



# Introduction to Clinical Trials, Clinical Trial Designs, and Statistical Terminology Used for Predictive Biomarker Research and Validation

Karla V. Ballman

## Introduction

Clinical research studies involving human patients or participants generally have two main variables of interest: *participant exposure* and *participant outcome*. In the context of biomarker studies in cancer research, the exposure would be the biomarker value for a patient, and the outcome might be survival. The distinguishing feature between a *retrospective study* and *prospective study* is what is known about the patient exposure and patient outcome at the time the study is designed. For a retrospective study, investigators look back into time to ascertain patient exposures (e.g., the biomarker value) and the patient outcome of interest (e.g., cancer survival). For a prospective study, the patient exposure of interest is known at the time the patient is included in the study (e.g., baseline biomarker value), and the patient is followed into the future to ascertain the outcome of interest (e.g., survival). As depicted in Fig. 2.1, in a retrospective study, the biomarker value and outcome for a patient are known by the start of the study.

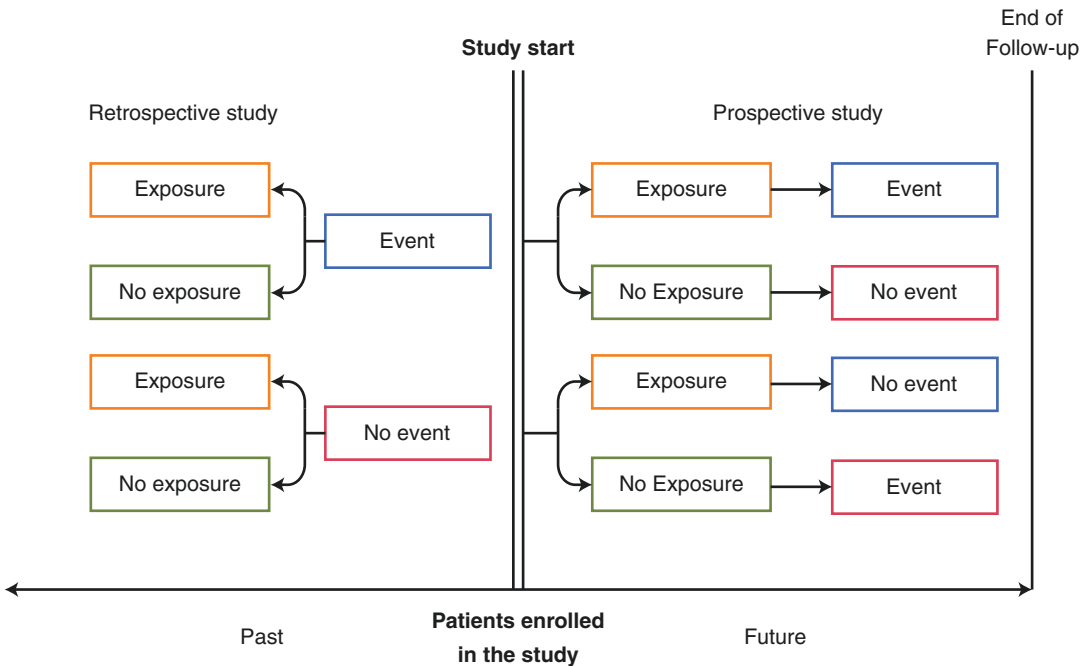
In contrast, in a prospective study the outcome of interest has not yet occurred at the start of the study, and patients are followed into the future

until the end of the study to determine their outcome.

Retrospective studies are limited by various confounding factors that introduce biases. In cancer biomarker studies, they are useful for the discovery of potential biomarkers to be explored in future studies but generally are not sufficient for biomarker validation. More definitive biomarker studies are based on data from prospective studies. For the purpose of establishing a treatment benefit of a predictive biomarker, the prospective study requires (1) a patient group that spans the biomarker outcomes (for a *dichotomous marker*, the study needs biomarker-positive and biomarker-negative patients; for a *continuous marker*, the study needs a group of patients that have biomarker values that represent the range of possible values), and across the biomarker values, it needs (2) patients treated with the treatment of interest and patient not treated with the treatment of interest (likely treated with a different treatment). The strongest design is one in which patients are *randomized* to the treatments as is done in a clinical trial. If patients are not randomized to treatment, the study will likely suffer from patient selection bias, similar to a retrospective study. The remainder of this chapter focuses on predictive biomarker studies in cancer that are based on clinical trial data. Sometimes, the biomarker study is conducted well after the clinical trial has been completed, but this still qualifies as a prospective study because at the time the patients were

---

K. V. Ballman  
Department of Healthcare Policy and Research,  
Division of Biostatistics and Epidemiology,  
Weill Cornell Medicine, New York, NY, USA  
e-mail: [kab2053@med.cornell.edu](mailto:kab2053@med.cornell.edu)



**Fig. 2.1** Prospective studies identify patients, determine or assign their exposure, and then follow patients forward from that time until they have an event of interest or the end of the study. Retrospective studies enroll patients and

then look backward in time from that point to ascertain their exposure status and whether they had the event of interest or not

enrolled on the trial, their baseline biomarker status was fixed (although it might not have been measured until much later), and patients were followed forward into the future for their outcomes.

A brief overview of the different phases of clinical trials is presented in section “[An Overview of Oncology Clinical Trial Designs](#).” Section “[Analysis of Clinical Trial Data](#)” provides a general description of clinical trial data analysis methods. The definition and characteristics of prognostic and predictive biomarkers are presented in section “[Biomarkers in Clinical Trials](#).” The interplay of biomarkers and clinical trial design is explored in section “[Use of Forest Plots](#).” Concluding remarks are made in section “[Biomarker Clinical Trial Designs](#).”

## An Overview of Oncology Clinical Trial Designs

Oncology clinical trials are performed in different settings and by different groups. Some trials

are initiated and led by an investigator that is a member of a cancer center within an academic medical center. These trials may be funded by a pharmaceutical company, the academic medical center, philanthropic funds, or a grant from the government (e.g., the National Cancer Institute, Department of Defense) or a nonprofit organization (e.g., Stand Up to Cancer). It is often the case that the funding comes from one or more of these sources. The principal investigator has control over the data, the data analyses, and the publication of results in investigator-initiated trials.

Pharmaceutical companies also conduct clinical trials. These trials are led and funded solely by the pharmaceutical company, and the company performs the data analysis and disseminates the trial results via publications. The National Cancer Institute (NCI) conducts the majority of government-funded trials, which includes internal trials as well as trials done by other institutions that are funded by NCI grants and contracts. Other government agencies that conduct or sponsor oncology clinical trials include the Department

of Defense and the Department of Veteran's Affairs. Finally, the NCI also funds and supports the National Clinical Trial Network (NCTN) that includes four groups that conduct trials for adult cancer patients (Alliance for Clinical Trials in Oncology, ECOG-ACRIN Cancer Research Group, NRG Oncology, and SWOG) and one group that conducts trials for pediatric cancer patients (Children's Oncology Group). About half of all patients who participate in a cancer clinical trial in a given year do so in a NCTN-led trial. Trials conducted by the NCI NCTN often receive additional support from pharmaceutical companies and/or nonprofit organizations. However, the data analyses leading to publications are conducted independently of the other funding sponsors. Data from any trial funded by a government agency is required to be deposited in a public repository.

There are four general types of clinical trial phases used for drug development in oncology. A drug development plan usually starts with a phase I trial and proceeds through the other phases in a sequential manner if the previous phase is deemed to be a positive trial. A phase I trial is the first time the drug regimen (e.g., a single drug or a new combination of drugs) is being used in humans. These trials are generally small and are designed to find a safe dose to be used in a phase II trial. Typically, sample sizes for a phase I trial are between 10 and 80 patients. The number of patients depends on the number of dose levels to be tested. A positive phase I trial establishes a dose level that is tolerable (has limited adverse events) and thought likely to be active.

