CLINICAL INVESTIGATION

Personalized medicine: challenges in biomarker-related clinical trial design

Pei He
Amgen, Inc. 1120 Veterans Boulevard, South San Francisco, CA 94080, USA
he.pei.hepei@gmail.com

Personalized medicine becomes an area of great interest following the recent development in human genetics, proteomics and metabolomics. There is an increasing need to embed the scientific discoveries from basic medical research into real life clinical practices, such that a patient's biological traits can be used to facilitate treatment. Those individualized biological traits, termed as biomarkers, are actively involved in developing personalized medicine and therefore bring out challenges to clinical trial designs. The necessity of biomarker validation and patient subgroup selection make the trial design more complex. This paper will first introduce different types of biomarkers and then review the challenges of clinical trial design with biomarkers from the clinical, statistical and regulatory perspectives.

Keywords: biomarkers • clinical trial design • personalized medicine • predictive • prognostic • targeted cancer therapy

Personalized medicine is an increasingly promising field of interest that connects biological research with clinical practices. Recent developments in human genetics and sequencing techniques make it possible to identify disease-related genes that affect the onset risk of a certain disease and influence the type and course of treatment. Information about each individual's unique genetic or biological traits can therefore be utilized for personalized disease prevention, diagnosis, monitoring and treatment. The ultimate goal in personalized medicine is to develop patient-specific medical procedures based on 'biomarkers' (clinically meaningful biological traits) and optimize medical efficiency for all.

Clinical trials are used as essential tools in medical research to validate new medical practices, and they therefore play an important role in the advancement of personalized medicine. Clinical trials must be conducted on newly developed preventative procedures, diagnostic tools, treatments and monitoring processes before such as these can be adapted into standard clinical practice. The emphasis on biomarkers in personalized medicine has recently added more complexity to clinical trial design and data analysis.

All biomarkers of different types require clinical trials to confirm their properties and to inform and influence daily clinical practices. Regulatory agencies, such as FDA in the U.S. and the EMA in Europe, have set requirements for approving the use of biomarkers in medicine [1,2]. Statistical analyses are used to confirm the exact function of biomarkers and their interactions with newly developed treatments. In some cases, retrospective analyses using previously conducted clinical trials are sufficient to verify the biomarker function. In other cases, prospective studies are needed in order to scrutinize the clinical effectiveness, safety and benefit/risk of the biomarkers. All emerging treatments and associated biomarkers require clinical trials and regulatory reporting before achieving approval for professional and/or commercial release.

We will begin this review with an overview of definitions, classifications, and functions of biomarkers and their roles in clinical applications. We will then review the issues and

FUTURE SCIENCE part of fsg

challenges of biomarker-related clinical trial design. Later sections of this review are structured as follows: the first section primarily describes the different classifications of biomarkers in the existing literature and proposes a function-based classification for further discussion; the following section discusses the challenges of biomarker-related clinical trial design when it comes to clinical concerns, statistical issues and regulatory difficulties. We then present an elaborate review of complications, when designing biomarker-related clinical trials, for each type of biomarker based on the proposed classification. The last section of this review presents concluding remarks.

Existing definitions & categories of biomarkers in literature

Although recent developments in genetics and molecular biology have highlighted the concept of 'biomarker', there have been various definitions of 'biomarker' in the research literature serving multiple purposes. For example, Gallo *et al.* have noted the most commonly adopted definition: 'any substance or biological structure that can be measured in the human body and may influence, explain or predict the incidence or outcome of disease' [3]. Some scholars argued that being 'measured in the human body' has been too limited that the more general alternative is preferred, which is also noted by Gallo *et al.*: 'any substance, structure or process that can be measured in bio-specimen and which may be associated with health-related outcomes' [3]. However, we found this definition too broad for our purpose. We prefer the NIH version, which defines the biomarker as 'a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention' [4]. This definition provides a foundation that accommodates a broad range of current biomarker applications while also maintaining specificity by limiting the focus on biologic, pathogenic and pharmacologic processes.

Ziegler *et al.* have defined various biomarkers based on their roles, such as cancer biomarker, copy number variant biomarker, DNA biomarker, epigenetic biomarker, safety biomarker, etc [5]. They also divide the biomarkers into three intrinsically different types: the DNA biomarker, the DNA tumor biomarker and the general biomarker [5]. The DNA biomarkers are patient-specific genetic information that remains stable through a patient's lifetime across different cell structures. Some examples are single nucleotide polymorphisms; simple sequence repeats; or insertion, deletions and other variations on the DNA level. DNA tumor biomarkers refer to the DNA changes specific to cancer cells that lead to the cellular or functional changes. These biomarkers are only found in cancer cells. The third type, the 'general' biomarkers, comprises 'all other forms of biomarkers, such as RNA, protein, or metabolite measurements which can be measured in biofluid, tissue, or even cell lines' [5]. The major difference among DNA biomarkers, DNA tumor biomarkers and general biomarkers is that the former are stable over time and location while the latter two are often only observed after specific mutations and in certain types of cells.

The preventative, diagnostic, prognostic, predictive biomarkers: definition & examples

Although Ziegler *et al.* provided a clear and intrinsic classification of the current biomarkers based on their various characteristics [5], we prefer to use a function-based categorization of biomarkers to identify their various roles in personalized medicine. We classify biomarkers into four categories: preventative, diagnostic, prognostic and predictive. Table 1 lists the general definition and related properties of each type of biomarkers.

Preventative

The definition of preventative biomarker is easily revealed by its name: a biological trait that can help in patient-specific disease prevention. This type of marker is most prevalent in cancer researches. For example, patients with BRCA1 or BRCA2 mutations have the lifetime risk of 56 and 87% to develop breast cancer [6]. As a general preventative practice, NIH recommends women carrying these biomarkers to consider a bilateral prophylactic mastectomy to reduce their risk of breast cancer [7], because studies show that mastectomy reduces the risk by at least 95% among the biomarker carriers [8–10]. Another example of preventative biomarker is the mutation in the tumor suppressor adenomatous polyposis coli gene, which poses a high risk for developing colorectal cancer [11]. For such biomarker carriers, one preventative recommendation is to have a prophylactic colectomy [12].

Both preventative biomarkers noted above have either a somatic mutation or germline mutation form. The somatic mutations are often acquired but not hereditary, and they only present in certain cells. The germline mutations, however, are often hereditary and can be passed on to the next generation. According to the definition of Ziegler *et al.*, a preventative marker might be a DNA biomarker or a DNA tumor biomarker, depending on whether they present only in cancer cells (somatic mutations) or in all cell types (germline mutations) [5].

