

# Integrating biomarkers in clinical trials

*Expert Rev. Mol. Diagn.* 11(2), 171–182 (2011)

Marc Buyse<sup>1</sup>,  
Stefan Michiels<sup>2</sup>,  
Daniel J Sargent<sup>3</sup>,  
Axel Grothey<sup>4</sup>,  
Alastair Matheson<sup>5</sup>  
and Aimery  
de Gramont<sup>6</sup>

<sup>1</sup>International Institute for Drug Development, 30 Avenue Provinciale, 1340 Louvain-la-Neuve, Belgium

<sup>2</sup>Institut Jules Bordet, 127 Boulevard de Waterloo, 1000 Brussels, Belgium

<sup>3</sup>Division of Biomedical Statistics and Informatics, Mayo Clinic Cancer Center, 200 First Street SW, Rochester, MN 55905, USA

<sup>4</sup>Department of Medical Oncology, Mayo Clinic College of Medicine, 200 First Street Southwest, Rochester, MN 55905, USA

<sup>5</sup>Fondation ARCAD, 22 rue Malher, 75004 Paris, France

<sup>6</sup>Hôpital Saint-Antoine, Pavillon Moïana, 184 rue du Faubourg Saint-Antoine, 75012 Paris Cedex 12, France

\*Author for correspondence:

Tel.: +32 106 144 44

Fax: +32 106 188 88

[marc.buyse@iddi.com](mailto:marc.buyse@iddi.com)

Biomarkers have a growing role in clinical trials. With the advent of the targeted therapy era, molecular biomarkers in particular are becoming increasingly important within both clinical research and clinical practice. This article focuses on biomarkers that anticipate the prognosis of individual patients ('prognostic' biomarkers) and on biomarkers that predict how individual patients will respond to specific treatments ('predictive' biomarkers, also called 'effect modifiers'). Specific Phase II and III clinical trial designs are discussed in detail for their ability to validate the biomarker and/or to establish the effect of an experimental therapy in patient populations defined by the presence or absence of the biomarker. Contemporary biomarker-based clinical trials in oncology are used as examples.

**KEYWORDS:** biomarker • predictive marker • prognostic marker • trial design • validation

Clinical trials have revolutionized medicine by providing reliable evidence on the efficacy and safety of new treatments. Until recently, clinical trials were designed and analyzed under the assumption that the effects of treatment were broadly similar in different individuals, and hence the goal of the clinical trial was primarily to provide precise and unbiased estimates of these common effects. Today, the advent of molecular biology has modified the fundamental tenet that the effect of treatment varies only randomly from patient to patient. There are indeed increasing numbers of treatments (especially targeted agents) whose effects vary widely as a function of individual molecular characteristics. When these characteristics are known early on in the course of developing of new treatments, clinical trials can be designed to target only those patients who are expected to benefit. However, more commonly, the exact molecular characteristics that drive a patient's response to a specific treatment are unknown and might only be discovered during the clinical development of the treatment or indeed following its approval. This fact has profound implications for clinical trials; it may indeed revolutionize the way in which trials are planned and executed [1–3]. The purposes of this article are, first, to discuss the ways in which biomarkers (including, but not limited to, molecular characteristics) are validated in clinical trials, and second, to consider how they can be incorporated into clinical trial designs with the goal of optimizing the use of new therapies

in biomarker-defined subgroups of patients. Our examples all come from the field of oncology, where efforts to develop new methodologies for clinical research have been most intense, but further examples abound in the other therapeutic areas [4]. We will focus our discussion on statistical issues, leaving aside practical considerations related to the biomarkers themselves, such as their accessibility in specific tissues (e.g., in fresh, frozen or formalin-fixed paraffin-embedded tissue) or the ease and reliability of their quantification (e.g., through reverse-transcriptase PCR or microarray), all of which greatly contribute to their potential for use in clinical practice.

## Biomarkers versus clinical end points

Following the definition adopted by the Biomarkers Definitions Working Group, a biomarker is defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [5]. First, a distinction can be made between pharmacokinetic/pharmacodynamic biomarkers, which we do not discuss in any detail in this article, and clinical biomarkers – that is, those having clinical utility, which are the focus of this article (TABLE 1). Pharmacokinetic/pharmacodynamic biomarkers are used in early drug development (*in vitro* studies, animal experiments and Phase I trials), while clinical biomarkers are used in Phase II and III clinical trials.

**Table 1. Classes and types of biomarkers.**

Biomarker class	Biomarker type	When measured	Biomarker function
PK/PD	PK	Post-treatment	Ensures that active drug concentrations are generated at tolerated doses
PK/PD	PD proof of mechanism	Post-treatment	Confirms that adequate drug–target interaction has been achieved
PK/PD	PD proof of concept	Post-treatment	Demonstrates that the desired effect is produced on tumor biology when the drug interacts with its intended target
Clinical	Prognostic	Pre-treatment or post-treatment	Forecasts the likely course of disease in a defined population due to the underlying tumor biology, irrespective of treatment
Clinical	Predictive or effect modifier	Pre-treatment or post-treatment	Predicts the likely effect or lack of effect of a specific treatment
Clinical	Surrogate	Post-treatment	Provides early and accurate prediction of both a clinical end point, and the effects of treatment on this end point

PD: Pharmacodynamic; PK: Pharmacokinetic.

Biomarkers can be contrasted with clinical end points, which capture information regarding how a patient feels, functions or survives [6]. Surrogate end points, which may themselves be based upon a biomarker, aim to replace a clinical end point with a faster and more sensitive evaluation of the effect of experimental treatments [7]. This article is concerned with biomarkers that forecast future states – namely prognostic and predictive biomarkers. Prognostic biomarkers predict the likely course of disease in a defined clinical population, irrespective of treatment; for example, lymph node involvement predicts a poor outcome in the management of solid tumors, even though treatment may prolong the survival of patients with and without evidence of nodal involvement. Predictive biomarkers forecast the likely response to treatment; for instance, hormone-receptor status predicts the response to endocrine therapies in breast cancer. Many biomarkers, such as hormone-receptor status in breast cancer, are in fact both prognostic and predictive. The present article will focus on biomarkers that are measured once, typically before a treatment is started, in order to guide treatment choice (TABLE 2). We will not discuss biomarkers that can be measured repeatedly during or after treatment, typically with standard techniques such as molecular imaging or blood sampling, even though these biomarkers arguably hold the greatest potential for future clinical research. Indeed, such dynamic biomarkers might be used to guide treatment choices if their prognostic or predictive ability could be demonstrated, not just for a single measurement taken before treatment, but also for repeated measurements taken over time during and after treatment. Of even greater interest would be dynamic biomarkers that could predict the treatment effects on the clinical end points of interest, as these could potentially be used as surrogate end points [7–11].

### Biomarkers in clinical trials

In this article, we will discuss the validation and use of prognostic and predictive biomarkers in clinical trials. It will be convenient to distinguish several situations, depending on the type of treatment (standard or experimental), the type of biomarker (prognostic or predictive) and whether the biomarker is validated or not (TABLE 2). ‘Validation’, which is discussed in greater detail elsewhere, is the

confirmation by robust statistical methods that a candidate biomarker fulfils a set of conditions that are necessary and sufficient for its use in the clinic [12].

Biomarkers are typically initially identified by retrospective analyses of existing patient series (possibly patients treated in clinical trials with long-term follow-up). Biological considerations obviously play a key role in the initial identification of prognostic and predictive biomarkers and remain important during a biomarker’s evaluation and hopefully eventual adoption into clinical practice. Importantly, however, while biological considerations can help strengthen the case for the adoption of a biomarker, they cannot dispense with statistical validation of the biomarker.