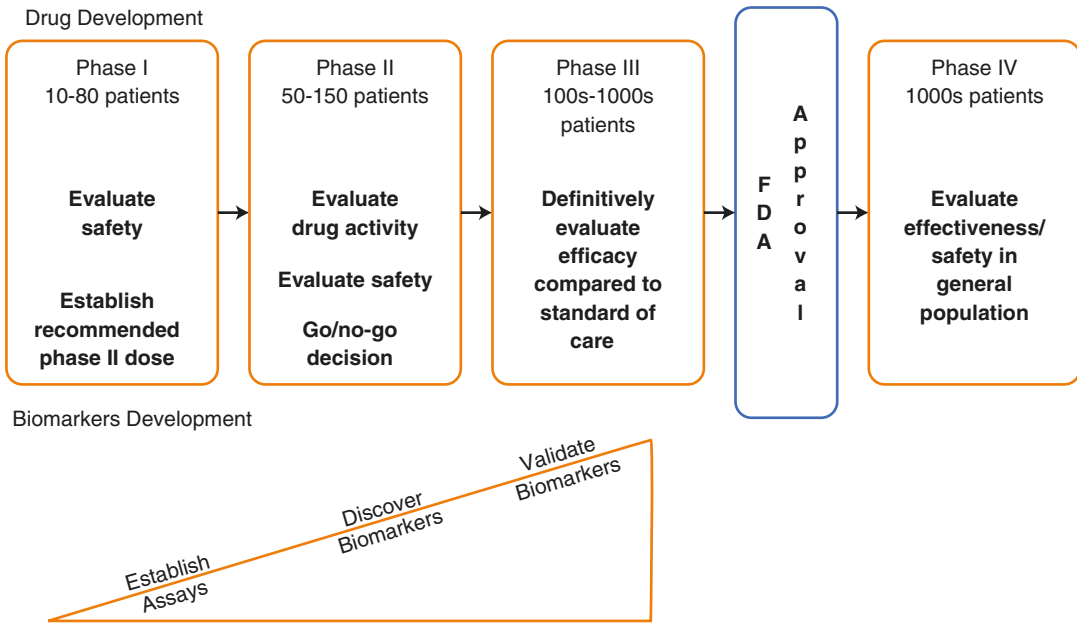
Phase II trials generally enroll on the order of 50–150 patients. The sample size is primarily driven by the number of treatment arms included in the trial. The purpose of a phase II trial is to further evaluate the safety of the drug regimen and to evaluate whether it has potential activity or efficacy. The decision rule is cast as a go/no-go decision. Specifically, if the clinical activity of the drug appears unpromising and/or the drug appears to be too toxic, the decision will be not to perform future trials with the regimen. On the other hand, if the activity level appears promising and the regimen appears to be relatively tolerable, the drug will likely be tested in a phase III

trial. Measures of clinical activity depend on the patient population and the postulated mechanism of action of the drug regimen. Some examples include tumor shrinkage, often measured as the tumor response rate, or a decrease in an established biomarker such as PSA for prostate cancer. Phase II trials can be single-arm trials where all patients receive the drug regimen, or they can be multi-armed where patients are randomized to the arms. Examples of multi-armed trials are a comparison among several different new regimens to select the best one to test in a phase III trial, a comparison of the new regimen to a control arm or a comparison of several different dosing regimens in order to optimize the regimen delivery for a phase III trial.

The sample size for a phase III trial is generally in the range of a few hundred patients to a few thousand patients. The goal is to evaluate the efficacy of the drug regimen. In a phase III trial, patients are randomized to a new regimen or to a control group. Depending on the disease, the control group could be treated with a placebo, if the disease is not life threatening or if there are no approved treatments available for the patient population, or standard of care, in the case of life-threatening disease for which there is an established treatment available. A phase III trial could test several different interventions but always has a control arm. Phase III trials are generally considered to be definitive trials. A positive phase III trial shows that a new regimen has a beneficial effect compared to the current standard of care, i.e., the control arm. If a phase III trial is positive, it usually changes the standard of care and could be the basis for FDA approval of the drug for use in the patient population in which the trial was conducted.

Phase IV studies are conducted after a drug regimen has been marketed and typically involves several thousand patients. The focus of these studies is to monitor the effectiveness of the drug regimen in the general population. It also collects information regarding adverse effects. Phase IV studies have uncovered adverse events that were not observed in previous clinical trials that are due to patient comorbidities or drug-drug interactions.

Within the phase I–IV paradigm of drug development, biomarker discovery may start in



**Fig. 2.2** Design phases for cancer drug development and an indication of the biomarker activities that parallel each of the drug development phases. The size of the triangle

for the biomarker development represents the level of evidence for the utility of the biomarker as well as the number of samples typically involved

phase I trials but is often limited to preliminary exploration or proof-of-concept because of the small sample sizes. Phase II studies are generally the platform for initial biomarker discovery studies and identify markers to be evaluated further in phase III trials. The most informative biomarker studies are part of phase III trials because their larger sample sizes afford more power and because they randomize patients to the drug regimen of interest and a control arm. A phase III study could be used for biomarker discovery, it could be used to validate a proposed biomarker, or the biomarker could be used to determine patient treatment. Figure 2.2 summarizes the roles of the different stages of clinical trial design and biomarker development.

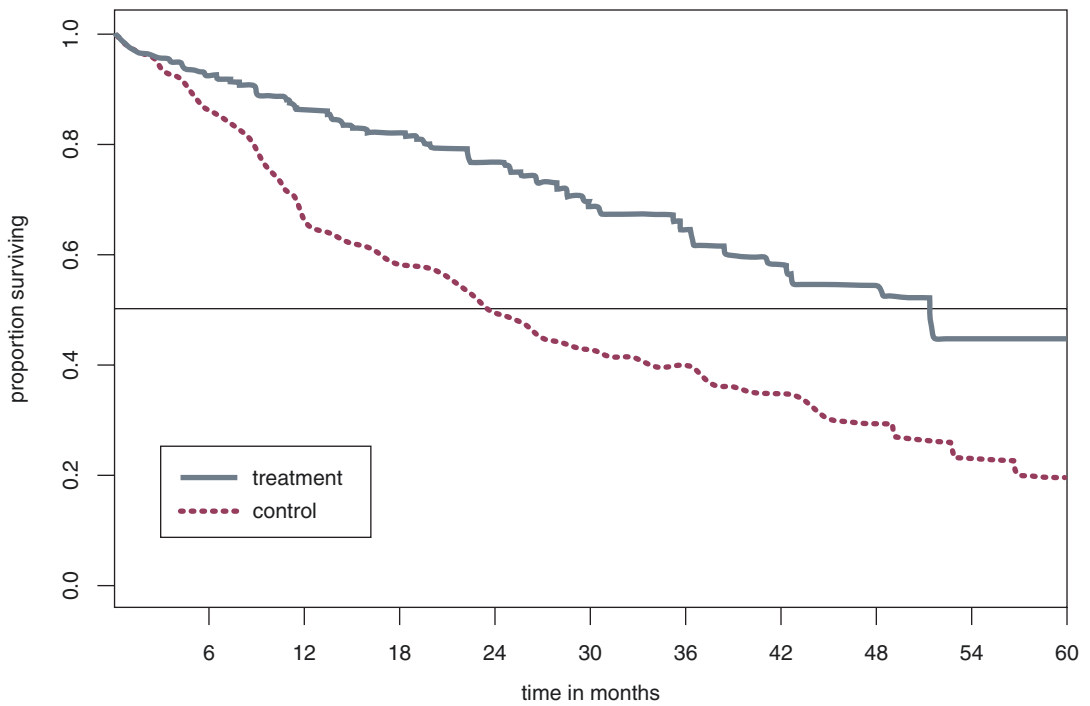
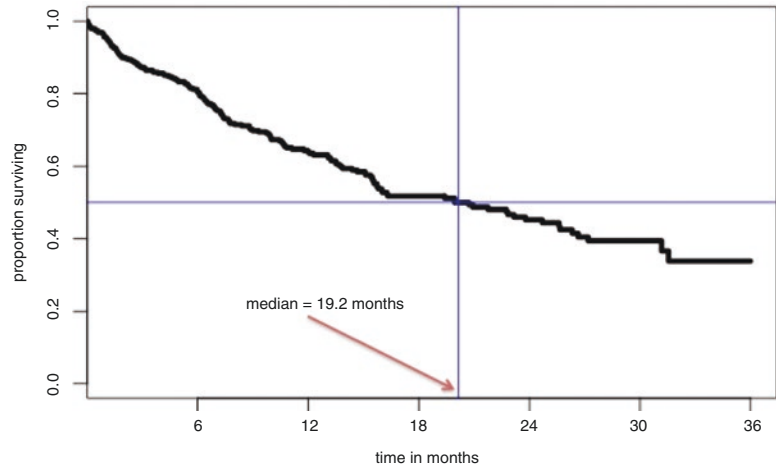
## Analysis of Clinical Trial Data

The statistical method to be used in evaluating data from a clinical trial depends on the outcome of interest. For the sake of brevity, it is assumed the outcome of interest is a time-to-event measure such as overall survival (OS), disease-free

survival (DFS), or progression-free survival (PFS). From this point the outcome will be described generically as survival but could be any measure that involves time from study start for a patient to an event where some patients are censored (i.e., they did not have the event by the end of the follow-up period). For a single-arm trial or the analysis of a single group, the survival time is summarized with a Kaplan-Meier (KM) curve. A KM curve estimates the proportion of patients who have survived as a function of time since treatment initiation (see Fig. 2.3). The median survival is often reported and represents the time point at which 50% of the patients have not survived (or had the event), implying that 50% have survived (or are event-free).