Diagnostic

Diagnostic biomarkers refer to those clinical indications that help doctors screen, diagnose or predetermine the

Table 1. Categorization of different biomarkers.

Term	Definitions	Properties	Examples
Preventative biomarkers	Biological traits that can help in patient-specific disease prevention	Often DNA biomarkers, stable over time and across tissue types	BRCA1 and BRCA2 in breast cancer prevention
Diagnostic biomarkers	Clinical indications that help doctors screen, diagnose, or measure the severity for a certain type of disease	Often exist after onset of the disease and disappear after the disease is cured	Citrullinated peptides/proteins (anti-CCP antibodies) in rheumatoid arthritis diagnosis
Prognostic biomarkers	A biomarker that monitors the disease progression or predicts the disease outcome	Often exist after the onset of the disease and may change over time; belong to either DNA cancer biomarkers or general biomarkers	MammaPrint to predict the metastasis in breast cancer
Predictive biomarkers	A biomarker that predicts the treatment response of a certain disease	Same as prognostic biomarkers.	Human epidermal growth factor receptor 2 (HER2) for breast cancer treatment of trastuzumab and lapatinib

severity of a certain type of disease. Already in general medical practices, biological molecules are detected and measured in blood, urine and cerebrospinal fluid samples to verify the existence of a disease. Pregnant women, for example, are given diagnostic tests for gestational diabetes, high blood pressure and so on. Such measurements from blood or urine samples can be regarded as a type of diagnostic biomarkers. Other tools such as imaging data are also considered as diagnostic biomarkers. For example, fluid and imaging biomarker have been employed to diagnose Alzheimer's disease, or even to identify the underlying pathological development of Alzheimer's disease before the patients become severely demented [13].

Prognostic versus predictive: distinct definitions & intricate interactions

Prognostic biomarkers and predictive biomarkers are the two most commonly used terms in biomarker-related trial designs. Prognostic biomarkers are often associated with disease outcome, whereas predictive biomarkers are associated with drug response. In the formal definitions of [14] and [5], prognostic biomarkers forecast the likely course of disease in a specific clinical population, irrespective of treatment, while predictive biomarkers predict the potential outcome in response to a specific treatment.

Unlike other types of biomarkers that can be either image-based or physiological indicators, most prognostic and predictive biomarkers refer to the cellular, molecular or genetic traits of a specific patient population. A clear example of an FDA-approved prognostic biomarker is MammaPrint, a test for breast cancer using a 70-gene to assess the risk of metastasis [15]. Another prognostic biomarker, Microsatellite instability, highlights the diversity of colorectal cancer and provides guidance for specialized treatment of colorectal cancer patients [16].

There are yet more examples of predictive markers that serve as treatment guidance for surgeons or predictors of a specific treatment. BluePrint, a test for breast cancer patients, can provide additional information to help predict patients' treatment response. Whether a patient could benefit from hormone therapy alone, and avoid chemotherapy, could be predicted by evaluating steroid hormone receptor proteins.

Predictive biomarkers are also prominent in targeted cancer therapies. For example, two breast cancer treatment drugs, trastuzumab and lapatinib, specifically target human epidermal growth factor receptor 2 (HER2), which promotes the growth of cancer cells. The HER2 gene, as a predictive biomarker, therefore plays an important role when selecting breast cancer treatment. HER2 positive patients often respond well to targeted therapies, but they are less responsive to hormone treatment [17].

In non-small cell lung cancer (NSCLC), both erlotinib and afatinib are tyrosine kinase inhibitors that target epidermal growth factor receptor (EGFR) and block the downstream signal transduction pathway. Only patients with specific EGFR mutations will benefit from such therapies [18]. Another example of a predictive biomarker in targeted cancer therapy is the KRAS gene. Both the American FDA and European EMA have approved the two EGFR inhibitors, cetuximab and panitumumab, for the treatment of colon cancer patients bearing the wild-type KRAS gene, although both treatments had little or no effect on patients with mutant KRAS. Even though KRAS is not the target of these two treatments, it plays an important role in predicting the treatment effects [19].

As noted in the examples above, the most prevalent prognostic/predictive biomarkers appear in recent advances in cancer treatment. These two types of biomarkers have also been widely used to help predict

patients' responses to treatments for noncancerous diseases. For example, an single nucleotide polymorphism located in the SLCO1B1 gene is associated with the increased risk of myopathy when treated by statin therapy [20].

Instead of having a clear role to be either prognostic or predictive, some biomarkers were first discovered to be prognostic but later were identified as predictive. For example, in standard glioblastoma treatments, O6-methylguanine-DNA-methyltransferase (MGMT) methylation used to serve as a prognostic biomarker for patients treated with radiotherapy combined with concomitant and adjuvant temozolomide (TMZ) [21]. However, Costa *et al.* presented doubts about the prognostic value of MGMT methylation based on a Portuguese multicenter study result [21]. More recently, Yin *et al.* have found that the MGMT methylation has a predictive but not prognostic value and that the survival benefits of TMZ-treated patients with MGMT methylation is significantly higher than for patients having TMZ-free therapies [22]. There are also biomarkers that are utilized for indicating both disease outcome as well as drug response. Buyse *et al.* have listed several examples: 'overexpression of HER2/neu in breast cancer, mutations in KIT in gastrointestinal stromal tumors and the presence of the fusion gene BCR-ABL (Philadelphia chromosome) in chronic myelogenous leukemia' [14]. All types of biomarkers become more and more essential in the era of personalized medicine, as they can give guidance to physicians and surgeons on how to treat patients specifically and efficiently.

Issues of biomarker-guided clinical trial designs

Due to the many biomarkers and their diverse functions and properties, designing clinical trials involving biomarkers is complicated. Not all trials have a similar purpose. Some trials are designed to validate the biomarker. Other trials are designed to select the patient subgroup that might benefit from a specific treatment. There are also designs meant to simultaneously fulfill the need to validate the biomarker, to select the subgroup and to confirm the efficacy of a treatment. Depending on the clinical evidence gathered about a specific biomarker, a series of questions need to be asked beforehand in order to determine the type of trial design most appropriate for the purpose. Given the relative newness of biomarker-driven trials, there are also regulatory challenges to consider such as companion assays for the biomarker and the diversity of trial designs.

In this section, we will review the challenges of biomarker-related trial designs from the clinical, statistical and regulatory perspective.

Clinical issues

Before designing a clinical trial, investigators need to clarify the purpose and scale of the trial. Clinical trials are generally partitioned into four phases in order to establish the pharmacodynamics, the pharmacokinetics, the required dose, and the clinical efficacy and safety of the new treatment. Depending on the type(s) of biomarker(s), trials can be used to retrospectively or prospectively validate a biomarker's impact and/or to confirm the treatment efficacy.