In the remainder of this article, we discuss trial designs involving biomarkers throughout the various phases of treatment and biomarker development. We begin our article by briefly discussing retrospective identification of biomarkers (often, although not always, using data from clinical trials: see first two rows of TABLE 2). We subsequently discuss prospective trial designs aimed at validating biomarkers using a standard therapy (third and fourth rows of TABLE 2). We then proceed to prospective trial designs aimed at testing an experimental therapy using validated biomarkers (fifth and sixth rows of TABLE 2). Finally, we will discuss designs that combine testing an experimental therapy along with the validation of corresponding biomarkers (last two rows of TABLE 2).

### Retrospective identification & validation of prognostic biomarkers

For a biomarker to be prognostic, an association must be demonstrated between the value of the marker at baseline, or changes in the biomarker over time, and a clinical end point, independently of treatment. For a putative prognostic biomarker to be validated, its association with the clinical end point of interest should be demonstrated repeatedly in independent studies, preferably across a range of clinical situations (since, contrary to common belief, heterogeneity is often an asset rather than a problem from a statistical point of view). Retrospective studies may be sufficient for the initial identification and statistical validation of prognostic biomarkers, although the biomarker’s clinical utility may need

**Table 2. Trial designs using biomarkers.**

Trial phase	Treatment	Biomarker type	Validated biomarker	Trial design	Examples
	Standard	Prognostic	No	Retrospective series	MammaPrint™ in early breast cancer Oncotype DX® in early breast cancer
	Standard	Predictive	No	Retrospective analyses of randomized trials	Oncotype DX in early breast cancer (SWOG-8814) KRAS mutations in advanced colorectal cancer (CRYSTAL) EGFR mutations in non-small-cell lung cancer (IPASS)
III	Standard	Prognostic	No	Clinical utility	MINDACT in early breast cancer TAILORx in early breast cancer
III	Standard	Predictive	No	Randomize-all Interaction Biomarker strategy	MARVEL in non-small-cell lung cancer P53 in advanced breast cancer ERCC1 in non-small-cell lung cancer
II	Experimental	Predictive	Yes	Targeted Bayesian	Herceptin in advanced breast cancer BATTLE in non-small-cell lung cancer I-SPY 2 in advanced breast cancer
III	Experimental	Predictive	Yes	Targeted	PETACC-8 in advanced colorectal cancer TOGA in advanced gastric cancer
II	Experimental	Predictive	No	Adaptive parallel Tandem two-step TTP ratio	Dovitinib in HER2-negative advanced breast cancer Saracatinib in pancreatic cancer Molecular profiling in various tumor types
III	Experimental	Predictive	No	Enrichment Prospective subset	IPASS in non-small-cell lung cancer SATURN in non-small-cell lung cancer

TTP: Time to progression.

to be confirmed in prospective studies, as described later in this article. The retrospective phase of the validation process can be illustrated by the MammaPrint™ microarray-based signature developed by The Netherlands Cancer Institute (Amsterdam, The Netherlands) in a small sample of 78 untreated patients with the goal of predicting the occurrence of distant metastases in women with early breast cancer [13]. A retrospective analysis identified a 70-gene signature as a strong prognostic marker for the occurrence of metastases within 5 years of resection. In a larger sample of patients treated at the same institution, patients with a poor-prognosis MammaPrint signature were confirmed to have a much higher risk of distant metastases within 5 years compared with patients with a good-prognosis signature [14]. An independent validation study of the signature was then conducted involving independent samples contributed by several European centers, with results confirming that the gene signature adds prognostic information over and above that provided by a binary risk classifier based on the other known clinical and pathological factors [15]. Although these results were impressive, the clinical usefulness of the signature was still in question, especially because the predictive accuracy of the signature was attenuated with longer follow-up (i.e., including patients who developed distant metastases after 5 years of follow-up) [16,17]. The negative-predictive value of the signature for distant-metastasis-free survival status at 5 years after diagnosis was relatively high (0.9 in the Amsterdam series; 0.84 in the validation series), but the positive-predictive value of the signature was rather modest (0.63 in the Amsterdam series; 0.30 in the validation series). Hence, the signature could not be claimed, in and of itself, to be a sufficiently accurate predictor of which

patients would develop metastases and could not provide the sole basis for a treatment decision. Overall, the clinical utility of this signature, that is, its ability to influence a therapeutic decision, remains to be confirmed in prospective trials [17,18]. In the USA, the development of the commonly used signature, Oncotype DX®, followed similar steps [19].

### Retrospective identification of predictive biomarkers

For a biomarker to be predictive, the baseline value, or changes in the values of the biomarker over time, must be shown to predict the efficacy or toxicity of a treatment, as assessed by a defined clinical end point. For a putative predictive biomarker to be validated, its ability to predict the effects of treatment (or lack thereof) should be demonstrated repeatedly in multiple studies. The statistical identification of predictive markers requires data from randomized trials that include patients with both high and low levels of the biomarker. Retrospective analyses may be sufficient to identify candidate predictive biomarkers and validate them to a degree that enables them to be incorporated into trial design and clinical practice, although definitive evidence may still require prospective clinical trials. The retrospective identification and provisional validation process can again be illustrated by Oncotype DX in early breast cancer. Using data from the Southwest Oncology Group (SWOG)-8814 trial (NCT00929591) [101], a higher recurrence score was demonstrated to predict a larger benefit of chemotherapy given in combination with tamoxifen in postmenopausal women with node-positive, estrogen-receptor-positive tumors [20]. Another notable example of retrospective identification of a predictive

biomarker is that of the *KRAS* mutation in advanced colorectal cancer, which was demonstrated in multiple trials to predict a lack of effect of two EGF receptor (EGFR)-directed monoclonal antibodies, cetuximab and panitumumab [21–24]. Not only had a tumor response to these drugs very rarely, if ever, been observed in patients with *KRAS*-mutated measurable colorectal tumors, but the statistical test for interaction repeatedly demonstrated that the benefit of these drugs on overall survival was significantly greater for patients with *KRAS* wild-type tumors than for patients with *KRAS* mutant tumors (who may not receive any benefit whatsoever) [25]. A reliable retrospective validation of a tissue biomarker requires the availability of biospecimens for all, or almost all, patients enrolled in the clinical trial in order to exclude potential sampling bias. For example, in a first biomarker analysis of the CRYSTAL trial (NCT00154102) [102], less than half of all randomized patients had been tested for the *KRAS* mutation [26]. The population of untested patients seemed to benefit less from cetuximab than the tested population, which suggested that the cetuximab benefit among patients with *KRAS* wild-type tumors might have been overestimated. In a second round of analyses, almost 90% of the randomized patients were tested for *KRAS* mutation, and the potential for bias was therefore almost entirely eliminated [25]. Some of the real-life problems with samples of convenience in gene-expression experiments have been recently illustrated by four trials with genomic subsets submitted to the US FDA [27]. It is also important to emphasize that even a retrospective validation requires a prospective analysis plan and prespecified biomarker cut points [28].