KM curves can be used to compare survival times of two or more groups when they are plotted on the graph. For example, Fig. 2.4 compares the survival times between patients randomized to a new experimental treatment (T) and patients randomized to a control group (C). It is clear that the T group has better survival in general than the C group. This is also demonstrated by comparing the estimated median survival times: 45.1 months

**Fig. 2.3** An example of a Kaplan-Meier curve plot for a group of patients that have a median survival of 20.1 months. The median is the value for which 50% of the patients are still alive (equivalent to 50% have died)



**Fig. 2.4** A display of two Kaplan-Meier curves for survival with one corresponding to patients in the treatment group (solid gray line) and one corresponding to patients

in the control group (dashed maroon line). The median survival for the treatment group is 49.2 months, and the median survival for the control group is 22.7 months

for group T compared to 26.3 months for group C. A *log-rank test* is used to determine whether the observed difference in the KM curves is likely due to chance alone ( $p\text{-value} \geq 0.05$ ) or is deemed statistical significant ( $p\text{-value} < 0.05$ ), which implies there is a treatment effect. The log-rank

$p\text{-value} = 0.0035$  for the curves in Fig. 2.4 shows that the patients in the treatment group appear to have a significantly better survival than patients in the control group. The log-rank test can also be used to evaluate whether there are differences in survival times among any number of groups.

Biomarker classification can also be used to define the patient groups to be compared. Suppose that a biomarker classifies patients into marker-positive (BM+) and marker-negative (BM-) groups. From Fig. 2.5 it appears as though the BM+ group has (very) slightly better survival compared to the BM- group; however, this difference is not statistically significant ( $p$ -value = 0.33). The conclusion in this case would be that the biomarker does not appear to be significantly associated with survival. An example of a biomarker that is not significantly associated with overall survival is PD-L1 protein expression in early-stage non-small cell lung cancer (NSCLC) [1] patients.

A question of interest might be whether there is an association of the biomarker and survival when adjusting for treatment group. Note that the biomarker analysis in Fig. 2.5 includes pooled patients across treatment groups meaning that the BM+ group contains patients in the treatment group as well patients in the control group and the BM- group contains patients in the treatment group as well as the control group. In the PD-L1 study referenced above, the BM+ group are all patients who are PD-L1 positive pooling across those who were and were not treated with adjuvant chemotherapy, and the BM- group are patients who are PD-L1 negative regardless of treatment. When the evaluation of the association with survival involves more than one variable,

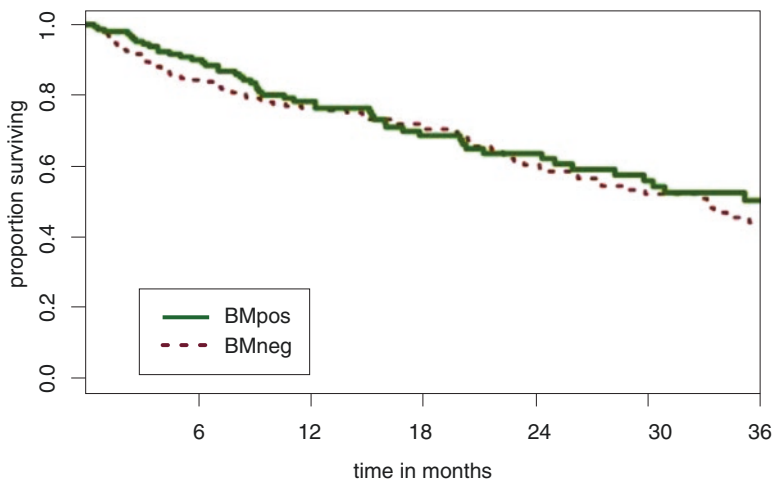
such as treatment group and biomarker status, statistical modeling is used, which in this case would be a *Cox proportional hazards model*. The relationship of each explanatory variable in the model and survival (the outcome variable) is summarized with a hazard ratio (HR), which is the ratio of the hazard of dying at a point in time for each group. The proportional hazard component of the model assumes that this ratio remains constant over all time points. A HR of 1.0 indicates there is no association between the variable and survival. Table 2.1 contains the *univariable HRs* for treatment group and biomarkers status.

The HR comparing the survival of the treatment group to the control group is  $HR = 0.62$ , which is less than one, and it is statistically significant ( $p$ -value = 0.0038). This means that patients in the treatment group are less likely to die than patients in the control group. (If the HR were greater than 1, this means that patients in

**Table 2.1** Univariable estimates of the hazard ratio (HR) for treatment group and biomarker status group with 95% confidence intervals (CIs) and  $p$ -values

Variable	HR (95% CI)	$p$ -value
Treatment group		0.0038
Control	1.00 (reference)	
Treatment	0.62 (0.45, 0.86)	
BM status		0.33
BM-	1.00 (reference)	
BM+	0.85 (0.61, 1.18)	

**Fig. 2.5** The Kaplan-Meier curves for survival for the BM+ group (green solid line) and the BM- group (red dashed line)



the treatment group are more likely to die than patients in the control group.) The best estimate of the treatment HR is 0.62, but there is uncertainty associated with the estimate. Confidence intervals (CIs) are used to convey the precision of the estimate, and 95% CIs are the most commonly used. This is an interval for which there is a 95% probability that it contains the true HR. The 95% CI for the HR = 0.62 is 0.45–0.86. This interval does not contain one, which is consistent with the conclusion that the association of treatment with survival is statistically significant. The conclusion of the univariable analysis of the treatment variable is that it appears that the treatment is associated with longer survival compared to standard of care (control arm).

The univariable HR for the biomarker is HR = 0.85 (95% CI, 0.61–1.18) with a  $p$ -value of 0.33. The 95% confidence interval contains 1 and the  $p$ -value is not statistically significant. It appears as though the biomarker is not associated with survival. Note that the conclusions based on the univariable Cox models are consistent with those from the KM analysis with the log-rank test, which is almost always the case.

A multivariable Cox model is used to evaluate the association of the biomarker with survival while adjusting for the treatment to which the patient was randomized. The multivariable model has both the treatment group and biomarker group as explanatory variables. Table 2.2 contains the *adjusted HRs* for the variables in the multivariable Cox model.

The multivariable HR for the biomarker classification is HR = 0.85 (95% CI: 0.61–1.19), and its

$p$ -value is 0.35. The estimate of the association between the biomarker and survival did not change (only the upper value of the 95% CI changed slightly) when adjusting for treatment assignment, and the  $p$ -value did change slightly but is still not significant. The conclusion would be that the biomarker does not appear to be associated with survival when adjusting for the treatment to a patient received. The lack of change between the univariable and multivariable HR estimates indicates that the effects of treatment and biomarker are not related. Returning to the PD-L1 and NSCLC example, the univariable HR for the BM+ patients (PD-L1 positive) compared to BM– patients is HR = 0.91 (95% CI, 0.75–1.30;  $p$ -value = 0.91). When the model includes treatment, the adjusted HR for PD-L1-positive versus PD-L1-negative patients, adjusting for adjuvant treatment (chemotherapy versus none), is HR = 1.01 (95% CI, 0.76–1.35;  $p$ -value 0.93) [1]. The conclusion would be that PD-L1 status (positive versus negative) is not associated with overall survival in early-stage NSCLC patients because there is no significant association between PD-L1 status and overall survival, even after adjusting for treatment.

## Biomarkers in Clinical Trials

A biomarker refers to a measurable indicator of a biological state. In cancer this includes indicators of cancer presence, of prognosis for patients with cancer, and of disease response to a specific treatment. A biomarker can be a single measurement

**Table 2.2** Univariable and multivariable estimates of the HRs (with 95% CIs) and  $p$ -values for treatment group and biomarker status group. The univariable values are the same as in Table 2.1 and are the estimate of the HR for models that only have the indicated variable. The multivariable estimates come from a model that contains both variables at the same time

Variable	Univariable models		Multivariable model	
	HR (95% CI)	$p$ -value	HR (95% CI)	$p$ -value
Treatment group		0.0038		0.0039
Control	1.00 (reference)		1.00 (reference)	
Treatment	0.62 (0.45, 0.86)		0.62 (0.45, 0.86)	
BM status		0.33		0.35
BM–	1.00 (reference)		1.00 (reference)	
BM+	0.85 (0.61, 1.18)		0.85 (0.61, 1.19)	