Clinical development process for preventative, diagnostic, prognostic & predictive biomarkers

Most reported preventative biomarkers are discovered in genomic research and verified through further observational studies involving different ethnic groups. Due to advances in molecular biotechnology, genetic mutations can be identified through standard sequencing techniques. After a genetic mutation is verified, the decision to carry out a prophylactic surgery becomes a complex issue requiring doctors to factor in the existence of any preventative biomarker as well as the patient's disease progression and risk.

Before introducing diagnostic and prognostic biomarkers into clinical practice, their validation is necessarily stringent. In 2009, the European Medicines Agency published guidelines defining biomarkers used for diagnosing or monitoring a disease as 'diagnostic agents' and established a development process to license such agents. The process is parallel to the four different phases of new drug development: Phase I must demonstrate the safety of human usage of the diagnostic agent; Phase II must establish an estimate of sensitivity and specificity to the diagnostic agent by enrolling healthy or diseased patients; Phase III must present credible validation of the diagnostic agent based on realistic and reproducible clinical settings in which the agent was and would be used, and then the efficacy of the diagnostic agent must be further verified; Phase IV must explore the possibility of clinical improvement of the diagnostic agent even after it has entered the market and clinical use [1].

The sequential phase process refers to the marketing approval of 'diagnostic agents' for *in vivo* use only. It is expected that regulatory agencies would have fewer requirements for *in vitro* diagnostic (IVD) tests compared with *in vivo* tests that involve direct contact with patients and might induce higher risk through implementation. Both FDA and EMA have published guidelines or legislation pertaining to IVD tests (further discussed later in this review). However, the FDA's guidelines focus on the use of an 'IVD companion diagnostic device' [23], whose definition corresponds to predictive biomarkers that identify a subgroup of

patients that would benefit from a newly discovered treatment. The filing for approval of such a therapeutic product is most often accompanied by the filing of a corresponding 'IVD device', although under certain (e.g., life-threatening) circumstances the FDA might grant approval of a treatment without the approval of an IVD companion device [23].

Unlike the typical, standard validation process for diagnostic or prognostic biomarkers involving the three-step sequence of assay development, retrospective validation and prospective accuracy analysis, the validation of a predictive biomarker often involves an extra step to confirm the 'predictiveness' of the biomarker with regard to the corresponding treatment. In some instances, such a validation might only involve a retrospective analysis of prospectively designed randomized trials; for example, the KRAS mutation as a predictive marker for cetuximab and panitumumab in treating metastatic colorectal cancer [24,25]. In most other cases, however, either a Phase II or a Phase III prospective trial is recommended to show the efficacy difference between biomarker positive and negative subpopulations identified by the assay developed. In this case, multifunctional prospective randomized trials to validate the biomarker and confirm the clinical efficacy simultaneously is preferred.

Other clinical issues related with predictive biomarker trials in targeted cancer therapies

The most prominent predictive biomarkers are mainly related to cancer therapies (as noted above). The following discussion, therefore, primarily focuses on targeted therapies for cancer treatment. For the sake of efficiency, the following discussion will not address how biomarkers are identified among millions of genetic traits but will assume the relationship between the molecular target and the corresponding predictive marker is well defined.

Sample collection

Cancer is a heterogeneous disease with phases and time-varying molecular properties of different cancer cells, such as primary tumor cells and metastases tumor cells. Given that most of the predictive biomarkers for targeted therapies are DNA tumor biomarkers, which exist only in tumor cells, biopsy samples from the primary tumor, the circulating tumor and the metastases tumor might all be necessary (at specific disease stages of treatment) in order to verify the existence of the marker [26,27].

End-point selection

When designing targeted therapies, end-point selection might pose another important issue to resolve. Generally speaking, the 'gold-standard' end point for cancer treatments is the overall survival. However, it might take too long to follow up and obtain the time-to-death for each patient, which could also impact the finances of a clinical trial. When overall survival is used as the end point of a trial, and the trial lasts for too long, another problem might be missing data due to patients lost to follow-up. In short-term trials, such as Phase I or Phase II, there might not be enough time allowed to gather the survival information before a Go/No-go decision for the next stage. When the clinical evidence of efficacy in terms of overall survival is unclear, early-stage trials can instead be used to explore whether a drug inhibits the targeted tumor and with some solid tumors the tumor's size can serve as a good indicator.

For various Phase II trials, tumor progression as classified by Response Evaluation Criteria In Solid Tumors (RECIST) standards becomes a common end point [28]. Other end points like tumor response are also widely used in the early stages. However, researchers have expressed some doubt about the prime valuation of objective response, because some molecularly targeted treatments have shown survival benefits despite low tumor response rates [29]. Other alternative end points have been proposed with relative advantages and disadvantages: the combination of tumor progression or response, the progression-free survival, the prognostic biomarkers indicating disease progression, imaging data, quality of life, or a continuous model of tumor measurement over time [30].

At the current stage, deriving statistical validation for an appropriate surrogate end point is still a goal of intensive statistical research and there is no consensus. Most surrogate end-point validations, through retrospective studies or meta-analysis, are treatment-specific. Therefore, whenever there is development of a new treatment with a new mechanism, whether the previously validated surrogate end point is still applicable becomes an unanswered question. There are very few universally accepted surrogate end points except the progression-free survival in fluoropyrimidine-based regimens for colon cancer treatment and the hematologic complete remission in patients with leukemia [14]. Further disease-specific research might be necessary to study the validity of potential surrogate end points. Clinical investigators need to carefully review the literature to determine the end point of a clinical trial depending on the purpose of the trial, the disease characteristics and other budget/scale constraint. Careful statistical calculations might also be carried out when a special type of end point is chosen.

Trials with multiple targets

Most targeted cancer therapies involve a complex signal transduction network of multiple receptor proteins

interacting with one another to drive the downstream pathway leading to malignant phenotypes. Sometimes, targeting a single protein or pathway might be insufficient to inhibit cancer cell activity. It is possible that the genetic instability of different receptors may be responsible for malignant phenotypes, at various time points, such that an effective treatment at early stage may later exhibit drug resistance due to the activation of alternative receptors/pathways.