### Prospective validation of prognostic biomarkers

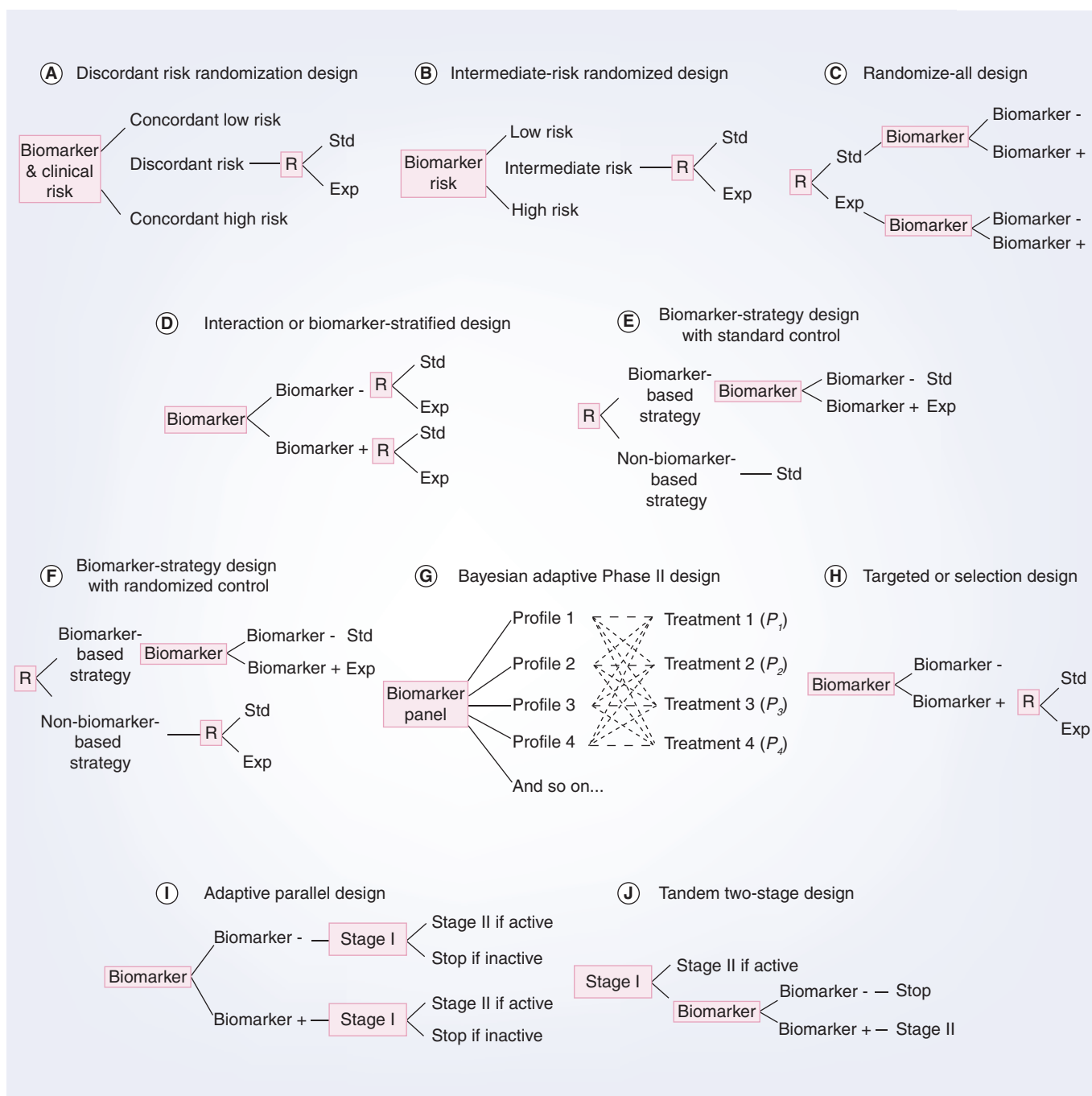
Even for a prognostic biomarker that has been statistically identified and validated in one or several retrospective analyses, the ultimate proof-of-usefulness in the clinic may still require prospective evidence from randomized trials. In particular, prospective studies are often performed to clarify the utility of the biomarker in patients for whom the optimal course of treatment is not apparent from established assessment methods. For example, patients with a low risk of disease recurrence might require only standard therapy (possibly with watchful follow-up) and patients at high risk require experimental therapy, but for patients with intermediate risk (based on the biomarker and/or clinicopathological factors), there may be uncertainty regarding the treatment decision. Such patients could be randomized to either standard or experimental treatment in prospective studies to clarify the role of the biomarker in determining treatment, thereby establishing the clinical utility of the biomarker. Examples of such biomarker-based treatment trials in early breast cancer include the ongoing MINDACT (NCT00433589) [103] and TAILORx trials (NCT00310180) [104]. Although both trials aim to confirm the usefulness of a gene signature (MammaPrint for MINDACT, Oncotype DX for TAILORx) by randomizing patients whose risk of disease progression is uncertain to chemotherapy versus no chemotherapy, they have taken rather different approaches to doing so (FIGURES 1A & 1B). MINDACT randomizes patients who have discordant risk assessments according to their MammaPrint gene

expression and traditional histopathological features [29], while TAILORx randomizes patients with an intermediate Oncotype DX recurrence score [30].

### Prospective validation of predictive biomarkers

Although the predictive potential of a putative biomarker can be suggested by the repeated lack of treatment efficacy in the subset of biomarker-positive patients (such as the lack of anti-tumor activity of cetuximab and panitumumab in patients with *KRAS*-mutated colorectal tumors), the ultimate proof that a biomarker is truly predictive comes from randomized trials in which patients are either biomarker positive or negative and receive either an experimental treatment or a standard treatment. The most straightforward situation is one in which all patients are randomized without regard to biomarker status ('randomize-all' design, FIGURE 1C), and all patients (or a subset of them) are tested afterwards to determine their biomarker status. Although this approach could be used prospectively, it is more commonly used retrospectively in an ongoing prospective trial. For instance, in the CRYSTAL trial (NCT00154102) [102], patients with advanced colorectal tumors were tested for the *KRAS* mutation after being randomized to chemotherapy with or without cetuximab [25,26]. The same approach was adopted for the European Organization for Research and Treatment of Cancer (EORTC) P53 trial (NCT00017095) [105], in which patients with locally advanced/inflammatory or large operable breast cancer were randomized between a taxane and a non-taxane chemotherapy regimen, and later tested for the tumor-suppressing protein P53 using a functional assay in yeast [31]. A potential problem with the delayed biomarker test is that the biomarker status may not be available for all patients (some patients may refuse consent or tissue may no longer be available), in which case it is important to verify that the subset of patients in whom the biomarker status is known is reasonably representative of the total population randomized.

In the prospective setting, that is, if the biomarker is available prior to commencing the trial, it is desirable (though by no means essential) to stratify the patients by their biomarker status and then randomize them ('interaction' design, FIGURE 1D). The benefits of stratification are balancing treatment groups with respect to biomarker status, while also making sure that the biomarker status is known for all patients. This approach was attempted in the MARVEL trial (NCT00738881) [106], in which patients were tested for *EGFR* status and then randomized between erlotinib or pemetrexed as second-line treatment of non-small-cell lung cancer. The analysis was planned to be conducted separately in marker-positive and marker-negative patients, with the use of an interaction test aimed at showing that the treatment effects differed in these two groups. Large numbers of events are generally required to reliably detect interactions and this, in turn, generally requires large patient populations [32]. Realistically, therefore, prospective 'interaction' trials capable of validating predictive markers are likely to be few in number. A meta-analysis of several 'randomize-all' designs is often a more feasible option to validate a predictive biomarker, such as that which has been performed for defective



**Figure 1. Trial designs.** (A) Discordant risk randomized design; (B) intermediate-risk randomized design; (C) randomize-all design; (D) interaction or biomarker-stratified design; (E) biomarker-strategy design with standard control; (F) biomarker-strategy design with randomized control; (G) Bayesian adaptive Phase II design ( $P_1$ ,  $P_2$  and so on: probabilities of allocating treatment 1, treatment 2 and so on); (H) targeted or selection design; (I) adaptive parallel design; (J) tandem two-stage design. +: Positive; -: Negative; Exp: Experimental; R: Randomization; Std: Standard.

mismatch repair as a predictor for 5-fluorouracil efficacy in stage II colon cancer [33]. If there are multiple candidate predictive biomarkers for a ‘randomize-all’ design, permutation tests have been proposed to control the error rate after suitable adjustment for multiple testing [34]. It should be noted that a significant statistical interaction between a biomarker and a treatment does not automatically imply that the biomarker will be useful for treatment

selection, since the treatment could be better for all patients but to an extent that depends on the biomarker value. A more sensible statistical test would exclude a magnitude of effect judged to be clinically worthwhile in subsets defined by the biomarker.