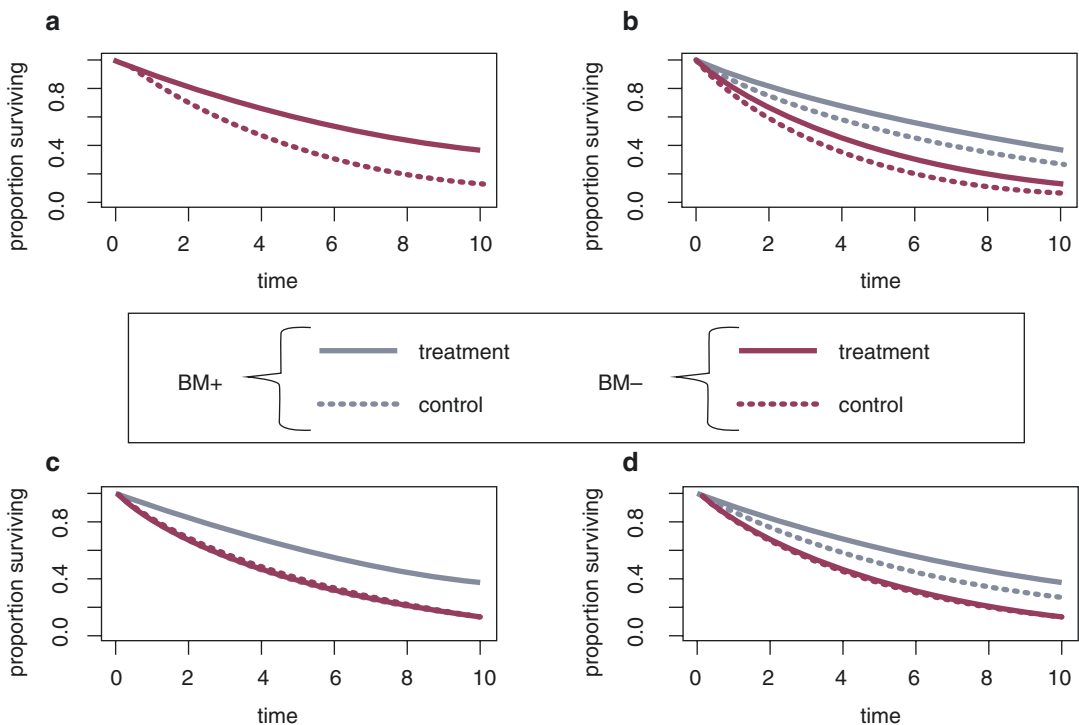
(e.g., PSA level for men), or it can be computed from numerous measurements (e.g., Oncotype Dx for women with early-stage breast cancer which is based on 21 genes). The two types of biomarkers commonly used in cancer clinical trials are *prognostic* and *predictive* biomarkers.

A prognostic biomarker informs about a likely cancer outcome regardless of what treatment a patient receives (including no treatment); it is thought to reflect the natural history of the disease. In other words, a prognostic biomarker is significantly associated with survival when adjusting for treatment a patient received. In Fig. 2.6b it can be seen that the biomarker is associated with survival for patients in the treatment group and for patients in the control group (Table 2.3).

The magnitudes of the association of the biomarker and survival are the same for both groups.

In Fig. 2.6d, it also can be seen that there is an association between the biomarker and survival for both groups. The difference between the scenarios depicted in Fig. 2.6d and that in 2.6b is that the magnitude of the association between the biomarker and survival depends on the treatment a patient received. For patients in the treatment arm, the magnitude of the biomarker association with survival is larger than for patients in the control group. In summary, if a biomarker is prognostic, there will be an association of the biomarker and survival regardless of treatment. If the magnitude of the association is the same in the groups, the biomarker is *purely prognostic*. If the magnitude differs between groups, the biomarker is both *prognostic and predictive*.

A biomarker is *predictive* when the treatment effect differs for BM+ patients and BM- patients. Figure 2.6c shows an association



**Fig. 2.6** Kaplan-Meier curves for different groups of patients where the color of the line denotes the biomarker group (BM+ is gray and BM- is maroon) and the line type denotes the treatment group (solid is the treatment group and dashed is the control group). (a) Illustrates the situation where the biomarker is neither prognostic nor

predictive. (b) Illustrates the situation where the biomarker is prognostic but not predictive. (c) Illustrates the situation where the biomarker is predictive but not prognostic. (d) Illustrates the situation where the biomarker is both prognostic and predictive



**Table 2.3** Definitions of different types of biomarkers with published examples of each

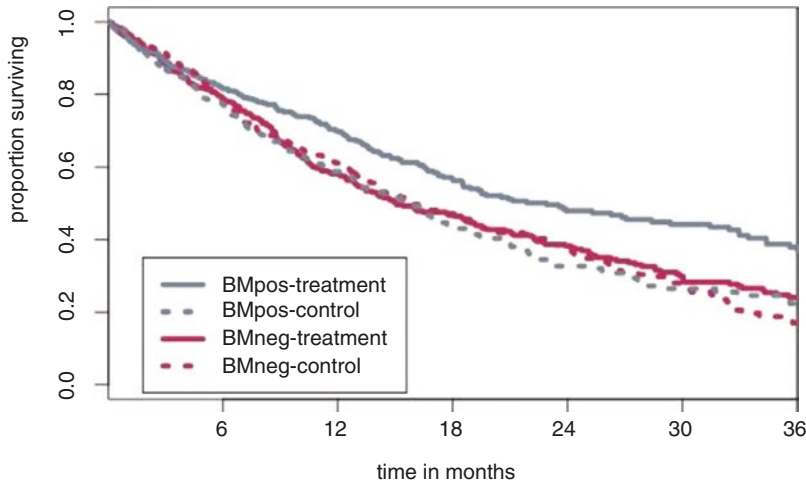
Biomarker	Definition	Example	Reference
Prognostic	A prognostic biomarker informs about a likely cancer outcome regardless of what treatment a patient receives (including no treatment)	Evaluation of PIK3CA mutation status for women with HER2-positive metastatic breast cancer	Baselga et al. [2]
Predictive	A biomarker is <i>predictive</i> when the treatment effect differs for biomarker-positive patients (BM+) and biomarker-negative (BM-) patients	RAS mutational status for treatment of metastatic colorectal cancer with an anti-EGFR antibody (cetuximab)	Van Cutsem et al. [3]
Prognostic and predictive	There is an association between the biomarker and survival for patients in the treatment and control groups	EGFR mutation status in NSCLC patients	Brugger et al. [4]
Neither prognostic nor predictive	Treatment is associated with survival, but within each treatment group, there is no association of the biomarker with survival	PD-L1 in early-stage NSCLC study	Tsao et al. [1]

between treatment and survival for BM+ patients; it appears as though patients in the treatment group have longer survival than patients in the control group. However, for BM- patients there is no association between treatment and survival. The same is true for Fig. 2.6d, where there appears to be a treatment benefit for BM+ patients but no treatment benefit for BM- patients. The difference between Fig. 2.6c, d is that the biomarker is *purely predictive* (and not prognostic) in Fig. 2.6c: there is no association between the biomarker and survival for patients in the control group. In Fig. 2.6d there is an association between the biomarker and survival for patients in the treatment and control groups indicating the biomarker is both predictive and prognostic. Figure 2.6a shows a case where the biomarker is neither predictive nor prognostic. Clearly, treatment is associated with survival, but within each treatment group, there is no association of the biomarker with survival.

In the era of precision medicine or individualized treatment, predictive biomarkers are more useful than prognostic biomarkers because they can be used to determine which patient will derive benefit from a treatment (say BM+ patients) and which will not (say BM- patients). In this case, a BM+ patient would receive the

treatment because he/she would likely garner benefit, and a BM- patient would not be treated because he/she would potentially experience adverse events with no benefit. The goal is to discover and validate more predictive biomarkers so that patients are treated with regimens from which they benefit and spared those from which they will not benefit and may only be harmed.

KM curves such as those in Fig. 2.6 can be used to gain a preliminary indication of whether a biomarker is potentially predictive. To be able to evaluate if a biomarker is predictive, all four groups of patients are necessary: BM+ treated with drug of interest, BM- patients treated with drug of interest, BM+ patients treated with control, and BM- patients treated with control. A biomarker is potentially predictive if the treatment is associated with survival in one biomarker group (e.g., BM+) and not the other (e.g., BM-). However, this is not sufficient. There needs to be a formal test of whether the treatment effect differs between the different biomarker groups. Such a test is performed with a statistical model, such as the Cox model for a survival outcome. The model contains the explanatory variables of treatment group and biomarker status with the addition of a variable for the interaction between the treatment and biomarker, the treatment by



**Fig. 2.7** A Kaplan-Meier plot summarizing the survival results for the four different biomarker and treatment group combinations. This plot suggests that the biomarker is predictive because BM+ patients derive benefit from

treatment and BM- patients do not. The predictive nature of the biomarker is confirmed with a statistically significant biomarker-by-treatment interaction term in the Cox model ( $p$ -value for interaction = 0.0049)

biomarker interaction variable. To determine whether a biomarker is predictive, the treatment-by-biomarker interaction term in the Cox model needs to be statistically significant (e.g.,  $p$ -value < 0.05). A significant treatment-by-biomarker interaction term indicates that the treatment effect differs by the biomarker group.

A Cox model that tests for an interaction between treatment groups by biomarker status will have three variables: treatment group, biomarker status, and the treatment-by-biomarker interaction. It is difficult to interpret and visualize the impact of the biomarker, treatment, and interaction based on the Cox model alone. In particular, the crude HRs that is produced by the software does not correspond to any of the four biomarker-by-treatment groups; the HRs for each of the four groups (one of which will be the reference group) are functions of the HRs of the model variables. KM curves can aid in understanding the relationship. Figure 2.7 contains the KM curves that correspond to a study of biomarkers and treatment. It appears as though BM+ patients drive benefit from treatment but BM- patients do not. The interaction term from the corresponding Cox model is statistically significant,  $p$ -value = 0.0049, indicating the biomarker is predictive.

