi.org/10.1016/j.ctrv.2015.12.008
0305-7372/

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