A final type of design addresses the clinical utility of a predictive biomarker. This design compares a standard treatment with a biomarker-strategy that adapts the treatment according to the

biomarker status (FIGURE 1E) [35]. For example, the GILT docetaxel trial (NCT00174629) [107] used DNA excision repair protein (ERCC1) overexpression in tumor RNA (a marker of cisplatin resistance) to customize chemotherapy in patients with advanced non-small-cell lung cancer. Patients were randomly assigned in a 1:2 ratio to either the control arm or the genotypic arm in which ERCC1 was assessed. Patients in the control arm received a standard regimen of docetaxel plus cisplatin. In the genotypic arm, patients with low ERCC1 levels received docetaxel plus cisplatin, and those with high levels received docetaxel plus gemcitabine [36].

There are two main concerns with this design: first, the difference between the two randomized arms is expected to be small, especially if the prevalence of a positive biomarker is low; and second, even if a difference was observed between the randomized arms, it could be due to a better efficacy of the experimental arm, regardless of the biomarker status [37–40]. The latter concern can be addressed in a modified design that compares the biomarker strategy with a randomized comparison of the same treatments, using the same allocation ratio to standard treatment or experimental treatment in the two strategies (FIGURE 1F). However, the former concern remains as only a small difference can be expected between the randomized strategies, and therefore the statistical power for comparing the strategy arm with the nonstrategy arm is very low [41]. Furthermore, owing to the potential inability to distinguish between a prognostic effect of the biomarker and an effect of treatment, this design cannot identify whether differences in outcome result from one or the other of these effects. In general, simulation studies suggest that the two marker-based strategy designs are less efficient than the randomize-all traditional design [40–43].

### Use of biomarkers to optimize treatment selection in Phase II trials

We now turn to the situation in which one or more predictive biomarkers are known or assumed to exist, where the purpose of the trial is not to formally validate these biomarkers, but rather to use them to optimize treatment selection. If a biomarker is truly predictive of the effect of an experimental treatment, then the best strategy will often be to target the subset of patients who are predicted to benefit most. For example, the clinical development of trastuzumab in breast cancer was restricted to patients with *HER2/neu*-amplified tumors, based both on biological considerations and the lack of tumor response in advanced tumors without *HER2/neu* amplification [44,45].

The combined use of information from several biomarkers is also likely to improve the predictive ability of any one of them. This is particularly relevant for biomarkers which, when present, have a high sensitivity for an outcome mandating a particular therapy (i.e., most of the patients having these markers should be treated), and for biomarkers which, when absent, mitigate with high specificity against a particular therapy (i.e., most of the patients not having the marker should not be treated). A Bayesian approach building on these ideas has been employed to design several recent Phase II trials. In the BATTLE trial (NCT00409968) [108], for example, patients with non-small-cell

lung cancer underwent a biopsy and were assigned to one of five mutually exclusive biomarker profiles based on the status of four biomarkers (*EGFR*, *KRAS*/serine/threonine protein kinase [BRAF], VEGF/VEGF receptor [VEGFR], retinoid X receptor [RXR]/cyclin 1) known or assumed to have predictive impact on the effect of the four drugs under investigation: erlotinib, sorafenib, vandetanib and bexarotene [46,47]. Depending on the biomarker group a patient belongs to, one of the drugs or a combination of them is *a priori* indicated most (represented by broken lines in FIGURE 1G). The design is ‘Bayesian adaptive’, that is, a Bayesian approach is used to prespecify the probabilities of allocating patients to any of the drugs being tested (denoted by  $P_1$ ,  $P_2$  and so on in FIGURE 1G), and the observed outcomes (e.g., tumor responses or proportions of progression-free patients at 8 weeks) of patients already treated are used to update these probabilities during the course of the trial. The BATTLE trial used a run-in period of equal randomization before switching to the adaptive randomization period. Early stopping rules were used to exclude possible biomarkers by treatment combinations based on posterior probabilities of obtaining clinical responses in particular subgroups. The I-SPY 2 trial (NCT01042379) [109] uses a similar Bayesian adaptive Phase II design to test five experimental agents versus a standard regimen given as neoadjuvant therapy in patients with operable breast cancer [48]. One downside of this type of design is that the predictive biomarkers have to be known at the start of the trial, which is not always the case in practice. In addition, biomarkers have to be ‘ranked’ with regard to their predictive value for specific treatment approaches to address the situation of when a tumor expresses two or more biomarkers at the same time.

### Use of biomarkers in Phase III trials

When a predictive biomarker exists for an experimental agent, the ‘targeted’ or ‘selection’ approach seems to be most appropriate, whereby only biomarker-positive patients enter the randomized trials aimed at establishing the worth of the new agent (FIGURE 1H). Among the many examples of targeted trials, the TOGA trial (NCT01041404) [110] compared chemotherapy with or without trastuzumab in patients with *HER2/neu*-positive advanced gastric cancer. Such trials have the capacity to confirm the usefulness of the marker in identifying a population in which there is a treatment benefit, but they do not imply that the marker is truly predictive since they provide no information regarding the lack of benefit among marker-negative patients. A key example of such a situation is the effect of trastuzumab in delaying or preventing recurrence in early breast cancer. In patients with *HER2/neu*-amplified tumors, the benefit of treatment has been established by several large randomized trials [49–53]. However, there remains an intriguing suggestion, based on patients without *HER2/neu* amplification who were accidentally entered into the large trials, that treatment may have similar effects in patients with *HER2/neu*-nonamplified tumors [54]. Unfortunately, the targeted nature of the designs used prohibits the development of formal proof of an interaction between the effect of trastuzumab and *HER2/neu* status.

Targeted trials can include different treatment options depending on biomarker values, as in the Eastern Cooperative Oncology Group (ECOG) E5202 trial (NCT00217737) [111] for patients with stage II colon cancer. In this trial, patients whose tumors have microsatellite instability (a putative predictive biomarker of resistance to fluoropyrimidines) and a normal 18q chromosome (a prognostic biomarker) will not receive adjuvant therapy, whereas patients whose tumors have microsatellite stability and 18q chromosomal abnormality will receive adjuvant therapy [55].

One of the key considerations when choosing between a targeted design versus a randomize-all design, besides the opportunity to confirm the predictive nature of the biomarker used to select patients, is the statistical power of each of these designs under different scenarios. While the targeted design may have a much higher power if the biomarker is truly predictive, a randomize-all design may accrue many more patients, especially if the biomarker prevalence is low, and may therefore gain power if the treatment had at least some activity in all patients [40,56–60]. Once a targeted trial is completed, the window of opportunity to conduct another trial in biomarker-negative patients may also have closed.

When trials of long duration are being conducted, knowledge that becomes available regarding predictive biomarkers may require the amendment of a randomize-all design into either a targeted design (with complete exclusion of a subset shown to derive no benefit or harm from treatment) or an enriched design (with preferential accrual of a subset presumed to benefit from treatment). Examples of such situations have recently occurred in trials testing *EGFR* inhibitors for patients with colorectal cancer in both advanced disease and the adjuvant setting, with the N0147 (NCT00079274) [112], PETACC-8 (NCT00265811) [113] and C80405 (NCT00265850) trials [114], progressively focusing on patients with *KRAS* wild-type tumors.

### Phase II codevelopment trials: combined studies of biomarker validation & patient selection

The Phase II and III designs discussed in the previous sections assumed that one or several biomarkers were available to predict treatment response (TABLE 2). However, completely validated biomarkers are rarely available. Typically, a predictive biomarker is proposed because of the assumed biological mechanism of a class of targeted agents, or because one or more retrospective analyses identified it as being a plausible candidate. However, even a strong biological rationale or a single statistically significant analysis requires confirmation, and often confirmatory trials, even those large in size, yield conflicting results that cause confusion and controversy [61]. It usually takes several prospective trials for a biomarker to be completely validated. Hence, future trial designs will typically aim to validate the biomarker and establish treatment benefit simultaneously [62].