If the treatment-by-biomarker interaction term is not statistically significant, then there is no evidence that the biomarker is predictive, even if it is the case that the log-rank test for treatment benefit is statistically significant in the BM+ group and not statistically significant in the BM- group. Often, investigators only analyze patients who were all treated with the drug of interest and conclude a biomarker is predictive if there is an association between the biomarker and survival. This is an inappropriate conclusion. Note that in Fig. 2.6b, for patients in the treatment group, there is an association between the biomarker and survival, BUT this is a purely prognostic biomarker because there is also an association between the biomarker and survival in the control group. Using only patients treated with the treatment of interest, it cannot be determined whether the situation is that in Fig. 2.6b (purely prognostic), Fig. 2.6c (purely predictive), or Fig. 2.6d (both prognostic and predictive).

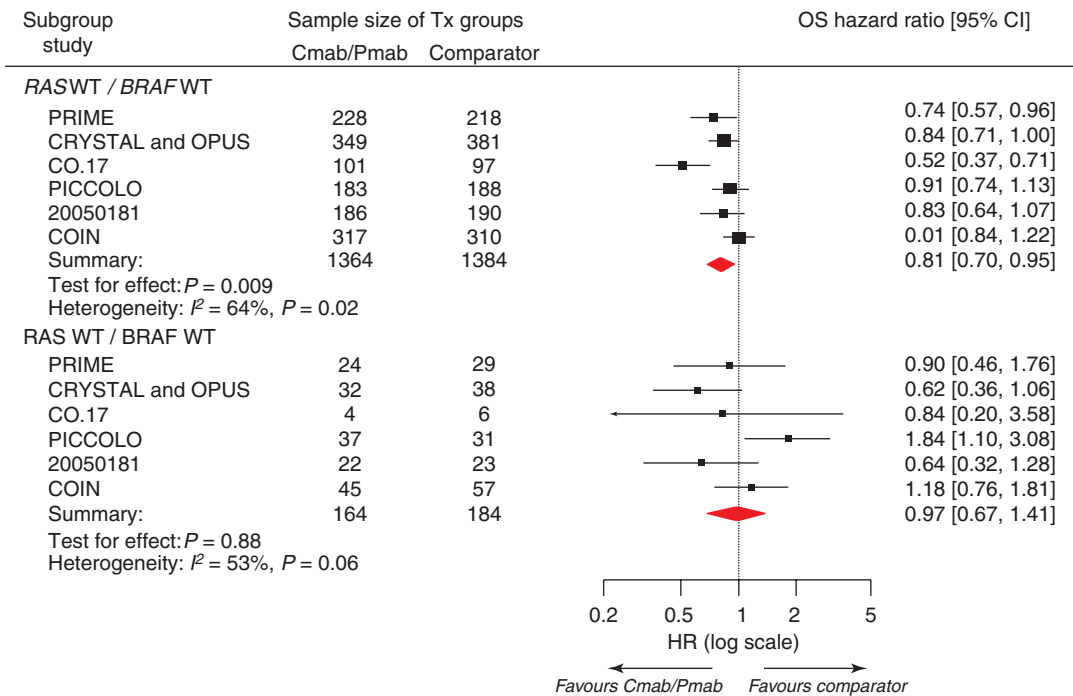
### The Use of Forest Plots

Often *meta-analysis* studies of predictive or prognostic biomarkers are conducted in order to garner more power, especially for testing for a

biomarker status by treatment interaction that is required to establish a biomarker is predictive. A forest plot is a graphical display of estimated results from randomized trials that investigate the same question. A forest plot typically lists the names of the included trials on the left-hand side. The content of the plot is the measure of the effect, which for overall survival is the HR, for each of the studies. The confidence intervals for the effect estimate is represented by horizontal lines and is often the numerical values for the effect estimate and confidence interval boundaries are provided on the right-hand side of the graphic. The graph may be plotted on the logarithmic scale when using a HR so that the confidence intervals are symmetric around the estimated effect. Each square is centered on the effect size, and the area of the square is proportional to the size of the study, which dictates the study's weight or influence in the analysis. The overall meta-analysis estimate of effect is represented by a diamond, with the width of the diamond corresponding to the confidence interval. A

vertical line corresponding to no effect (e.g., HR = 1) is often plotted.

Figure 2.8 is a forest plot taken from a study performed by Rowland et al. [5]. The authors performed a meta-analysis of randomized clinical trials that evaluated the effect of BRAF V600E mutation status, mutated (MT) versus wild type (WT), and benefit from anti-EGFR monoclonal antibody treatment (anti-EGFR mAb) in patients with metastatic colorectal cancer that was RAS wild type. From the figure, it can be seen that within these studies, patients with BRAF WT tumors obtained benefit from anti-EGFR mAb treatment, with a few studies yielding statistically significant results. On the other hand, it appears as though patient with BRAF MT tumors did not garner benefit from anti-EGFR mAb treatment with none of the studies having statistically significant results in this group. The meta-analysis estimate of anti-EGFR mAb benefit in patients with BRAF WT tumors is 0.81 (95% CI, 0.70–0.95;  $p$ -value = 0.009) and in patients with BRAF MT tumors is 0.97 (95% CI, 0.67–1.41;



**Fig. 2.8** A forest plot taken from Rowland et al. [5]. Note that there is an error in this plot in that the bottom portion is for RAS WT/BRAF MT patients. (Reprinted from Rowland et al. [5]. With permission from Nature Publishing Group)

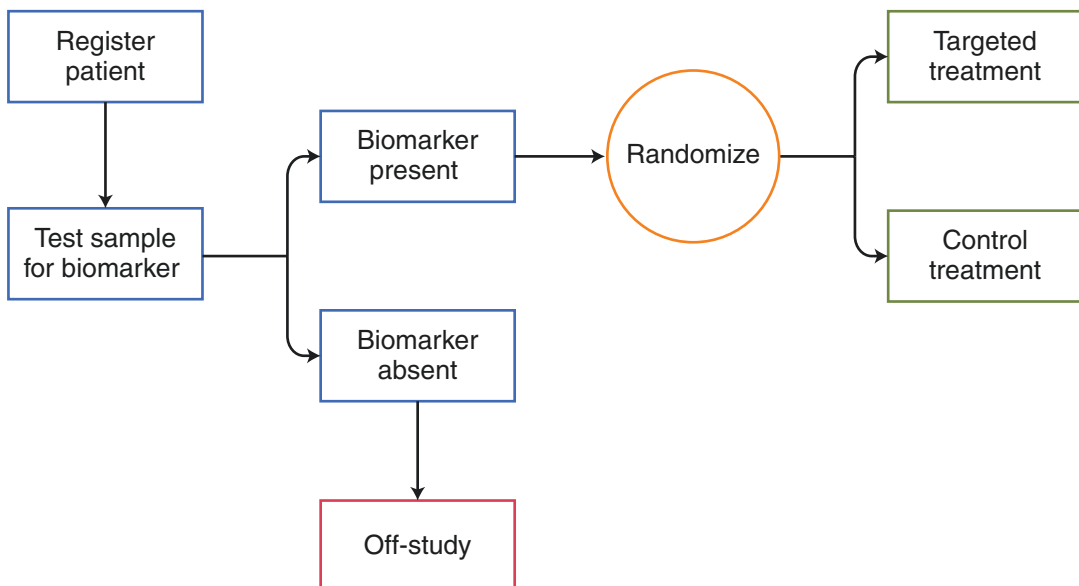
$p$ -value = 0.88). Although there appears to be differential treatment effects in the two biomarker groups, the test for interaction between BRAF status (WT versus MT) and treatment (anti-EGFR mAb treatment versus no anti-EGFR mAb treatment) was not statistically significant,  $p$ -value = 0.43. Hence, there is no evidence from this study that BRAF mutation status is a predictive biomarker for benefit from anti-EGFR mAb in patients with RAS WT metastatic colorectal cancer.