In targeted cancer treatment, there are examples of clinically effective single-agent inhibitors. For example, monoclonal antibodies have been successfully found in breast cancer patients with HER2 mutations [31], NSCLC harboring EGFR receptors [18] or anaplastic lymphoma kinase [32], etc. However, for the majority of cancer types, there has been no established single treatment proven to be effective. Researchers are therefore urged to consider targeting concomitantly or sequentially on multiple agents. But mechanisms leading to tumor activities are so complex and interactive that designing a prospective randomized trial, with either all-possible combinations or sequential application of targeted treatments, might be infeasible to carry out among the well-selected patient subpopulations. The prolonged trial time of sequential regimens could also create major cost constraints, and trials designed with multiple goals and treatment combinations also often require a large sample size to achieve statistically meaningful power. More scientific research is needed to study the molecular changes when disease progresses, so that it can provide more insight to design trials with multiple agents.

Statistical issues

Typical clinical trials are mostly of fixed sample size design, where enrollment and the length of trials are planned ahead to achieve certain statistical power and to test whether the newly developed treatment is superior to the pre-existing standard of care. In cancer clinical trials and other types of trials where end points are time-to-events (death, progression, recurrence, etc.), the design involves more uncertainty so that assumptions have to be made on accrual rate, censoring pattern and other parameters related with the sample-size calculation. Fixed sample-size designs are easy to implement logistically, compared with group sequential designs or adaptive designs, where the trial might stop or vary before the specified end date. However, the latter types of flexible trials might involve more statistical complexities. Investigators should determine the trial type based on the specific clinical problem, the related biomarker and the realistic constraints.

For preventative/diagnostic/prognostic biomarkers

The initial verification of a preventative/diagnostic/prognostic biomarker begins with retrospective analysis to establish a correlation with the biomarker and the corresponding outcome. For example, in preventative biomarker analysis, researchers hope to find a correlation between the presence/absence of a certain genetic marker and the incidence rate of a certain disease. In validating diagnostic biomarkers, the sensitivity and specificity are two gold standards to measure the diagnostic accuracy.

The validation of a prognostic biomarker is similar to that of a preventative biomarker. A straightforward statistical standard for analyzing a prognostic biomarker is to demonstrate the association between the presence/absence of the biomarker and certain treatment-independent outcomes, such as the odds ratio of disease progression or death in cancer trials.

Strong association does not, however, necessarily mean strong prediction accuracy. For example, the FDA-approved prognostic biomarker known as MammaPrint, used to predict the outcome in breast cancer, was identified through retrospective studies to have an odds ratio of 15.0 for metastases within 5 years between the patients with poor or good prognosis. Although the association is quite strong, the specificity is on the modest size of 59% [33]. When more retrospective data were gathered from different research centers, a disparity was noted in the duration of follow-up [34]. Although the prognostic effect of MammaPrint was confirmed through different statistical analyses, researchers still suggested the need for prospective studies to provide further confirmation. There is still, generally, no consensus regarding statistical rules or procedures for validating preventative/diagnostic/prognostic biomarkers except to check the sensitivity, specificity or correlation between the biomarker and a clinical outcome.

For predictive biomarkers

When it comes to validating a predictive biomarker, however, the statistical literature clearly recognizes the need for at least one prospectively designed clinical trial with or without the efficacy confirmation. Depending on the stage of clinical research for the specific predictive biomarker, various forms of Phase II or Phase III designs with multiple goals have been proposed by researchers. This review will next discuss the available clinical trial designs regarding predictive biomarkers, their corresponding pros and cons, and then summarize the major statistical challenges with such trial designs A list of designs with their pros and cons discussed in this manuscript is presented in Table 2.

Enrichment design

The significance of enrichment design is highlighted in trial designs involving biomarkers. This is evident in the recent review of trial designs for biomarker validation in a Phase II setting by Mandrekar *et al.* [35] and by Tajik *et al.* [36] regarding cancer clinical trial design.

In enrichment design, all patients are tested for the biomarker before enrollment. Only patients with the positive biomarker are recruited and randomly assigned to the trial. The main purpose of enrichment design is to evaluate the safety and clinical efficacy of the treatment within the biomarker-positive group. This type of trial design is most often implemented when sufficient clinical evidence has been established to indicate that only patients with the positive biomarker would benefit from the treatment. This is more common with targeted cancer therapies, where the treatment is targeting a specific molecular pathway; an example would be the trial for trastuzumab for HER2-positive breast cancer patients [37]. In instances where the biomarker prevalence is low, an enrichment design is preferable to an all-comer design or adaptive design, because the treatment efficacy of the biomarker-positive group could be diluted by a notably larger biomarker-negative group. However, if trials only enroll biomarker-positive patients, the major drawback of an enrichment design is that clinical questions would be left unanswered for the general population. In addition, the evaluation of the assay accuracy and reproducibility of the companion test would remain uncertain.

All-comer design

Compared to the enrichment design, the all-comer design enrolls patients without the restriction on patients' biomarker status of being positive or negative. However, such designs only enroll patients with a valid biomarker status, because one of the goals of such a design is to explore the interaction of the biomarker with the treatment effect. In a Phase II setting, a preliminary interaction test can first be performed to explore the treatment-by-marker interaction. A prespecified threshold value might be defined before determining the patient population for the next stage. If the interaction test is significant, only the patients with positive biomarker values will be compared with treatment and control arms, and, if not significant, the treatment effect of the overall population will be evaluated. Other types of arrangements are feasible, such as simultaneously starting with the two separate evaluations of treatment effects in biomarker-positive and -negative subgroups followed by a comparison between the two within-group treatment effects. In such cases, the issue of multiple testing needs to be addressed either through bonferroni correction or some other procedure.

An all-comer design for a Phase III trial might differ from a Phase II all-comer design due to a different

Table 2. Designs of predictive biomarker related trials: advantages and disadvantages.

Designs	Properties	Advantages	Disadvantages
Enrichment design	Only enroll biomarker positive patients	Good for biomarker with clear evidence and/or low prevalence	Cannot gather treatment information for all population; cannot test for the companion diagnostic tool validity
All-comers design	Enroll both biomarker-positive and -negative patients	Complete treatment information for overall population	Large sample size with high cost; biomarker effect might be diluted
Biomarker strategy design	Patients are randomized to biomarker strategy group and control group	Complete information gathered; able to test companion diagnostic tools	Scale and cost might be large
Adaptive enrichment design	Two-stage design with subpopulation selection at the second stage	Flexible design with smaller expected sample size than all-comers but still contain information for biomarker-negative group	Logistically complicated; need complex simulation studies to determine sample size; multiple prerequisite needed to proceed
Group sequential design	Designs with several interims to make go/no-go decisions	Smaller expected sample size; decisions can be made during interims to save cost; proven effective treatment can be accessible to all patients early	Logistically challenging; extra complexity involved in the biomarker-related trials
Seamless design	A design in combination of two phases	Smaller expected sample size; shortened duration of the drug development process	Logistically challenging; complexity in sample size calculation