In the Phase II setting, several designs have been proposed to make use of a putative predictive biomarker. A first design, termed ‘adaptive parallel’, conducts two two-stage Phase II trials in parallel; one in the biomarker-positive group (expected to benefit more from treatment) and one in the biomarker-negative group of patients [63,64]. After the first stage, the trial may continue in

all patients or only in the biomarker-positive group (FIGURE 1I). One such trial that is currently recruiting is a multicenter, open-label, two-stage, Phase II trial of dovitinib in FGF receptor 1 (FGFR1)-amplified and FGFR1-unamplified metastatic HER2-negative breast cancer (NCT00958971) [65,115].

A second design, termed ‘tandem two-step’, uses a predefined pharmacogenomic biomarker [64,66]. All patients are entered in the first step (or stage), regardless of the biomarker (FIGURE 1J). If the number of clinical responses that are observed in the first stage is large enough, the study proceeds to the second stage in the overall population. If the number of responses observed in the first stage is insufficient, the study accrues only patients in the subgroup predicted by the pharmacogenomic biomarker to be responders, and the study termination is governed by a standard optimal two-stage Phase II trial design in that subgroup of patients. The tandem two-step design was implemented in a Phase II clinical trial of saracatinib as monotherapy in previously treated metastatic pancreatic cancer patients, with a primary 6-month survival end point (NCT00735917) [116]. The primary end point was not reached in the overall population, but a predefined pharmacodiagnostic strategy has now been employed to enrich patients most likely to benefit [67].

Bayesian adaptive randomization methods have been extended to accommodate the development of targeted therapies for which the companion biomarkers are only putative or not known at all at the start of the Phase II trial. The trial is divided into two stages: a learning stage to identify the biomarkers and an adaptive stage, during which these biomarkers are used for adaptive randomization. A follow-up study of the BATTLE trial, the BATTLE-2 trial, will test four treatments in advanced-stage lung cancer using this type of trial design [41]. Another Bayesian adaptive design uses biomarker data and clinical outcome as they become available during the course of the trial to continuously ‘learn’ about the most appropriate biomarkers and update the randomization probabilities. This design is of an exploratory nature as it does not control for type I errors [68].

A completely different type of design has recently been proposed that uses molecular profiling of tumor biopsies from patients who are refractory to conventional chemotherapy in order to identify, among a large panel of approved drugs, those that could potentially be active given their molecular target(s). Such a trial (NCT00530192) [117] was conducted in 106 patients with various metastatic cancers who had failed at least two lines of chemotherapy. The primary objective was to demonstrate that their time to disease progression when on the therapy selected with molecular profiling was longer than the time to progression (TTP) on the last conventional treatment regimen they received; in other words, their TTP ratio would be greater than one [69]. Although the trial demonstrated a TTP ratio greater than one in about a quarter of all patients, the absence of a randomized control group made these results difficult to interpret. The TTP ratio has been used with some success to define the optimal dose of imatinib therapy in patients with gastrointestinal stromal tumors [70,71]. An attractive feature of this outcome measure is that patients act as their own control in terms of their TTP, which is appropriate for cytostatic

agents. However, the variability inherent in TTP for successive lines of treatment may however limit the ability of the TTP ratio to yield reliable information, and a poor correlation between the TTP over successive lines of treatment makes the design inefficient [72]. Use of the TTP ratio is also unlikely to reach definite conclusions in nonrandomized designs [73].

### Phase III codevelopment trials

Sometimes a new agent requires testing against standard therapy in a population thought to be especially responsive based on clinical observations, rather than a known biomarker. This situation is illustrated by the *EGFR* tyrosine kinase inhibitors gefitinib and erlotinib, which had shown higher response rates in certain subsets of patients with non-small-cell tumors of the lung (especially in East Asian female patients who had never smoked and presented with an adenocarcinoma histology) [61]. A Phase III worldwide trial (NCT00242801) [118] comparing gefitinib to best supportive care in 1692 unselected patients with previously treated advanced non-small-cell lung cancer could only detect a nonsignificant trend in favour of gefitinib in spite of the large sample size ( $p = 0.087$ ) [74]. By contrast, the Phase III IPASS trial (NCT00322452) [119] was carried out in Asia to compare single-agent gefitinib with conventional chemotherapy in 1217 Asian patients who were either nonsmokers or former light smokers and had a previously untreated advanced adenocarcinoma [75]. After the trial started, it was discovered that *EGFR* mutations associated with responsiveness to gefitinib were more prevalent in Asian populations, and indeed the better outcome of gefitinib over chemotherapy in the IPASS trial was entirely owing to the subset of patients with *EGFR* mutations [75,76]. Hence the IPASS trial had been enriched in mutation-positive patients, who represented approximately 60% of patients in whom the mutation status could be retrospectively determined. In patients positive for the *EGFR* mutation, progression-free survival was significantly longer for patients treated with gefitinib than chemotherapy (hazard ratio [HR] for progression or death: 0.48) but in mutation-negative patients it was significantly shorter (HR for progression or death: 2.85). Similarly, with respect to overall survival, HRs for death in the gefitinib group in comparison with chemotherapy were 0.78 and 1.38, respectively [75]. These results were prospectively confirmed in at least two more Japanese trials targeting the mutated-*EGFR* population: these trials showed HR for progression or death of 0.36 ( $p < 0.01$ ) and 0.49 ( $p < 0.001$ ), respectively [77,78].

Finally, an increasingly frequent dilemma for the design of Phase III trials in the presence of a potential biomarker is to decide whether the primary analysis of the trial will include all randomized patients or the presumed subset of responsive patients (e.g., the biomarker-positive patients). The trade-off is clear: if all patients benefit from the new treatment, albeit perhaps to different degrees, then the power of a test that compares all randomized patients is likely to be higher; while if biomarker-positive patients benefit far more from treatment, then the power of a test that compares only biomarker-positive patients is likely to be greater. The dilemma can be solved, at the expense of an increase in sample size,

by performing both tests at a lower significance level, in such a way that the overall (also called 'experiment-wise') type I error remains controlled. This approach, in which a 'prospective subset' analysis is planned, was used in the SATURN trial (NCT00556712) [120] for patients with advanced non-small-cell lung cancer. After standard treatment with four cycles of platinum-based chemotherapy, patients who had not yet progressed were randomly allocated to receive erlotinib or placebo until progression or unacceptable toxicity were reached [79]. Progression-free survival after randomization was tested in all patients at a significance level of 0.03 and in the patients whose tumors had *EGFR* protein overexpression at a significance level of 0.02. In this trial, the overall significance level was clearly maintained at 0.05 (the sum of 0.03 and 0.02), but the approach was overly conservative because of the correlation between the two tests (overall and in the subset). A vast amount of literature has been devoted to less conservative, yet properly controlled, ways of adjusting the significance level of both tests [80–84].

In many cases, the assay used to measure biomarker positivity is imperfect. One possible approach in this case is to incorporate the positive-predictive value of the assay to estimate and test the treatment effects in patients who are truly biomarker positive by application of the expectation-maximization algorithm [85]. If the biomarker is known at the start of the trial but its cutoff value has not yet been fully established to define biomarker-positive patients, a biomarker-adaptive design can be applied. This design combines a test for overall treatment effect in all randomly assigned patients with the establishment and validation of a cutpoint for the biomarker-positive patients [81].

Other adaptive designs have also been proposed to help make the decision for enrichment based on the results of interim analyses. The approach consists of starting the trial in the entire patient population, and then stopping accrual in the nonresponsive subgroup based either on a Bayesian decision rule [86] or on evidence of low conditional power [87]. The advantage of the latter approach is that it is based on established group-sequential design methodology. Some authors have adopted the converse approach: they start the trial only in the putatively 'targeted' subgroup of patients in the first stage and allow the trial either to be terminated owing to futility in that subgroup after stage 1 or they start recruitment of the entire patient population in stage 2 [88]. A two-stage design has also been proposed to identify a predictive gene-expression profile and to validate it in a single prospective trial [89]. In the first stage, the gene-expression profile is identified to predict whether a patient is more likely to benefit from the experimental treatment compared with the standard one, by using an interaction test. The gene-expression profile is prospectively applied to identify the subset of sensitive patients among stage 2 patients, rather than to restrict the entry of stage 2 patients. The final analysis of the trial consists of a comparison of the experimental treatment with the standard treatment in the whole trial population, as well as in the subset of the stage 2 sensitive patients, with proper adjustment of the significance level of each test to keep the overall significance level under an acceptable value, such as 0.05. A more recent version of



this design introduces a cross-validation extension of the adaptive signature design that optimizes the efficiency of both the classifier development and the validation components [40]. A challenge for all these adaptive designs is the potential heterogeneity of treatment effects before and after the adaptation because of changes in patient recruitment.