### Biomarker Clinical Trial Designs

There are numerous clinical trial designs that incorporate biomarkers, validate biomarkers, and discover biomarkers. The *enrichment design* is used when there is compelling evidence that treatment benefit (if any) will be restricted to a subgroup of patients who do (or do not) have a particular biomarker. In this design, all patients are screened for the biomarker, and only those in the subgroup of interest (either have or do not

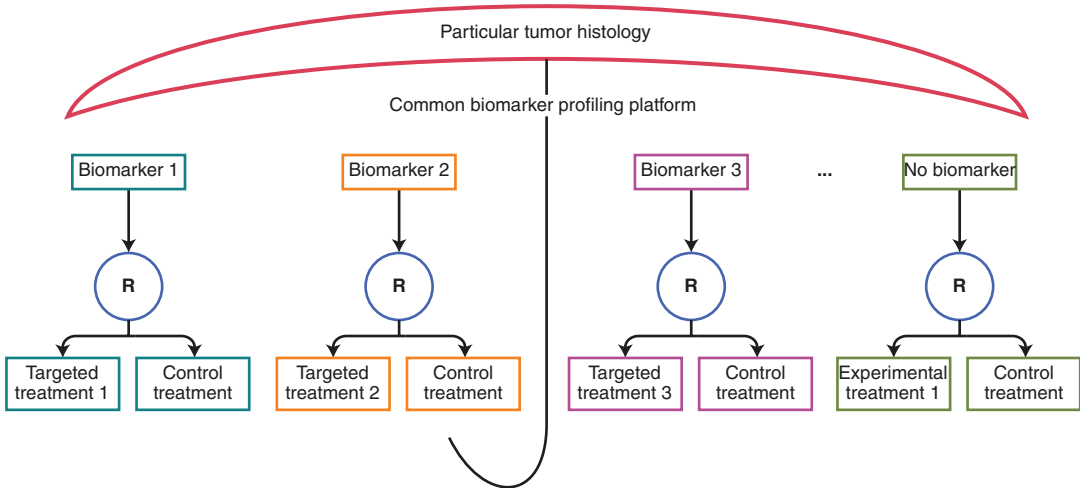
have the biomarker) are enrolled on the trial (see Fig. 2.9).

This trial design cannot validate whether the biomarker is predictive for the treatment benefit since all patients are in the same biomarker subgroup. It can only provide evidence whether there is a treatment benefit in the selected biomarker subgroup. If there is benefit, it is unknown whether patients in the nonselected biomarker group may also have derived treatment benefit. Such a design should only be used in cases where there is persuasive evidence that the biomarker is predictive. A successful example of the use of this design was the trials for trastuzumab in patients with HER2+ breast cancer: the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-31 and the North Central Cancer Treatment Group (NCCTG) N9831 trials [6]. These trials only included women with tumors that were found to be HER2 positive. There were strong preclinical data to indicate that only these patients would derive benefit from trastuzumab. The trials were successful and led to FDA approval for the use of trastuzumab to treat



**Fig. 2.9** A diagram of the schema for an *enrichment* trial design. Patients are registered (and consented) prior to their sample being tested for the biomarker. If the biomarker is “present” (either deemed positive or negative),

the patient is then enrolled and randomized to the targeted treatment or the control treatment (usually standard of care). If the biomarker is “absent,” the patient goes off-study and is no longer followed



**Fig. 2.10** A diagram for an umbrella trial. Tumor of a specific histologic type is tested for a panel of biomarkers on a common testing platform. Tumors that have biomarkers of interest are then randomized to a treatment

that targets the biomarker or to a control treatment. If tumors have none of the biomarkers of interest, they either are off-study or are randomized between control and another (untargeted) experimental treatment

HER2-positive breast cancer in the adjuvant setting. The question of whether patients with HER2-negative tumors would benefit from trastuzumab is currently being investigated.

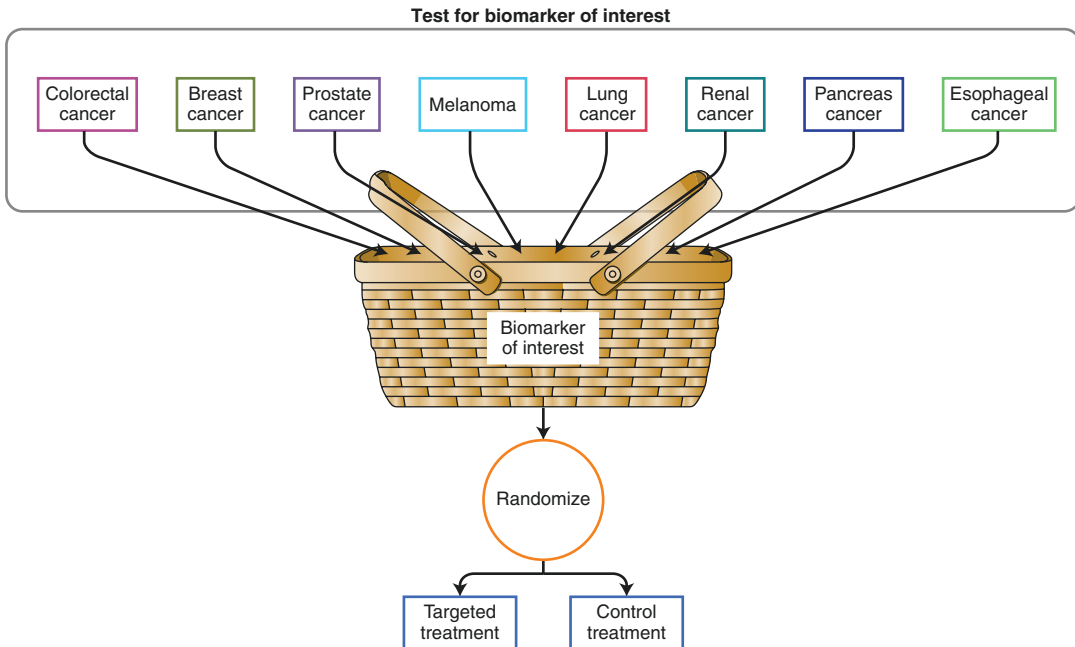
Two different enrichment designs have recently gained popularity: the umbrella trial and the basket (or bucket trial). The *umbrella design* tests the treatment benefit of multiple drugs on different mutations in a single tumor type or histology (see Fig. 2.10).

It provides a common infrastructure to facilitate patient screening and accrual. Patients are assigned or randomized to treatment arms based on their biomarker status. The intent of the trial is to evaluate the benefit of different drugs matched to their mutation in a single type of cancer. The biomarker testing is usually done at a central location prior to patient enrollment and randomization. Examples of recent umbrella trials include I-SPY2 [7, 8], BATTLE [9, 10], and Lung-MAP [11]. A *basket* or *bucket* trial includes cancers of different types that each has the same biomarker of interest (see Fig. 2.11).

This trial design tests the benefit of a treatment for which the biomarker is thought to be predictive. The design includes many different cancer types that belong to the same biomarker subgroup, and one targeted treatment (usually) is tested.

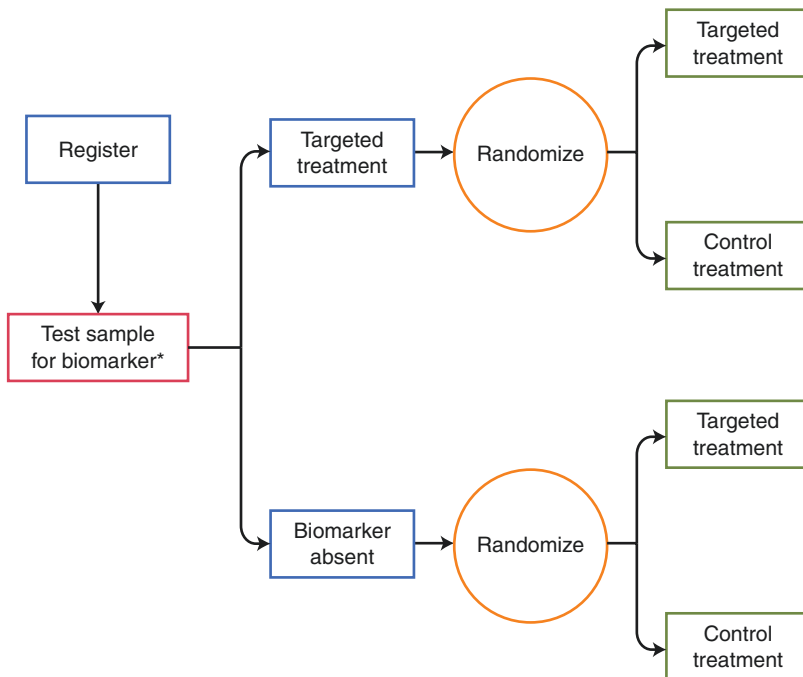
Patients are tested for the biomarker prior to enrollment to the trial since the biomarker subgroup is an eligibility criterion. Examples of basket trials are MPACT [12], MATCH [13], and a vemurafenib trial for cancers with BRAF V600 mutations [14]. These are versions of enrichment trials and are designed to realize benefits of efficiency of using a single platform (umbrella trial) or to increase the number of patients eligible for treatment with a particular biomarker and to determine if the benefit is similar across tumor types (basket).

The *all-comer* (or *unselected*) design tests all patients for their biomarker status and enrolls all patients regardless of biomarker status. An eligibility criterion for this trial is adequate specimen availability and quality to perform the biomarker assay. The patients are randomized to the same set of treatment arms, for all the biomarker groups (see Fig. 2.12). The SATURN (sequential Tarceva in unresectable non-small lung cancer) trial [15] is an example of an all-comer trial. In this trial, all eligible NSCLC patients were randomly assigned to erlotinib or placebo plus standard of care, regardless of the EGFR status of their tumor. The trial was designed to evaluate the efficacy of erlotinib in all randomized patients as well as in the subgroup of patients that had EGFR-positive tumors.