scale and purpose. Most often, a Phase III trial will have a much larger sample size and will involve regulatory filing. Therefore, a formal testing procedure will be carried out either in the overall population or the biomarker-positive population. In general, regulatory agencies provide less room for flexibility in Phase III trials. For the investigators, the major trade-off of a biomarker-guided, Phase III, all-comer trial is the possibility of targeting the right population for regulatory approval and, at the same time, maximizing the number of patients who will benefit from the new treatment. Sometimes, trials with coprimary end points are designed and treatment effects are evaluated for both the overall study population and the biomarker-positive population [38]. In other cases, an interim analysis might first be carried out to decide which population to choose and the final analysis focuses on this chosen population. Jiang *et al.* proposed a Phase III threshold design that combines the coprimary end points and subset selection together with the extra purpose of selecting the cut-point value for the biomarkers instead of assuming a binary biomarker [39]. Biomarker-guided trial design is more statistically complicated when cut-point selection is involved. Threshold design often falls into the all-comer design category, because all patients need to be included first to determine the right subpopulation.

Biomarker-strategy design

The biomarker-strategy design involves a new management strategy such that a combination of marker testing and treatment allocation is performed in different management types. This design type is similar to the all-comer design, as most often patients are enrolled before the test of biomarkers. Patients are randomized into a biomarker-strategy group and a control group. In the biomarker-strategy group, the biomarker-positive patients receive the experimental treatment while the biomarker-negative patients receive the standard-of-care.

After the first patient enrollment, whether the test of the biomarker will be carried out is up to the different design types [36]. For the patients who were randomized to the control strategy, they will either be assigned to standard-of-care or randomized to experimental or standard-of-care treatment just like the biomarker-positive group. This type of design becomes particularly useful when there is clear evidence that only biomarker-positive patients will benefit from a certain new treatment, and therefore it is unethical to randomize the biomarker-negative patients to the new treatment group.

Unlike an enrichment trial design, a biomarker-strategy design provides researchers with the opportunity to compare the validity of the companion testing assay. Also, potential future prognostic markers can be tested with data collected from such a design [35]. Both classic all-comer designs and biomarker strategy designs are preferred by researchers, because they can gather clinical information from whole populations, and this is the most comprehensive approach when pursuing multiple purposes. However, the major drawback lies in the scale and cost of the trial. If a new treatment were only effective in the biomarker-positive group, an all-comer or biomarker-strategy design would need to recruit more patients than an enrichment design in order to achieve the same statistical power. Also, the significant efficacy in the biomarker-positive group might be diluted in an all-comer or biomarker-strategy design.

Adaptive enrichment design

The three types of design (noted above) have their pros and cons, and researchers need to weigh the trade-off between the cost/sample size of a trial and the comprehensive information a trial can gather. Given its flexibility, an adaptive enrichment design provides researchers with a compromise. In such a design, trials are normally separated into two stages. In the first stage, the whole population is enrolled. Then, at interim, a decision is made to choose the all-comer or enrichment design for the second stage. How does this adaptive enrichment design differ from the type of 'all-comer' design with interim for subpopulation selection? The difference lies in the 'adaptiveness', which means there is interim adjustment for either the sample size, randomization ratio or the enrichment hypothesis, or even the change of end points in adaptive enrichment design (but not the all-comer design) with subpopulation selection. Sometimes futility stopping will also be combined in order to increase the efficiency of the design. Song proposed a two-stage design of subpopulation selection in a Phase II setting [40]. Mackey and Bengtsson proposed a sample size and cut-off threshold estimation scheme for the adaptive enrichment setting where subgroup-selection and clinical efficacy testing were carried out sequentially [41]. Wang and Hung provided an overview of the adaptive enrichment design and discussed the statistical and regulatory aspects of the design challenges using examples under different settings of composite null hypothesis testing [42]. Bayesian methods were also proposed by Zhou *et al.* to adjust the randomization scheme based on the interim data [43]. The most prominent advantage of an adaptive enrichment design is, clearly, its flexibility. Investigators do not need to make groundless decisions at the design stage but can wait until usable information is gathered. In addition, to increase the efficiency of the

clinical trial, adaptations can be applied after taking into account a variety of perspectives.

There are several prerequisites for an adaptive enrichment design. First of all, a reliable end point is required and should be easy to measure for the interim analysis. Otherwise, it would be hard to make decisions at the interim stage. When it comes to targeted cancer therapies, this brings out challenges due to the fact that the overall survival is the most well-acknowledged end point but requires long follow-up time and may not be available during interim. In such situations, a surrogate end point might be helpful when a quick estimate can be made at the interim stage and when the prediction of the final end point, based on the surrogate end point, is of sufficient accuracy. It should be noted, however, that establishing a sufficiently accurate surrogate end point can sometimes be quite difficult (see above).

In order for an adaptive enrichment design to proceed smoothly, the second requirement is the availability of clinical and biological data throughout the study duration. This logistical issue might not be easy to manage for multinational or multicenter trials, but it might be achievable in the future when all medical records are electronic. There are other challenges of adaptive enrichment design. For example, there could be imbalances of treatment and control arms that lead to a biased estimation of a specific treatment effect if not adjusted appropriately. The extra flexibility in potential changes to the patient population, the randomization scheme or the null hypothesis tested could also lead to further complications in statistical inference and power calculations. The addition of futility stopping would also affect the statistical power. Careful considerations thus need to be taken, with thorough simulation studies at the design stage, in order to formulate a reasonable and feasible adaptive enrichment trial.

Other designs (group sequential, Phases I–II, II–III)

Aside from the four types of trial designs noted above, other types of designs for predictive biomarkers also exist in the literature. Compared with adaptive designs, a group sequential design provides another option for general clinical trial design that allows for early adjustment. Unlike the adaptive designs mentioned above, the interim analyses in group sequential design are used to make Go/No-Go decisions for efficacy or futility but not for parameter adjustments. Under the setting of biomarker-guided trials, group sequential designs can also be combined with the patient subpopulation testing. Lai *et al.* proposed a group sequential design to test multiple composite hypotheses in the overall population and the biomarker-positive populations [44].

In recent studies of clinical trial designs, there is a growing interest in developing a seamless design that combines two phases of trial development to achieve efficiency. These designs harbor the qualities of both group sequential designs and adaptive designs, such that the interim results from the earlier phase is combined with the final testing in the later phase. Such design type could certainly be applied to biomarker-guided trials where different schemes of subpopulation selection can be carried out. Jenkins *et al.* suggested a seamless Phase II/III design with subgroup selection based on biomarkers [45].