### Expert commentary

The incorporation of biomarkers into clinical trials is likely to soon become the norm rather than the exception. We have reviewed the most frequently used designs in oncology, but we expect that many more designs will be proposed in the near future, especially adaptive ones. Although much of the methodological research in trial design is exciting and promising, one should keep in mind some key statistical features that are desirable in almost all situations: first and foremost, the presence of an appropriate control group; second, a sample size sufficient for the results of the trial to be reliable; and, last but not least, the capacity to question the assumptions implicit in the trial design, for example, in terms of the predictive value of a biomarker used to select therapies.

### Five-year view

Research and medical practice will be driven by combinations of targeted therapies based on biomarkers obtained from primary or metastatic tissue biopsies, other tissues and imaging techniques. Biomarkers will be measured at baseline to select patients and during the course of the trial to assess treatment efficacy and safety. Novel clinical trial designs will allow go/no go decisions to be made earlier, while the use of more sensitive end points will accelerate new drug registration. Large-scale randomized evidence will continue to be needed for reliable validation of early findings.

### Financial & competing interests disclosure

Marc Buyse is a shareholder of the International Drug Development Institute. Daniel J Sargent declares consultancy fees from the following companies: Almac, DiagnoCure, Exiqon, Genomic Health and Precision Therapeutics. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

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

### Key issues

- Retrospective analyses of patient series or randomized trials can be used to identify prognostic and predictive biomarkers, but prospective designs are required to validate them.
- Various Phase II trial designs have been proposed to test potential biomarkers using known drugs, or to test experimental drugs using known biomarkers.
- In Phase III trials, clinical utility designs can be used to validate prognostic biomarkers, while randomize-all, interaction and biomarker-strategy designs can be used to validate predictive biomarkers.
- The latter three designs suffer from low statistical power, hence trials using these designs generally require large numbers of patients.
- Targeted trial designs are preferred when a biomarker is sufficiently reliable to exclude patients unlikely to respond to therapy, whereas enrichment and prospective subset designs are more suitable when a biomarker is suspected but still requires prospective validation.

### References

Papers of special note have been highlighted as:

- of interest
- of considerable interest

- Simon R. An agenda for clinical trials: clinical trials in the genomic era. *Clin. Trials* 1, 468–470 (2004).
- Simon R. New challenges for 21st Century trials. *Clin. Trials* 4, 167–169 (2007).
- Simon R. Clinical trials for predictive medicine – new challenges and paradigms. *Clin. Trials* 7(5), 516–524 (2010).
- Simon R. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. *Per. Med.* 7, 33–47 (2010).
- Biomarkers Definitions Working Group. Biomarkers and surrogate end points: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* 69, 89–95 (2001).
- Temple RJ. A regulatory authority's opinion about surrogate end points. In: *Clinical Measurement in Drug Evaluation*. Nimmo WS, Tucker GT (Eds). Wiley, NY, USA, 3–22 (1995).
- Buyse M, Michiels S. Biomarkers and surrogate end points in clinical trials. In: *Fundamentals of Oncology Clinical Trials*. Kelly WK, Halabi S (Eds). Demos Medical Publishing, VA, USA 215–225 (2010).
- Burzykowski T, Molenberghs G, Buyse M. (Eds). *The Evaluation of Surrogate End Points*. Springer, NY, USA (2005).
- Buyse M, Vangeneugden T, Bijmens L *et al*. Validation of biomarkers as surrogates for clinical end points. In: *Biomarkers in Clinical Drug Development*. Bloom JC, Dean RA (Eds). Marcel Dekker, NY, USA 149–168 (2003).
- Weir CJ, Walley, RJ. Statistical evaluation of biomarkers as surrogate end points: a literature review. *Stat. Med.* 25, 183–203 (2006).
- Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate end points. *Stat. Methods Med. Res.* 17, 303–340 (2008).
- Buyse M, Sargent D, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points – the challenge of validation. *Nat. Rev. Clin. Oncol.* 7, 309–317 (2010).
- van 't Veer LJ, Dai Y, van de Vijver MJ *et al*. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530–536 (2002).
- 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, 1999–2009 (2002).
- 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, 1183–1192 (2006).

- 16 Desmedt C, Piette F, Loi S *et al.* Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin. Cancer Res.* 13, 3207–3214 (2007).
- 17 Hayes DF, Trock B, Harris AL. Assessing the clinical impact of prognostic factors: when is 'statistically significant' clinically useful? *Breast Cancer Res.* 52, 305–319 (1998).
- 18 Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. *N. Engl. J. Med.* 360, 790–800 (2009).
- 19 Subramanian J, Simon R. What should physicians look for in evaluating prognostic gene-expression signatures? *Nat. Rev. Clin. Oncol.* 7, 327–334 (2010).
- 20 Albain KS, Barlow WE, Shak S *et al.*; The Breast Cancer Intergroup of North America. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol.* 11, 55–65 (2009).
- 21 Di Fiore F, Blanchard F, Charbonnier F *et al.* Clinical relevance of *KRAS* mutation detection in metastatic colorectal cancer treated by cetuximab plus chemotherapy. *Br. J. Cancer* 96, 1166–1169 (2007).
- 22 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, 1626–1634 (2008).
- 23 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, 1757–1765 (2008).
- 24 Van Cutsem E, Dicoat M, Arber N *et al.* Molecular markers and biological targeted therapies in metastatic colorectal cancer: expert opinion and recommendations derived from the 11th ESMO/World Congress on Gastrointestinal Cancer, Barcelona, 2009. *Ann. Oncol.* 21(Suppl. 6), vi1–vi10 (2010).
- 25 Van Cutsem E, Lang I, Folprecht G *et al.* Cetuximab plus FOLFIRI: final data from the CRYSTAL study on the association of *KRAS* and *BRAF* biomarker status with treatment outcome. ASCO Annual Meeting Proceedings. *J. Clin. Oncol.* 28, 3570 (2010).
- 26 Van Cutsem E, Kohne C.H, Hitre E *et al.* Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N. Engl. J. Med.* 360, 1408–1417 (2009).
- 27 Wang SJ, O'Neill RT, Hung HJ. Statistical considerations in evaluating pharmacogenomics-based clinical effect for confirmatory trials. *Clin. Trials* 7(5), 525–536 (2010).
- 28 Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J. Natl Cancer Inst.* 101, 1446–1452 (2009).
- 29 Bogaerts J, Cardoso F, Buyse M *et al.* Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. *Nat. Clin. Pract. Oncol.* 3, 540–551 (2006).
- 30 Kim C, Paik S. Gene-expression-based prognostic assays for breast cancer. *Nat. Rev. Clin. Oncol.* 7, 340–347 (2010).
- 31 Therasse P, Carbonnelle S, Bogaerts J. Clinical trials design and treatment tailoring: general principles applied to breast cancer research. *Crit. Rev. Oncol. Hematol.* 59, 98–105 (2006).
- 32 Peterson B, George SL. Sample size requirements and length of study for testing interaction in a 2 x k factorial design when time-to-failure is the outcome. *Control. Clin. Trials* 14, 511–522 (1993).
- 33 Sargent D, Marsoni S, Monges G *et al.* Defective mismatch repair as a predictive marker for lack of efficacy of FU-based adjuvant therapy in colon cancer. *J. Clin. Oncol.* 28, 3219–3227 (2010).
- 34 Michiels S, Pothoff RF, George SL. Multiple testing of treatment-effect-modifying biomarkers in a randomized clinical trial with a survival end point. *Stat. Med.* (2011) (In Press).
- 35 Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. *J. Clin. Oncol.* 23, 2020–2027 (2005).
- 36 Cobo M, Isla D, Massuti B *et al.* Customizing cisplatin based on quantitative excision repair cross-complementing 1 mRNA expression: a Phase III trial in non-small-cell lung cancer. *J. Clin. Oncol.* 25, 2747–2754 (2007).
- 37 Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. *J. Clin. Oncol.* 27, 4027–4034 (2009).
- **Clinically useful discussion of trial designs for predictive biomarker validation.**
- 38 Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: one size does not fit all. *J. Biopharm. Stat.* 19, 530–542 (2009).
- 39 Mandrekar SJ, Sargent DJ. Genomic advances and their impact on clinical trial design. *Genome Med.* 1(7), 69 (2009).
- 40 Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. *J. Natl Cancer Inst.* 102, 152–160 (2010).
- **Discussion and examples of various biomarker-based Phase III trial designs.**
- 41 Lee JJ, Gu X, Liu S. Bayesian adaptive randomization designs for targeted agent development. *Clin. Trials* 7(5), 584–596 (2010).
- 42 Young KY, Laird A, Zhou XH. The efficiency of clinical trial designs for predictive biomarker validation. *Clin. Trials* 7(5), 557–566 (2010).
- 43 Hoering A, LeBlanc M, Crowley JJ. Randomized Phase III clinical trial designs for targeted agents. *Clin. Cancer Res.* 14, 4358–4367 (2008).
- **Detailed comparison of the properties of different biomarker-based designs based on a simulation study.**
- 44 Slamon DJ, Godolphin W, Jones LA *et al.* Studies of the *HER-2/neu* protooncogene in human breast and ovarian cancer. *Science* 244, 707–712 (1989).
- 45 Marty M, Cognetti F, Maraninchi D *et al.* Randomized Phase II trial of the efficacy and safety of trastuzumab combined with docetaxel in patients with human epidermal growth factor receptor 2-positive metastatic breast cancer administered as first-line treatment: the M77001 study group. *J. Clin. Oncol.* 23, 4265–4274 (2005).
- 46 Lee JJ, Liu DD. A predictive probability design for Phase II cancer clinical trials. *Clin. Trials* 5, 93–106 (2008).
- 47 Zhou X, Liu S, Kim ES, Herbst RS, Lee JJ. Bayesian adaptive design for targeted therapy development in lung cancer – a step toward personalized medicine. *Clin. Trials* 5, 181–193 (2008).
- **Detailed discussion of a Bayesian adaptive Phase II trial design and comparison with traditional randomized designs.**
- 48 Barker AD, Sigman CC, Kelloff JG *et al.* I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin. Pharmacol. Ther.* 86, 97–100 (2009).
- 49 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, 783–792 (2001).