**Fig. 2.11** A diagram for a basket trial. All types of tumors are tested for a specific biomarker. If they have the biomarker of interest, they are randomized to a targeted treatment or to a control treatment. If they do not have the biomarker of interest, they are not registered to the trial

ment or to a control treatment. If they do not have the biomarker of interest, they are not registered to the trial



\*Can be tested at any point prior to analysis

**Fig. 2.12** A diagram of the schema for an *all-comers* trial design. Patients are registered and entered onto the trial regardless of their biomarker status. In the diagram, they are tested prior to randomization, but this does not have to

be the case; the biomarker status only needs to be known prior to the analysis of the trial data. Both types of patients, those with the biomarker present or absent, are randomized to targeted treatment or to control treatment

The test for the biomarker can be performed before or after randomization. If the biomarker is a *stratification variable*, then to ensure the same distribution of biomarker subgroups among the treatment arms, it needs to be performed prior to patient randomization. If it is not used as a stratification factor, it can be performed at any time prior to the pre-planned trial analyses. There are several different ways the trial data could be analyzed, but the analysis method must be pre-specified at the time of trial design. If the primary interest is to validate that the biomarker is predictive, a biomarker by treatment interaction analysis will be the primary analysis. This formally tests for a biomarker by treatment interaction term in a Cox model as described above.

Another type of analyses determines which patient subgroups defined by the biomarker benefit from treatment, if any, by performing sequential analyses. One approach is to test for a treatment effect in the entire trial cohort (ignoring biomarker group). If this is not significant, then a test of treatment benefit will be done in a planned biomarker subset, which is the subset thought to be the most likely to derive benefit a priori. Another approach is to first test for treatment benefit in a biomarker subset (the one with the strongest a priori evidence it would benefit), and if this is statistically significant, perform a test of treatment benefit in the entire clinical trial cohort. The type of analysis plan that will be done is pre-specified during the trial planning stage, and the level of significance used for the planned sequential analyses are set to ensure the overall trial type I error is maintained at 0.05.

It is best to use the marker-by-treatment interaction analysis when there is uncertainty whether the biomarker is predictive or not. However, this analysis requires the largest sample size. The sequential testing approaches are also relevant for situations where there is uncertainty of whether the biomarker is predictive or not, but they are not powered to detect a biomarker by treatment interaction. The intent for the latter two approaches is to find subgroup(s) that benefit from treatment without formally establishing whether the biomarker is predictive. These trials are generally a bit smaller than what is needed for the maker-by-treatment interaction analysis.

Finally, there are refinements to the designs discussed above that incorporate a Bayesian aspect to perform exploratory analyses meant to discover biomarkers as the trial proceeds. These designs are sometimes called *exploratory platform designs* and usually are early phase (I or II) trials. Such designs are useful when there is uncertainty regarding the best biomarkers for the treatments under study. In this design, drug arms are pre-specified, and patients are initially randomized equally across the arms, regardless of the biomarker status of their tumor. Biomarker testing is performed on a tumor biopsy prior to randomized, and pre-specified biomarker cohorts are stratified evenly across treatment arms. After a sufficient number of patients have been assigned to each arm, the efficacy for each biomarker-treatment combination is evaluated, and the randomization is adapted so that future patients have a higher probability of being assigned to a treatment group that appears favorable for the biomarker in their tumor. Drugs that do not appear to be beneficial for any biomarker group are dropped. Biomarker-treatment combinations that surpass a pre-defined threshold of efficacy are brought forward in a larger enrichment trial (e.g., phase II or III). In these trials, only patients with tumors that have the identified biomarker are enrolled, and the patients are randomized to the experimental treatment or standard of care. Examples of exploratory platform trials with Bayesian adaptive randomization are BATTLE [16], for patients with previously treated lung cancer, and I-SPY2 [17], a neoadjuvant trial for breast cancer patients.

## Concluding Remarks

For cancer treatments to be more individualized to patient and/or disease characteristics, it is necessary to develop predictive biomarkers. However, the success rate for finding predictive biomarkers has been disappointing. To increase the success rate, it is important to understand the evidence that is needed to determine whether a biomarker is predictive of treatment benefit. It is also important to understand the different roles of biomarkers in clinical trials and the implications of the different clinical trial designs for the evaluation of biomarkers.

## Glossary

### **Adjusted (or multivariable hazards ratio**

**(HR)** A multivariable Cox model allows the evaluation of the association of multiple variables on the outcome (e.g., survival). This allows a more accurate assessment of the relationship of a variable of interest to overall survival by accounting for other variables that may be associated with survival. For example, when evaluating the association of a biomarker with survival, a treatment variable may be added to the model. This would allow the evaluation of the association of the biomarker with survival after accounting for the association of treatment with survival. The hazard ratio for a variable from a multivariable Cox model is referred to as a multivariable HR or an adjusted HR.

**Continuous (bio)marker** A continuous biomarker is one that has an infinite number of possibilities; in other words, it can take on any value between its minimum and maximum value if it could be measured to any desired degree of precision. An example of a continuous biomarker is PSA level for prostate cancer. The minimum value is 0 and there is no absolute maximum. If PSA could be measured to any desired degree of precision, all nonnegative values are possible.

**Cox proportional hazards model** A Cox proportional hazards model is a regression technique for time-to-event data (e.g., survival) where there is censoring (when some patients are alive at the time of analysis). It is a way to evaluate the association of a variable with the time-to-event outcome such as survival. The method is semi-parametric; that is, it does not assume a model for  $t$  survival but does assume that the effect of a variable on survival is constant over time. The association is measured by a hazard ratio (HR) where  $HR = 1$  means no association, a  $HR < 1$  means increasing values of the variable reduces the chance of death, and  $HR > 1$  means that increasing value of the variable increases the chance of death.

**Dichotomous (bio)marker** A dichotomous biomarker is one that takes one of two possible values. It is used to split patient cohorts into two categories or groups. An example of a

dichotomous biomarker is estrogen receptor (ER) status for women with breast cancer: ER positive versus ER negative.

**Log-rank test** A log-rank test is used to compare the survival distributions of two or more groups. The null hypothesis is that there is no difference among the groups. If the  $p$ -value is significant (e.g., less than 0.05), this is evidence that the groups have different survival experiences. Note this is only a test for a difference among the survival experiences and does not provide an estimate regarding the size of the differences between any two groups.

**Meta-analysis** A meta-analysis encompasses techniques for combining data from multiple studies. An underlying assumption is that the treatment effect is consistent across studies and combining results across studies yields increased power. Most meta-analysis approaches essentially compute a weighted average from the results of the individual studies, and larger studies tend to be given more weight.

**Randomization or random assignment** In randomized trials, the participants are assigned by chance to the treatment groups (arms) rather than by choice. Randomization serves to make the groups similar with respect to variables (e.g., patient characteristics, tumor traits) other than the treatment. This means if differences are observed for the outcome variable (primary endpoint), it can be attributable to the treatment since the groups balanced for the other variables. Randomization is accomplished with a chance procedure (e.g., flipping a coin) or a random number generator.

**Stratification variable** A stratification variable in a clinical trial is a variable that is used to group patients into strata corresponding to the values of the variable. Randomization is performed separately within each stratum. An example of a stratification variable is whether a patient has disease in his/her lymph nodes or not (e.g., lymph node status with values of lymph node positive and lymph node negative). Variables selected for stratification are those where it is important there is no imbalance between the treatment arms because they are highly prognostic of outcome.



**Type I error** Type I error is the error that occurs when the null hypothesis is rejected although it is true. It is a false-positive result. For example, suppose in reality there is no difference between the experimental treatment and standard of care with respect to overall survival. However, a clinical trial is performed, and it is found that the treatment arm had superior survival compared to the standard of care arm with a  $p$ -value of 0.03. The investigators conclude that the experimental treatment is better than the standard of care. In reality, this is an incorrect conclusion and an example of a type I error. (Note that the investigators would not know that their conclusion is incorrect.)

**Univariable hazards ratio (HR)** A univariable hazard ratio is the ratio of hazard rates for an event (e.g., death) corresponding to the different values of one variable of interest. For example, in a Cox model that contains only a treatment variable (experimental versus control), a HR = 0.50 for survival indicates that patients in the treatment group die at half the rate per unit of time as patients in the control group.