The major advantage of the group sequential trial, or seamless Phase II/III trial, is the shortened expected study duration at which an early stop of efficacy or futility might be possible. Compared with the adaptive design, group sequential and seamless Phase II/III make no adjustment based on sample size or randomization schemes, and therefore these designs reduce the possibility of treatment imbalance or biased statistical inference. However, the extra layers of interims present challenges in trial logistics, and the involvement of biomarkers would further complicate the implementation and decision-making process. Clinical investigators should take into account all the constraints of such flexible trial designs and cater to the specific purpose of different treatments for various diseases.

Other statistical issues

Plenty of statistical challenges have been mentioned in the previous section when each type of design was introduced, such as the possible treatment arm imbalance, the complexity in bias correction and statistical inference, and the possible involvement of the surrogate end points, etc. One key issue in most biomarker-guided trials is multiple testing. Except in enrichment designs, where only the efficacy in biomarker-positive patients is tested, all other designs involve calculating the test statistics for more than one trial time. This necessity leads to the multiple testing problem, which is created because incorrectly rejecting the null hypotheses is more likely to occur when more than one null hypothesis is tested [46]. Bonferroni corrections are often used in confirmatory trials to split the significance level alpha into different parts for subgroups [47], and examples of this can be seen in an all-comer biomarker-guided coprimary end-point trial for NSCLC [48]. However, such a correction might be too conservative, because it does not take into account the dependence structure between the two test hypotheses. Spiessens and Debois proposed an adjusted significance level for subgroup analysis by applying a bivari-

ate normal model between the two test hypotheses and adjusting for the significance level based on the covariance structure, which is similar in fashion to that of group sequential trials [49]. Such an approach can provide a less conservative solution for the multiple testing issues in biomarker-guided designs.

Another possible statistical issue derives from the complexity of biomarkers. As mentioned above, certain biomarkers are both prognostic and predictive. When verifying the interaction between a biomarker and treatment with survival end points, a cox proportional regression is often performed with the interaction term of biomarker and treatment indicator [50]. However, the cox proportional regression model implicitly assumes independence between the biomarker value and the survival rate of the control arm that it omits the occasion where a biomarker might be both prognostic and predictive. Such difficulties could be solved by approximating the constant baseline hazard functions in cox proportional model using piecewise models and estimating the baseline hazard accordingly for further testing [51].

Regulatory issues

The regulatory challenges for predictive biomarker-guided trials first arise when the companion assays of the biomarker need to be validated together with the treatment effect of the targeted therapy [2,23]. In the US, the FDA has separate divisions to review the drug (Center for Drug Evaluation and Research) and the companion IVD assay (Center for Devices and Radiological Health). Furthermore, the two FDA divisions might have different standards for the review process. It is therefore challenging to design trials that can provide enough evidence for the efficacy and safety of both the drug and the companion testing kit that satisfy the requirements for these two divisions.

Another challenging issue exists when the development of a formal companion IVD test might be later than the pivotal clinical trial for the drug. Sometimes a more convenient laboratory-developed test is used to identify the subpopulation before the development of an IVD test. In such a case, bridging studies might be necessary to establish the 'equivalence' between both tests for regulatory reporting. The simplest route would be to retest all the tissue samples using the newly developed IVD. However, it is possible that tissue samples would not be available for all patients due to missing samples or the lack of consent forms for retesting. Therefore, the retesting population might not be a random sample from the original population of the pivotal trial. More in-depth statistical analysis, based on the missing data, should therefore be employed to adjust for such biases.

Problems can also arise during the regulatory decision process for the approval of a newly developed targeted therapy. Given the rapid development of new drugs, regulatory agencies must minimize the conflict between making a newly effective treatment available to patients as early as possible and, a the same time, detecting and deleting inefficient molecules. Both the American FDA and European EMA suggest, at the least, a positive result from a randomized Phase III trial [52]. In cancer therapies, as we mentioned in previous sections, overall survival is the most widely accepted end point for such a confirmatory Phase III trial. However, the extra follow-up time for overall survival might lead to a confounded result of treatment effect when post-study therapies come into play. Therefore, besides overall survival, both FDA and EMA would also consider other surrogate end points such as time to progression [53] and progression free survival [30], depending on the specific tumor types and whether a clinical and statistical valid surrogate end point is well established or not. Identifying the valid surrogate end points for each submitted trial becomes an extra task for the regulatory agencies.

The recent success of cancer therapy in single-arm early phase clinical trials with antitumor activities as end points brings out extra complexity to the decision-making process, as seen in the crizotinib treating NSCLC patients with the EML-ALK gene and vermurafenib treating metastatic melanoma patients with BRAF mutation [32,54]. Whether a regulatory agency should grant approval to promising treatments or to let them go through the traditional process of Phase I, II and III is a tough issue. These days, the FDA provides an accelerated approval process for new therapies with promising efficacy results treating life-threatening diseases, and surrogate end points are allowed in such a process [55]. This type of approval process has to be followed up by confirmatory trials to monitor the post-approval activities. However, it might be difficult to enroll patients after the treatment has been put on market, because the new treatment has already become available to all patients.

When there is no equivalent standard-of-care available, another challenge for investigators is determining whether they should carry out a randomized or single-arm trial. This problem becomes more prominent among targeted therapies in patients with rare tumor types or mutations, because the recruitment for a randomized trial might not be feasible, as with the use of sunitinib for metastatic alveolar soft part sarcomas [56]. The cost of the targeted cancer therapy trials, especially with the extra cost of expensive molecular tests for biomarkers, makes it more difficult for new drug development process. Researchers are now pro-

posing a flexible approval process, featuring a broader tolerance for innovative clinical trial designs, in order to give more incentives to pharmaceutical companies to develop new treatment and to allow more efficient therapies to become available to patients.