- 50 Romond EH, Perez EA, Bryant J *et al.* Trastuzumab plus adjuvant chemotherapy for operable HER-2 positive breast cancer. *N. Engl. J. Med.* 353, 1673–1686 (2005).
- 51 Piccart-Gebhart M, Procter M, Leyland-Jones B *et al.* Trastuzumab after adjuvant chemotherapy in HER-2 positive breast cancer. *N. Engl. J. Med.* 353, 1659–1672 (2005).
- 52 Slamon DJ, Eiermann W, Robert N *et al.* Phase III randomized trial comparing doxorubicin and cyclophosphamide followed by docetaxel with doxorubicin and cyclophosphamide followed by docetaxel and trastuzumab with docetaxel, carboplatin and trastuzumab in HER2 positive early breast cancer patients: BCIRG 006 study. *Breast Cancer Res. Treat.* 94(Suppl. 1) (2005) (Abstract 1).
- 53 Joensuu H, Kellokumpu-Lehtinen P-L, Bono P *et al.* Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer. *N. Engl. J. Med.* 354, 809–820 (2006).
- 54 Paik S, Kim C, Wolmark N. HER2 status and benefit from adjuvant trastuzumab in breast cancer. *N. Engl. J. Med.* 358(13), 1409–1411 (2008).
- 55 Gangadhar T, Schilsky R. Molecular markers to individualize adjuvant therapy for colon cancer. *Nat. Rev. Clin. Oncol.* 7, 318–325 (2010).
- 56 Sargent D, Allegra C. Issues in clinical trial design for tumor marker studies. *Semin. Oncol.* 29, 222–230 (2002).
- 57 Maitournam A, Simon R. On the efficiency of targeted clinical trials. *Stat. Med.* 24, 329–339 (2005).
- 58 Simon R, Maitournam A. Evaluating the efficiency of targeted designs for randomized clinical trials. *Clin. Cancer Res.* 10, 6759–6763 (2004); correction and supplement 12, 3229 (2006).
- 59 Simon R. Using genomics in clinical trial design. *Clin. Cancer Res.* 14, 5984–5993 (2008).
- 60 George SL. Statistical issues in translational cancer research. *Clin. Cancer Res.* 14, 5954–5958 (2008).
- **Easy-to-understand formulae for the comparison of targeted versus nontargeted trial designs.**
- 61 Govindan R. INTERESTing biomarker to select IDEAL patients for epidermal growth factor receptor tyrosine kinase inhibitors: yes, for *EGFR* mutation analysis, others, I PASS. *J. Clin. Oncol.* 28, 713–715 (2010).
- 62 Taube SE, Clark GM, Dancey JE, McShane LM, Sigman CC, Gutman SI. A perspective on challenges and issues in biomarker development and drug and biomarker codevelopment. *J. Natl Cancer Inst.* 101, 1453–1463 (2009).
- 63 Jones CL, Holmgren E. An adaptive Simon two-stage design for Phase 2 studies of targeted therapies. *Contemp. Clin. Trials* 28, 654–661 (2007).
- 64 McShane LM, Hunsberger S, Adjei AA. Effective incorporation of biomarkers into Phase II trials. *Clin. Cancer Res.* 15, 1898–1905 (2009).
- 65 André F, Baselga J, Ellis MJ *et al.* Study CTKI258A2202: a multicenter, open-label Phase II trial of dovitinib (TKI258) in *FGFR1*-amplified and nonamplified HER2-negative metastatic breast cancer. *J. Clin. Oncol.* 28(Suppl. 15) (2010) (Abstract TPS122).
- 66 Pusztai L, Anderson K, Hess KR. Pharmacogenomic predictor discovery in Phase II clinical trials for breast cancer. *Clin. Cancer Res.* 13, 6080–6086 (2007).
- 67 Nallapareddy S, Arcaroli J, Touban B *et al.* A Phase II trial of saracatinib (AZD0530), an oral Src inhibitor, in previously treated metastatic pancreatic cancer. Presented at: 2010 ASCO *Gastrointestinal Cancers Symposium*. Orlando, FL, USA, 22–24 January 2010 (Abstract 165).
- 68 Eickhoff JC, Kim K, Beach J, Kolesar JM, Gee JR. A Bayesian adaptive design with biomarkers for targeted therapies. *Clin. Trials* 7(5), 546–556 (2010).
- 69 Von Hoff DD, Stephenson JJ Jr, Rosen P *et al.* Pilot study using molecular profiling of patients' tumors to find potential targets and select treatments for their refractory cancers. *J. Clin. Oncol.* 28, 4877–4883 (2010).
- 70 Debiec-Rychter M, Sciort R, Le Cesne A *et al.* *KIT* mutations and dose selection for imatinib in patients with advanced gastrointestinal stromal tumours. *Eur. J. Cancer* 42, 1093–1103 (2006).
- 71 Zalberg JR, Verwij J, Casali PG *et al.* Outcome of patients with advanced gastro-intestinal stromal tumours crossing over to a daily imatinib dose of 800 mg after progression on 400 mg. *Eur. J. Cancer* 41, 1751–1757 (2005).
- 72 Mick R, Crowley JJ, Carroll RJ. Phase II clinical trial design for noncytotoxic anticancer agents for which time to disease progression is the primary end point. *Control. Clin. Trials* 21, 343–359 (2000).
- 73 Doroshow JH. Selecting systemic cancer therapy one patient at a time: is there a role for molecular profiling of individual patients with advanced solid tumors. *J. Clin. Oncol.* 28, 4869–4871 (2010).
- 74 Thatcher N, Chang A, Parikh P *et al.* Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *Lancet* 366, 1527–1537 (2005).
- 75 Mok TS, Wu Y-L, Thongprasert S *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary carcinoma. *N. Engl. J. Med.* 361, 947–957 (2009).
- 76 Lynch TJ, Bell DW, Sordella R *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* 350, 2129–2139 (2004).
- 77 Maemondo M, Inoue A, Kobayashi K *et al.* Gefitinib or chemotherapy for non-small-cell lung cancer with mutated *EGFR*. *N. Engl. J. Med.* 362(25), 2380–2388 (2010).
- 78 Mitsudomi T, Morita S, Yatabe Y *et al.* Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised Phase 3 trial. *Lancet Oncol.* 11(2), 121–128 (2010).
- 79 Cappuzzo F, Ciuleanu T, Stelmakh L *et al.* Erlotinib as maintenance treatment in advanced non-small-cell lung a multicentre, randomised, placebo-controlled Phase 3 study. *Lancet Oncol.* 11(6), 521–529 (2010).
- 80 Wang SJ, O'Neill RT, Hung HM. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. *Pharm. Stat.* 6, 227–244 (2007).
- 81 Jiang, Freidlin B, Simon R. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J. Natl Cancer Inst.* 99, 1036–1043 (2007).
- 82 Wang SJ, Hung HM, O'Neill RT. Adaptive patient enrichment designs in therapeutic trials. *Biomed. J.* 51, 358–374 (2009).
- 83 Alesh M, Huque MF. A flexible strategy for testing subgroups and overall population. *Stat. Med.* 28, 3–23 (2009).
- 84 Spiessens B, Debois M. Adjusted significance levels for subgroup analyses in clinical trials. *Contemp. Clin. Trials* 31(6), 647–656 (2010).