## References

1. Tsao MS, Le Teuff G, Shepherd FA, Landais C, Hainaut P, Filipits M, Pirker R, Le Chevalier T, Graziano S, Kratze R, Soria JC, Pignon JP, Seymour L, Brambilla E. PD-L1 protein expression assessed by immunohistochemistry is neither prognostic nor predictive of benefit from adjuvant chemotherapy in resected non-small cell lung cancer. *Ann Oncol*. 2017;28(4):882–9. <https://doi.org/10.1093/annonc/mdx003>. PubMed PMID: 28137741.
2. Baselga J, Cortés J, Im SA, Clark E, Ross G, Kiermaier A, Swain SM. Biomarker analyses in CLEOPATRA: a phase III, placebo-controlled study of pertuzumab in human epidermal growth factor receptor 2-positive, first-line metastatic breast cancer. *J Clin Oncol*. 2014;32(33):3753–61. <https://doi.org/10.1200/JCO.2013.54.5384>. Epub 2014 Oct 20. PubMed PMID: 25332247.
3. Van Cutsem E, Lenz HJ, Köhne CH, Heinemann V, Tejpar S, Melezínek I, Beier F, Stroh C, Rougier P, van Krieken JH, Ciardiello F. Fluorouracil, leucovorin, and irinotecan plus cetuximab treatment and RAS mutations in colorectal cancer. *J Clin Oncol*. 2015;33(7):692–700. <https://doi.org/10.1200/JCO.2014.59.4812>. Epub 2015 Jan 20. PubMed PMID: 25605843.
4. Brugger W, Triller N, Blasinska-Morawiec M, Curescu S, Sakalauskas R, Manikhas GM, Mazieres J, Whittom R, Ward C, Mayne K, Trunzer K, Cappuzzo F. Prospective molecular marker analyses of EGFR and KRAS from a randomized, placebo-controlled study of erlotinib maintenance therapy in advanced non-small-cell lung cancer. *J Clin Oncol*. 2011;29(31):4113–20. <https://doi.org/10.1200/JCO.2010.31.8162>. Epub 2011 Oct 3. Erratum in: *J Clin Oncol*. 2011 Dec 10;29(35):4725. PubMed PMID: 21969500.
5. Rowland A, Dias MM, Wiese MD, Kichenadasse G, McKinnon RA, Karapetis CS, Sorich MJ. Meta-analysis of BRAF mutation as a predictive biomarker of benefit from anti-EGFR monoclonal antibody therapy for RAS wild-type metastatic colorectal cancer. *Br J Cancer*. 2015;112(12):1888–94. <https://doi.org/10.1038/bjc.2015.173>. Epub 2015 May 19. Review. PubMed PMID: 25989278; PubMed Central PMCID: PMC4580381.
6. Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE Jr, Davidson NE, Tan-Chiu E, Martino S, Paik S, Kaufman PA, Swain SM, Pisansky TM, Fehrenbacher L, Kutteh LA, Vogel VG, Visscher DW, Yothers G, Jenkins RB, Brown AM, Dakhil SR, Mamounas EP, Lingle WL, Klein PM, Ingle JN, Wolmark N. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med*. 2005;353(16):1673–84. PubMed PMID: 16236738.
7. Rugo HS, Olopade OI, DeMichele A, Yau C, van 't Veer LJ, Buxton MB, Hogarth M, Hylton NM, Paoloni M, Perlmutter J, Symmans WF, Yee D, Chien AJ, Wallace AM, Kaplan HG, Boughey JC, Haddad TC, Albain KS, Liu MC, Isaacs C, Khan QJ, Lang JE, Viscusi RK, Pusztai L, Moulder SL, Chui SY, Kemmer KA, Elias AD, Edmiston KK, Euhus DM, Haley BB, Nanda R, Northfelt DW, Tripathy D, Wood WC, Ewing C, Schwab R, Lyandres J, Davis SE, Hirst GL, Sanil A, Berry DA, Esserman LJ, I-SPY 2 Investigators. Adaptive randomization of veliparib-carboplatin treatment in breast cancer. *N Engl J Med*. 2016;375(1):23–34. <https://doi.org/10.1056/NEJMoa1513749>.
8. Park JW, Liu MC, Yee D, Yau C, van 't Veer LJ, Symmans WF, Paoloni M, Perlmutter J, Hylton NM, Hogarth M, DeMichele A, Buxton MB, Chien AJ, Wallace AM, Boughey JC, Haddad TC, Chui SY, Kemmer KA, Kaplan HG, Isaacs C, Nanda R, Tripathy D, Albain KS, Edmiston KK, Elias AD, Northfelt DW, Pusztai L, Moulder SL, Lang JE, Viscusi RK, Euhus DM, Haley BB, Khan QJ, Wood WC, Melisko M, Schwab R, Helsten T, Lyandres J, Davis SE, Hirst GL, Sanil A, Esserman LJ, Berry DA, I-SPY 2 Investigators. Adaptive randomization of Neratinib in early breast cancer. *N Engl J Med*. 2016;375(1):11–22. <https://doi.org/10.1056/NEJMoa1513750>.
9. Kim ES, Herbst RS, Wistuba II, Jack Lee J, Blumenschein GR Jr, Tsao A, Stewart DJ, Hicks ME, Erasmus J Jr, Gupta S, Alden CM, Liu S, Tang X, Khuri FR, Tran HT, Johnson BE, Heymach JV, Li M, Fossella F, Kies MS, Papadimitrakopoulou V,

- Davis SE, Lippman SM, Hong WK. The BATTLE trial: personalizing therapy for lung cancer. *Cancer Discov.* 2011; <https://doi.org/10.1158/2159-8274.CD-10-0010>.
10. Papadimitrakopoulou V, Jack Lee J, Wistuba II, Tsao AS, Fossella FV, Kalhor N, Gupta S, Byers LA, Izzo JG, Gettinger SN, Goldberg SB, Tang X, Miller VA, Skoulidis F, Gibbons DL, Li S, Wei C, Diao L, Andrew Peng S, Wang J, Tam AL, Coombes KR, Ja SK, Mauro DJ, Rubin EH, Heymach JV, Hong WK, Herbst RS. The BATTLE-2 study: a biomarker-integrated targeted therapy study in previously treated patients with advanced non-small-cell lung cancer. *JCO.* 2016;34(30):3638–47.
  11. Steuer CEI, Papadimitrakopoulou V, Herbst RS, Redman MW, Hirsch FR, Mack PC, Ramalingam SS, Gandara DR. Innovative clinical trials: the LUNG-MAP study. *Clin Pharmacol Ther.* 2015;97(5):488–91. <https://doi.org/10.1002/cpt.88>.
  12. Lih CJ, Sims DJ, Harrington RD, Polley EC, Zhao Y, Mehaffey MG, Forbes TD, Das B, Walsh WD, Datta V, Harper KN, Bouk CH, Rubinstein LV, Simon RM, Conley BA, Chen AP, Kummur S, Doroshow JH, Williams PM. Analytical validation and application of a targeted next-generation sequencing mutation-detection assay for use in treatment assignment in the NCI-MPACT trial. *J Mol Diagn.* 2016;18(1):51–67. <https://doi.org/10.1016/j.jmoldx.2015.07.006>.
  13. Moore KN, Mannel RS. Is the NCI MATCH trial a match for gynecologic oncology? *Gynecol Oncol.* 2016;140(1):161–6. <https://doi.org/10.1016/j.ygyno.2015.11.003>. Review.
  14. Cappuzzo F, Ciuleanu T, Stelmakh L, Cicenias S, Szczesna A, Juhász E, Esteban E, Molinier O, Brugger W, Melezínek I, Klingelschmitt G, Klughammer B, Giaccone G. SATURN investigators. Erlotinib as maintenance treatment in advanced non-small-cell lung cancer: a multicentre, randomised, placebo-controlled phase 3 study. *Lancet Oncol.* 2010;11(6):521–9. [https://doi.org/10.1016/S1470-2045\(10\)70112-1](https://doi.org/10.1016/S1470-2045(10)70112-1). Epub 2010 May 20. PubMed PMID: 20493771.
  15. Hyman DM, Pazanov I, Subbiah V, Faris JE, Chau I, Blay JY, Wolf J, Raje NS, Diamond EL, Hollebecque A, Gervais R, Elez-Fernandez ME, Italiano A, Hofheinz RD, Hidalgo M, Chan E, Schuler M, Lasserre SF, Makrutzki M, Sirzen F, Veronese ML, Taberero J, Baselga J. Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. *N Engl J Med.* 2015;373:726–36. <https://doi.org/10.1056/NEJMoa150230>.
  16. Kim ES, Herbst RS, Wistuba II, Lee JJ, Blumenschein GR Jr, Tsao A, Stewart DJ, Hicks ME, Erasmus J Jr, Gupta S, Alden CM, Liu S, Tang X, Khuri FR, Tran HT, Johnson BE, Heymach JV, Mao L, Fossella F, Kies MS, Papadimitrakopoulou V, Davis SE, Lippman SM, Hong WK. The BATTLE trial: personalizing therapy for lung cancer. *Cancer Discov.* 2011;1(1):44–53. <https://doi.org/10.1158/2159-8274.CD-10-0010>. Epub 2011 Jun 1. PubMed PMID: 22586319; PubMed Central PMCID: PMC4211116.
  17. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther.* 2009;86(1):97–100. <https://doi.org/10.1038/clpt.2009.68>. Epub 2009 May 13. PubMed PMID: 19440188.