Conclusion

The fast-paced research in human genomics, proteomics and metabolomics makes it feasible to develop personalized treatment for individual patients based on their biological traits. Such distinguishing biomarkers can nowadays be applied in real clinical practices to aid in disease prevention, diagnosis, monitoring and treatment selections. However, the use of biomarkers brings out challenges in clinical trial designs due to their complex roles in disease development and the need to test and validate such roles. Clinical investigators must face new validation processes for preventative/diagnostic/prognostic biomarkers that are set by regulatory agencies. With predictive biomarkers, the sample tissue collection and testing, the end-point selection and the combination of different targets increase the difficulties of trial design. Statistical issues arise when there are more variations of trial design, each with pros and cons, proposed in the field of biomarker-guided trials. Finding the right design for specific needs in order to validate the efficacy of both the biomarker and the corresponding treatment, while working within statistical constraints, becomes a significant challenge. Issues such as multiple testing and validating dual biomarkers that are both prognostic and predictive add more complexity to the trial design. Finally, regulatory agencies place added pressure upon the need to make promising treatments for life-threatening diseases available to patients as soon as possible, while at the same time maintaining good surveillance on inefficient treatments. The complexity of biomarker-related clinical trials increases the challenges for meeting regulatory approval. Research communities are working together to design efficient and flexible trials, which provide

Executive summary

- The NIH definition of biomarker is 'a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention'.
- We classify all biomarkers according to their roles. In this paper, we mainly talk about four categories of biomarkers: preventative, diagnostic, prognostic and predictive biomarkers.
- Preventative biomarkers are biological traits that can help in patient-specific disease prevention.
- Diagnostic biomarkers refer to those clinical indications that help doctors screen, diagnose or measure the severity for a certain type of disease.
- Prognostic biomarkers forecast the likely course of disease in a specific clinical population irrespective of treatment.
- Predictive biomarkers predict the potential outcome in response to a specific treatment.
- Some biomarkers are both prognostic and predictive, which means they are not only related with the disease outcome but also the drug response.
- Retrospective analyses based on previously randomized trials are necessary to first establish a correlation between the preventative/diagnostic/prognostic biomarker and a clinical end point, and sensitivity and specificity are measured to provide validation.
- For predictive biomarkers, a formal prospective study is preferred to validate the predictiveness of the biomarker while at the same time showing the treatment efficacy.
- Other clinical challenges in designing biomarker-related trials are sample collections from various stages of cancer tissues, the end-point selection problem and how to design trials with multiple targeted agents.
- Enrichment design refers to the designs that only enroll patients with positive biomarkers and is applicable when enough clinical evidence supports the conclusion that only biomarker-positive patients would benefit from the new treatment.
- All-comers design and biomarker-strategy design enroll patients with either positive or negative biomarker status and have different schemes of randomization and marker testing procedure afterward.
- Adaptive enrichment design allows two stages in the trial, where the first stage is to gather information to make the decision of whether all patients or biomarker-positive patients should be enrolled in the second stage.
- Multiple testing is one of the most important statistical issues in biomarker-related design. Alpha spending and other types of adjustment are proposed to tackle this issue.
- Trials need to be designed to provide evidence for the efficacy and safety of both the drug and the companion testing kit that satisfy the requirement for two separate divisions in FDA to approve for the drug and the companion kit.
- Regulatory agencies are facing the dilemma of allowing accelerated approval of efficient drugs while at the same time control the quality of drug surveillance.

statistical validity to meet regulatory approval, in order to benefit the general public.

Future perspective

The involvement of biomarkers in clinical practices will be more and more common in the next 5–10 years because of the development in medical-related biological research. More clinical questions need to be answered about the biomarker and its role in disease process, and therefore more biomarker-related clinical trials will be designed to answer those specific questions. More flexible trials serving multiple purposes are expected due to the intricate relation between biomarkers and the disease. Also trials with shorter length are preferred due to the fast pace in new drug development. Pharmaceutical companies need to keep a balance of minimizing the cost and duration of a clinical trial while at the same time guarantee the clinical validity of the trial. Regulatory agencies also face with the dilemma of making effective treatment available to the right patients as soon as possible while screening out those ineffective ones. More flexible confirmatory trial designs are expected, as long as the clinical and statistical validity become well established.

Financial & competing interests disclosure

The author has no 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. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties

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

References

1 Guideline on clinical evaluation of diagnostic agents. European Medicines Agency. www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003580.pdf

2 Companion diagnostics. Food and Drug Administration (FDA). www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/ucm407297.htm

3 Gallo V, Egger M, Mccormack V *et al.* STrengthening the Reporting of OBservational studies in Epidemiology-Molecular Epidemiology (STROBE-ME): an extension of the STROBE statement. *PLoS medicine* 8(10), e1001117 (2011).

4 Atkinson AJ, Colburn WA, Degruttola VG *et al.* Biomarkers and surrogate endpoints: preferred definitions and conceptual framework*. *Clin. Pharmacol. Ther.* 69(3), 89–95 (2001).

5 Ziegler A, Koch A, Krockenberger K, Grosshennig A. Personalized medicine using DNA biomarkers: a review. *Hum. Genet.* 131(10), 1627–1638 (2012).

6 Srivastava A, Mckinnon W, Wood ME. Risk of breast and ovarian cancer in women with strong family histories. *Oncology* 15(7), 889–902, (2001).

7 Surgery to reduce the risk of breast cancer. National Cancer Institute. www.cancer.gov/cancertopics/factsheet/Therapy/risk-reducing-surgery

8 Rebbeck TR, Friebel T, Lynch HT *et al.* Bilateral prophylactic mastectomy reduces breast cancer risk in *BRCA1* and *BRCA2* mutation carriers: the PROSE Study Group. *J. Clin. Oncol.* 22(6), 1055–1062 (2004).

9 Domchek SM, Friebel TM, Singer CF *et al.* Association of risk-reducing surgery in BRCA1 or BRCA2 mutation carriers with cancer risk and mortality. *JAMA* 304(9), 967–975 (2010).

10 Hartmann LC, Schaid DJ, Woods JE *et al.* Efficacy of bilateral prophylactic mastectomy in women with a family history of breast cancer. *N. Engl. J. Med.* 340(2), 77–84 (1999).

11 Familial adenomatous polyposis. US National Library of Medicine. http://ghr.nlm.nih.gov/condition/familial-adenomatous-polyposis

12 Guillem JG, Wood WC, Moley JF *et al.* ASCO/SSO review of current role of risk-reducing surgery in common hereditary cancer syndromes. *Ann. Surg. Oncol.* 13(10), 1296–1321 (2006).

13 Souslova T, Marple TC, Spiekerman AM, Mohammad AA. Personalized medicine in Alzheimer's disease and depression. *Contemp. Clin. Trials* 36(2), 616–623 (2013).

14 Buyse M, Sargent DJ, Grothey A, Matheson A, De Gramont A. Biomarkers and surrogate end points – the challenge of statistical validation. *Nat. Rev. Clin. Oncol.* 7(6), 309–317 (2010).

15 Slodkowska EA, Ross JS. MammaPrint 70-gene signature: another milestone in personalized medical care for breast cancer patients. *Expert Rev. Mol. Diagn.* 9(5), 417–422 (2009).

16 Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology* 138(6), 2073–2087 (2010).

17 Davoli A, Hocevar BA, Brown TL. Progression and treatment of HER2-positive breast cancer. *Cancer Chemother. Pharmacol.* 65(4), 611–623 (2010).