- 85 Liu JP, Lin JR, Chow SC. Inference on treatment effects for targeted clinical trials under enrichment design. *Pharm. Stat.* 8(4), 356–370 (2009).
- 86 Brannath W, Zuber E, Branson M *et al.* Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Stat. Med.* 28, 1445–1463 (2009).
- 87 Mehta C, Gao P, Bhatt DL *et al.* Optimizing trial design: sequential, adaptive, and enrichment strategies. *Circulation* 119, 597–605 (2009).
- **Group sequential framework to enrich for a targeted subgroup in a Phase III trial after interim analysis.**
- 88 Liu A, Liu C, Li Q, Yu KF, Yuan VW. A threshold sample-enrichment approach in a clinical trial with heterogeneous subpopulations. *Clin. Trials* 7(5), 537–545 (2010).
- 89 Freidlin B, Simon R. Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clin. Cancer Res.* 11(21), 7872–7878 (2005).
- http://clinicaltrials.gov/ct2/results?term=NCT00310180
- 105 Clinicaltrials.gov: NCT00017095. Biomarker (*p53* gene) analysis and combination chemotherapy followed by radiation therapy and surgery in treating women with large operable or locally advanced or inflammatory breast cancer http://clinicaltrials.gov/ct2/results?term=NCT00017095
- 106 Clinicaltrials.gov: NCT00738881. Pemetrexed or erlotinib as second-line therapy in treating patients with advanced non-small cell lung cancer http://clinicaltrials.gov/ct2/results?term=NCT00738881
- 107 Clinicaltrials.gov: NCT00174629. GILT docetaxel – non-small cell lung cancer http://clinicaltrials.gov/ct2/results?term=NCT00174629
- 108 Clinicaltrials.gov: NCT00409968. BATTLE program: umbrella protocol for patients with NSCLC http://clinicaltrials.gov/ct2/results?term=NCT00409968
- 109 Clinicaltrials.gov: NCT01042379. I-SPY 2 trial: neoadjuvant and personalized adaptive novel agents to treat breast cancer http://clinicaltrials.gov/ct2/results?term=NCT01042379
- 110 Clinicaltrials.gov: NCT01041404. TOGA study - a study of herceptin (Trastuzumab) in combination with chemotherapy compared with chemotherapy alone in patients with HER2-positive advanced gastric cancer http://clinicaltrials.gov/ct2/results?term=NCT01041404
- 111 Clinicaltrials.gov: NCT00217737. Oxaliplatin, leucovorin, and fluorouracil with or without bevacizumab in treating patients who have undergone surgery for stage II colon cancer http://clinicaltrials.gov/ct2/results?term=NCT00217737
- 112 Clinicaltrials.gov: NCT00079274. Comparison of combination chemotherapy regimens with or without cetuximab in treating patients who have undergone surgery for stage III colon cancer http://clinicaltrials.gov/ct2/results?term=NCT00079274
- 113 Clinicaltrials.gov: NCT00265811. Combination chemotherapy with or without cetuximab in treating patients with stage III colon cancer that was completely removed by surgery http://clinicaltrials.gov/ct2/results?term=NCT00265811
- 114 Clinicaltrials.gov: NCT00265850. Cetuximab and/or bevacizumab combined with combination chemotherapy in treating patients with metastatic colorectal cancer http://clinicaltrials.gov/ct2/results?term=NCT00265850
- 115 Clinicaltrials.gov: NCT00958971. Safety and efficacy of TKI258 in FGFR1 amplified and non-amplified metastatic HER2 negative breast cancer http://clinicaltrials.gov/ct2/results?term=NCT00958971
- 116 Clinicaltrials.gov: NCT00735917. AZD0530 in treating patients with previously treated metastatic pancreatic cancer http://clinicaltrials.gov/ct2/results?term=NCT00735917
- 117 Clinicaltrials.gov: NCT00530192. Molecular profiling protocol (SCRI-CA-001) http://clinicaltrials.gov/ct2/results?term=NCT00530192
- 118 Clinicaltrials.gov: NCT00242801. Iressa vs best supportive care – 2nd/3rd line survival study http://clinicaltrials.gov/ct2/results?term=NCT00242801
- 119 Clinicaltrials.gov: NCT00322452. First line IRESSA™ versus carboplatin/paclitaxel in Asia http://clinicaltrials.gov/ct2/results?term=NCT00322452
- 120 Clinicaltrials.gov: NCT00556712. A study of tarceva (erlotinib) following platinum-based chemotherapy in patients with advanced, recurrent, or metastatic non-small cell lung cancer (NSCLC) http://clinicaltrials.gov/ct2/results?term=NCT00556712

### Websites

- 101 Clinicaltrials.gov: NCT00929591. Tamoxifen with or without combination chemotherapy in postmenopausal women who have undergone surgery for breast cancer http://clinicaltrials.gov/ct2/results?term=NCT00929591
- 102 Clinicaltrials.gov: NCT00154102. Cetuximab combined with irinotecan in first-line therapy for metastatic colorectal cancer (CRYSTAL) http://clinicaltrials.gov/ct2/results?term=NCT00154102
- 103 Clinicaltrials.gov: NCT00433589. Genetic testing or clinical assessment in determining the need for chemotherapy in women with breast cancer that involves no more than 3 lymph nodes http://clinicaltrials.gov/ct2/results?term=NCT00433589
- 104 Clinicaltrials.gov: NCT00310180. Hormone therapy with or without combination chemotherapy in treating women who have undergone surgery for node-negative breast cancer (the TAILORx trial)