18 Antonicelli A, Cafarotti S, Indini A *et al.* EGFR-targeted therapy for non-small cell lung cancer: focus on EGFR oncogenic mutation. *Int. J. Med. Sci.* 10(3), 320–330 (2013).

19 Price TJ, Peeters M, Kim TW *et al.* Panitumumab versus cetuximab in patients with chemotherapy-refractory wild-type KRAS exon 2 metastatic colorectal cancer (ASPECCT): a randomised, multicentre, open-label, non-inferiority Phase 3 study. *Lancet Oncol.* 15(6), 569–579 (2014).

20 Group SC, Link E, Parish S *et al.* SLCO1B1 variants and statin-induced myopathy-a genomewide study. *N. Engl. J. Med.* 359(8), 789–799 (2008).

21 Dunn J, Baborie A, Alam F *et al.* Extent of MGMT promoter methylation correlates with outcome in glioblastomas given temozolomide and radiotherapy. *Br. J. Cancer* 101(1), 124–131 (2009).

22 Yin AA, Zhang LH, Cheng JX *et al.* The predictive but not prognostic value of MGMT promoter methylation status in elderly glioblastoma patients: a meta-analysis. *PloS ONE* 9(1), e85102 (2014).

23 *In vitro* companion diagnostic devices. Food and Drug Administration (FDA). www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM262327.pdf

24 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)

25 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).

26 Sieuwerts AM, Mostert B, Bolt-De Vries J *et al.* mRNA and microRNA expression profiles in circulating tumor cells and primary tumors of metastatic breast cancer patients. *Clin. Cancer Res.* 17(11), 3600–3618 (2011).

27 Sleijfer S, Gratama JW, Sieuwerts AM, Kraan J, Martens JW, Foekens JA. Circulating tumour cell detection on its way to routine diagnostic implementation? *Eur J. Cancer* 43(18), 2645–2650 (2007).

28 Eisenhauer EA, Therasse P, Bogaerts J *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* 45(2), 228–247 (2009).

29 Bradbury P, Seymour L. Tumor shrinkage and objective response rates: gold standard for oncology efficacy screening trials, or an outdated end point? *Cancer J.* 15(5), 354–360 (2009).

30 Dhani N, Tu D, Sargent DJ, Seymour L, Moore MJ. Alternate endpoints for screening phase II studies. *Clin. Cancer Res.* 15(6), 1873–1882 (2009).

31 Slamon DJ, Leyland-Jones B, Shak S *et al.* Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* 344(11), 783–792 (2001).

32 Kwak EL, Bang YJ, Camidge DR *et al.* Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N. Engl. J. Med.* 363(18), 1693–1703 (2010).

33 Van De Vijver MJ, He YD, Van't Veer LJ *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* 347(25), 1999–2009 (2002).

34 Buyse M, Loi S, Van't Veer L *et al.* Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J. Natl Cancer Inst.* 98(17), 1183–1192 (2006).

35 Mandrekar SJ, An MW, Sargent DJ. A review of Phase II trial designs for initial marker validation. *Contemp. Clin. Trials* 36(2), 597–604 (2013).

36 Tajik P, Zwinderman AH, Mol BW, Bossuyt PM. Trial designs for personalizing cancer care: a systematic review and classification. *Clin. Cancer Res.* 19(17), 4578–4588 (2013).

37 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).

38 Redman MW, Crowley JJ, Herbst RS, Hirsch FR, Gandara DR. Design of a phase III clinical trial with prospective biomarker validation: SWOG S0819. *Clin. Cancer Res.* 18(15), 4004–4012 (2012).

39 Jiang W, Freidlin B, Simon R. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J. Natl Cancer Inst.* 99(13), 1036–1043 (2007).

40 Song JX. A two-stage patient enrichment adaptive design in phase II oncology trials. *Contemp. Clin. Trials* 37(1), 148–154 (2014).

41 Mackey HM, Bengtsson T. Sample size and threshold estimation for clinical trials with predictive biomarkers. *Contemp Clin. Trials* 36(2), 664–672 (2013).

42 Wang SJ, Hung HM. Adaptive enrichment with subpopulation selection at interim: methodologies, applications and design considerations. *Contemp. Clin. Trials* 36(2), 673–681 (2013).

43 Gu X, Yin G, Lee JJ. Bayesian two-step Lasso strategy for biomarker selection in personalized medicine development for time-to-event endpoints. *Contemp. Clin. Trials* 36(2), 642–650 (2013).

44 Lai TL, Liao OY, Kim DW. Group sequential designs for developing and testing biomarker-guided personalized therapies in comparative effectiveness research. *Contemp. Clin. Trials* 36(2), 651–663 (2013).

45 Jenkins M, Stone A, Jennison C. An adaptive seamless Phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. *Pharm. Stat.* 10(4), 347–356 (2011).

46 Miller RG. Springer series in statistics. In: *Simultaneous statistical inference (2nd Edition).* Springer, Verlag, Germany (1981).

47 Moye LA, Deswal A. Trials within trials: confirmatory subgroup analyses in controlled clinical experiments. *Control. Clin. Trials* 22(6), 605–619 (2001).

48 Cappuzzo F, Ciuleanu T, Stelmakh L *et al.* Erlotinib as maintenance treatment in advanced non-small-cell lung cancer: a multicentre, randomised, placebo-controlled Phase 3 study. *Lancet Oncol.* 11(6), 521–529 (2010).

49 Spiessens B, Debois M. Adjusted significance levels for subgroup analyses in clinical trials. *Contemp. Clin. Trials* 31(6), 647–656 (2010).

50 Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine-reporting of subgroup analyses in clinical trials. *N. Engl. J. Med.* 357(21), 2189–2194 (2007).

51 He P. Identifying cut points for biomarker defined subset effects in clinical trials with survival endpoints. *Contemp. Clin. Trials* 38(2), 333–337 (2014).

52 The FDA's drug review process: ensuring drugs are safe and effective. Food and Drug Administration (FDA).

www.fda.gov/drugs/resourcesforyou/consumers/ucm143534.htm,

53 Pazdur R. Endpoints for assessing drug activity in clinical trials. *Oncologist* 13(Suppl. 2), 19–21 (2008).

54 Flaherty KT, Puzanov I, Kim KB *et al.* Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.* 363(9), 809–819 (2010).

55 Guidance for industry expedited programs for serious conditions – drugs and biologics. Food and Drug Administration (FDA). www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm358301.pdf, (2014)

56 Stacchiotti S, Tamborini E, Marrari A *et al.* Response to sunitinib malate in advanced alveolar soft part sarcoma. *Clin. Cancer Res.* 15(3), 1096–1104 (2009).